CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 124, No. 4, Pt. 1 of 2

October 2008

(Continued)

CONTENTS—*Continued from preceding page*

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

TRANSDUCTION [38]

STRUCTURAL ACOUSTICS AND VIBRATION [40]

NOISE: ITS EFFECTS AND CONTROL [50]

ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION [58]

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

(Continued)

CONTENTS—*Continued from preceding page*

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

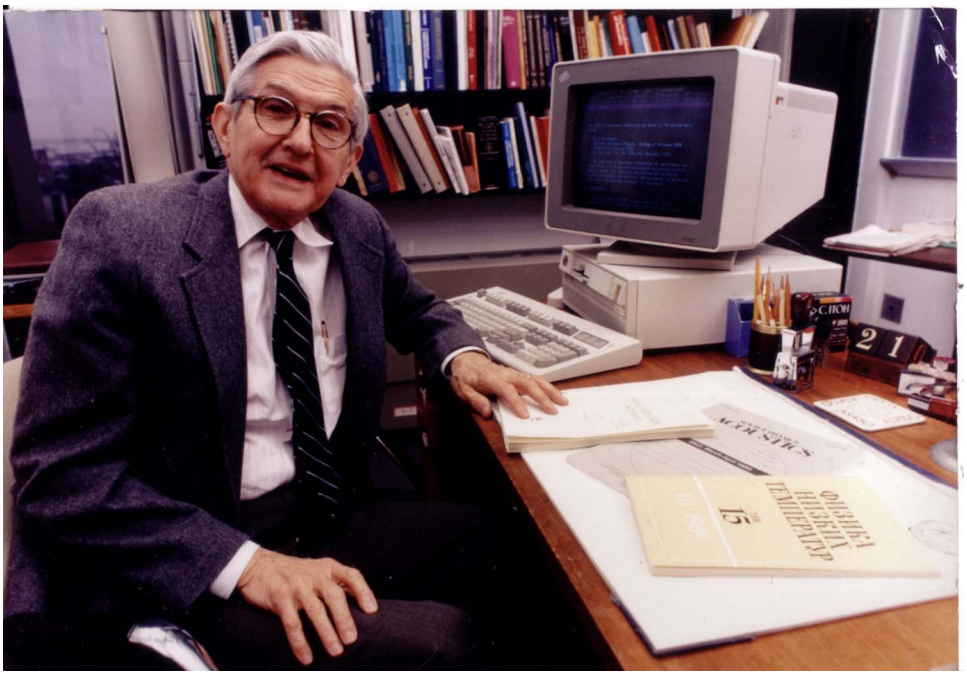
SPEECH PERCEPTION [71]

MUSIC AND MUSICAL INSTRUMENTS [75]

BIOACOUSTICS [80]

(Continued)

CONTENTS—*Continued from preceding page*



Robert T. Beyer
1920–2008

As this issue goes to press, we must report with sadness the death on Wednesday, 20 August, of Robert T. Beyer, whose distinguished career included extensive leadership and service to the Acoustical Society of America. He was President of the Society in 1968–1969, received the Gold Medal in 1984, and the Distinguished Service Citation in 1989. For two decades, 1974–1994, he served as Treasurer of the Society. After R. Bruce Lindsay's death in 1985, Beyer served in addition as the Editor-in-Chief ad interim, until Daniel Martin was selected as Lindsay's successor.

Beyer's funeral in Providence on Saturday, 23 August, was attended by many friends and colleagues, including four former students, all of whom are now Fellows of this Society. Other attendees included Peter Westervelt, Beyer's close former colleague in the Physics Department at Brown, and the upcoming recipient of the Society's Silver Medal in Physical Acoustics, and the Editor-in-Chief. Among the several messages to the family that were read at the collation following the service were ones by David Blackstock and Konstantin Naugolnykh.

Beyer's multi-faceted career will undoubtedly be further commemorated within future pages of the Society's publications and at future meetings of the Society.

Extracting the local Green's function on a horizontal array from ambient ocean noise

S. E. Fried and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego La Jolla, California 92093-0238
sefried@ucsd.edu, wkuperman@ucsd.edu

Karim G. Sabra
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405
karim.sabra@me.gatech.edu

Philippe Roux
Lab. de Géophysique Interne et Tectonophysique, Université J. Fourier & CNRS, BP53, 38041 Grenoble, France
philippe.roux@obs.ujf-grenoble.fr

Abstract: Using only ocean ambient noise recordings it is possible to approximate the local time domain Green's function (TDGF) and extract the time delays associated with different ray path between the elements of a bottom hydrophone array. Comparing the strength of the noise correlation function taken over increasing time windows with residual fluctuations points to an optimum time window to use in the noise correlation function. Through comparison with computer simulations the resulting time series is shown to accurately approximate noise responses in the environment. Analysis of the TDGF gives accurate environmental detail, specifically the critical angle at the water-sediment interface.

PACS numbers: 43.30.Re [JL]
Date Received: March 25, 2008 **Date Accepted:** June 14, 2008

1. Introduction

There has been significant work recently in extracting coherent signals from random noise data. Early work successfully extracted the time domain Green's function (TDGF) from an aluminum cube by correlating the Brownian noise recorded at sensors on two faces of the cube.[1] The same theory of correlating the random signals to extract their coherent aspects has been applied theoretically and experimentally to both land seismic and ocean environments. For land environments, seismograms were passively obtained in the stationary environments that accurately reflected active seismic results.[2,3] The noise correlation processing of ambient ocean noise has been shown to successfully extract coherent propagating wavefronts between hydrophone arrays[4] though correlation times are limited due to the dynamic medium. Practical applications of this technique include using a vertical array suspended from a towed buoy to create a passive fathometer[5] and array element localization of hydrophone receivers in a horizontal array.[6]

This paper will illustrate that it is possible to passively extract the local TDGF from ambient ocean noise recording taken along a bottom-mounted horizontal array where each receiver is turned into a virtual source. An ideal window of processing time is determined which allows for the maximum average signal correlation strength over the distance of the array. The resulting time series is shown to be consistent with previous active acoustic experiments of the region, and allows us to extract individual environmental details. This is illustrated by identifying the critical angle at the water-sediment interface.

2. Theory

The time derivative of the noise correlation function (NCF) has been proved both theoretically and experimentally[4,7,8] to be approximately the sum of the TDGF and the time-reversed TDGF between the two hydrophones[8]

$$\frac{d\langle C_{a,b}(\tau)\rangle}{d\tau} \approx -G(\vec{r}_a,\vec{r}_b;\tau) + G(\vec{r}_b,\vec{r}_a;-\tau),$$

where C is the cross correlation of the time-aligned signal between receivers a and b located at $\vec{r}_a$ and $\vec{r}_b$, respectively, and τ is the time delay between those two receivers. The cross correlation isolates and accumulates the time average of those source events in the environment which propagate through both hydrophones. The specifics of the length of time necessary for performing the cross correlation are dependent on the environment around the hydrophones and especially the distribution of noise sources.

The data used in the present analysis are the same data that were used for array element localization using the NCF technique.[6] In that analysis 11 min time averages of data were used to extract the coherent wavefronts between close (less than 28 m apart) hydrophone pairs. Here we have increased the length of time averaged in order to resolve the greatest distance between hydrophones (up to 115 m) on the array.

3. Experimental results

3.1 Data

The data used in this analysis were originally taken in May 1995, 3.4 km off the coast of southern California for the Adaptive Beach Monitoring experiment.[9,10] This work used the ambient noise data recordings from one of the two bottom mounted horizontal arrays located parallel to the shore in roughly 21 m of water. The environment around the hydrophone arrays was measured and modeled using active geoacoustic methods; two nearby conductivity, temperature, and density casts provided the sound speed profile for the water column. The measurements of the environment found this area to be essentially range independent. Each array of 64 hydrophones recorded almost continuously for 2.5 weeks at a sampling frequency of 1500 Hz. Throughout that time the hydrophones recorded significant biological activity, dominated by noise from the croaker fish (*Sciaenidae*) family.[9] The ambient noise levels were especially high at night as the fish migrated from the surf zone out to the area where the hydrophones were located. That the noise field was dominated by biological activity within the water column (as opposed to physical effects like waves breaking which localize sound at the surface) made this data set particularly useful for extracting the approximate TDGF. That is, because the noise sources were volumetric a more accurate representation of the amplitude of the time of arrivals can be expected versus having only surface noise sources.[4]

3.2 Results

Approximating the local Green's function using only the ambient noise recordings is limited by the characteristics of the noise that is recorded. Ideally the noise should be fairly isotropic, or at least spread throughout the area of the hydrophones. The recording should not contain individual loud events (such as source pings, motor noises, etc.) which may bias the correlation process. To that end, the noise recordings are filtered and smoothed (both the time series amplitude and in frequency spectrum) in order to minimize the influence of isolated unique loud events.

Deciding the length of time to use in the NCF is one of the challenges of this technique. To get the cleanest result the strength of the strongest arrival versus the residual fluctuations was tracked as successive additional successive 5 min samples were averaged into the correlation. The correlation of hydrophones with a greater distance between each other required a longer time correlation to produce a clear result. The changing ocean environment (and coastal pumps which turned on and off every few hours) meant using as short a continuous time

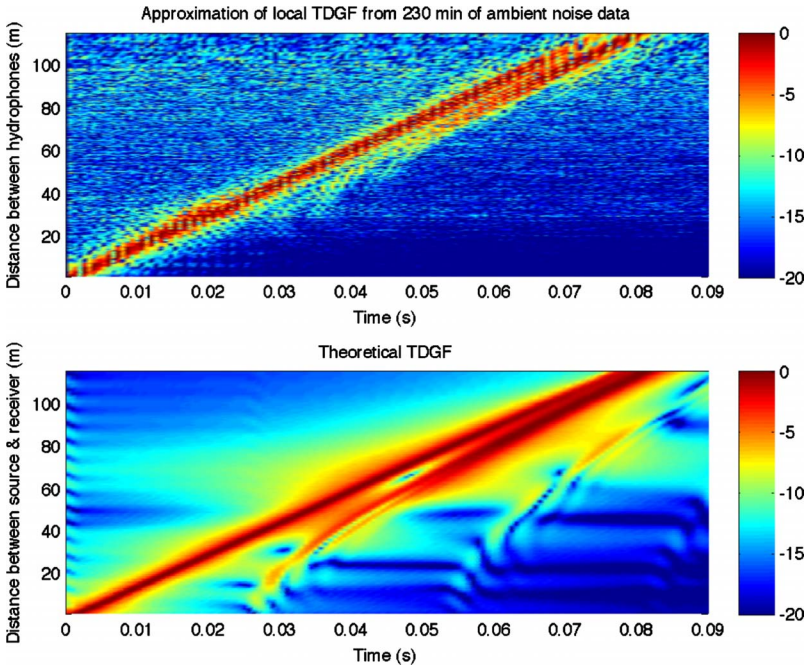


Fig. 1. (Color online) Comparison of the calculated and theoretical local TDGF shown in normalized dB. (a) Time averaged noise correlation function result: *y* axis is the distance between hydrophones, *x* axis is the correlation time. The location of the peaks show the propagation time of sound between the two hydrophones for the direct and surface reflected paths. (b) Spectral simulation of TDGF assuming a source and receiver at the given distance (along *y* axis) separation. Peaks show time of arrival (along *x* axis) of sound from theoretical source to receiver. Environment used for simulated Green's function is based on previous active acoustic inversions and measurements.

recording as possible. In order to minimize interference from pumps and coastal noise the data was filtered to the 250–750 Hz frequency band. Each 5 min segment of data was individually processed and then averaged until the strength of the strongest arrival found in the NCF compared with the residual fluctuations of the correlation function began to deteriorate. The cleanest result appeared to be a 3 h, 50 min window in the early morning hours. After this point there appears to be a change in the environment or the ambient noise such that averaging additional time began degrading the desired signal.

Since the environment was fairly homogeneous throughout the time of the recordings, different pairs of hydrophones with the same separation from each other give essentially the same response. The Green's function approximations for distinct pairs of hydrophones were then sorted by the distance between the pair of hydrophones. When stacked in order of increasing separation distance between hydrophones the NCF exhibits a time of arrival structure consistent with what would be expected from the theoretically calculated (using wavenumber integration techniques[11]) Green's function. Figure 1 shows the processed data stacked by separation distance between pairs of hydrophones and a simulation of the Green's function as though there were sources and receivers at the same positions along the array. The environment for the simulation is taken from the geoacoustic inversion of the same environment.[10] This comparison of the processed data with the simulated TDGF demonstrates that the processed data give an accurate time of arrival sequence for a signal traveling the given distance.

Clearly visible are multiple returns (most clear at greater distances) in the NCF of the data, which match the expected time of arrival of different ray paths between a source and receiver from the simulated Green's function. To verify this, a Monte Carlo noise simulation was made of a two dimensional noisy ocean environment model of the region around the hydrophone array. Discrete sound events were simulated randomly in time and range (up to 2 km)

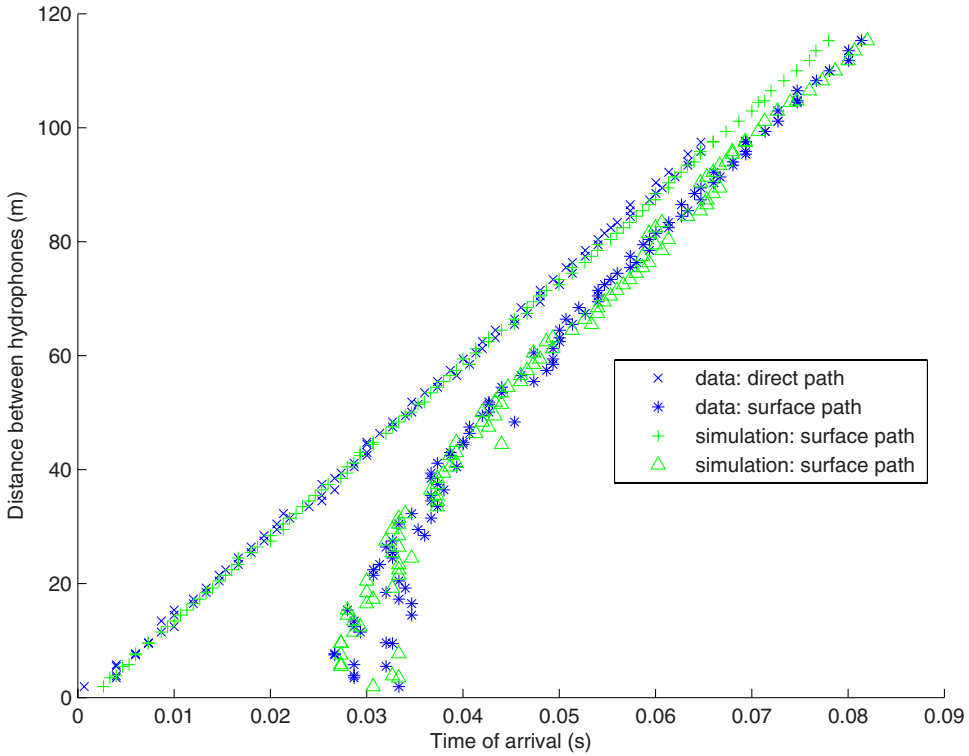


Fig. 2. (Color online) Comparison of peak time of arrivals (x axis) for direct and surface reflection paths for both the NCF processed data and the Monte Carlo noise simulation for increasing distance between hydrophone pairs (y axis). At smaller distances (less than around 60 m) the surface reflection does not tend to give a strong enough return to be accurately extracted. At greater distances the two paths approach each other enough that the stronger path can dominate. There are still two paths but at distances greater than approximately 100 m the surface path begins to dominate the data TDGF and it becomes difficult to isolate the peak arrival of the direct path arrival as it is overshadowed by the surface reflection return.

and located at depths within 2–3 m of the surface and bottom and theoretically propagated along the array. The noise was modeled in this way for computational efficiency and to better simulate the actual noise field which was primarily fish noise near the bottom and physical interactions along the surface. The resulting simulated time series was then NCF processed in exactly as the real data. Figure 2 shows the time of arrival of the peak strength for the direct and surface reflected ray paths for both the NCF processed data and the noise simulation. Both accurately extract the time of arrival for the direct and surface reflection paths, although the surface reflection path is most visible for shallower reflection angles (i.e., greater separation distance between hydrophones). At distances greater than 100 m, the surface reflection path begins to dominate over the direct path. At distances less than approximately 40 m, the strength of the surface reflection path return dissipates and a distinct second arrival becomes undetectable.

When isolated and plotted by distance the strengths of the surface path arrival by the theoretical reflection angle along the ocean bottom the resulting curve begins to approximate a curve for reflection coefficient. This is because of the greater attenuation of steeply reflected ray paths from sources far from the array.[12] Due to clipping and smoothing in the NCF processing though, it is only possible to isolate the relative strengths of the reflection at different angles of incidence and not the actual reflection coefficient. The strength of the surface reflection return was normalized by the distance the sound had traveled, and then normalized from zero to one.

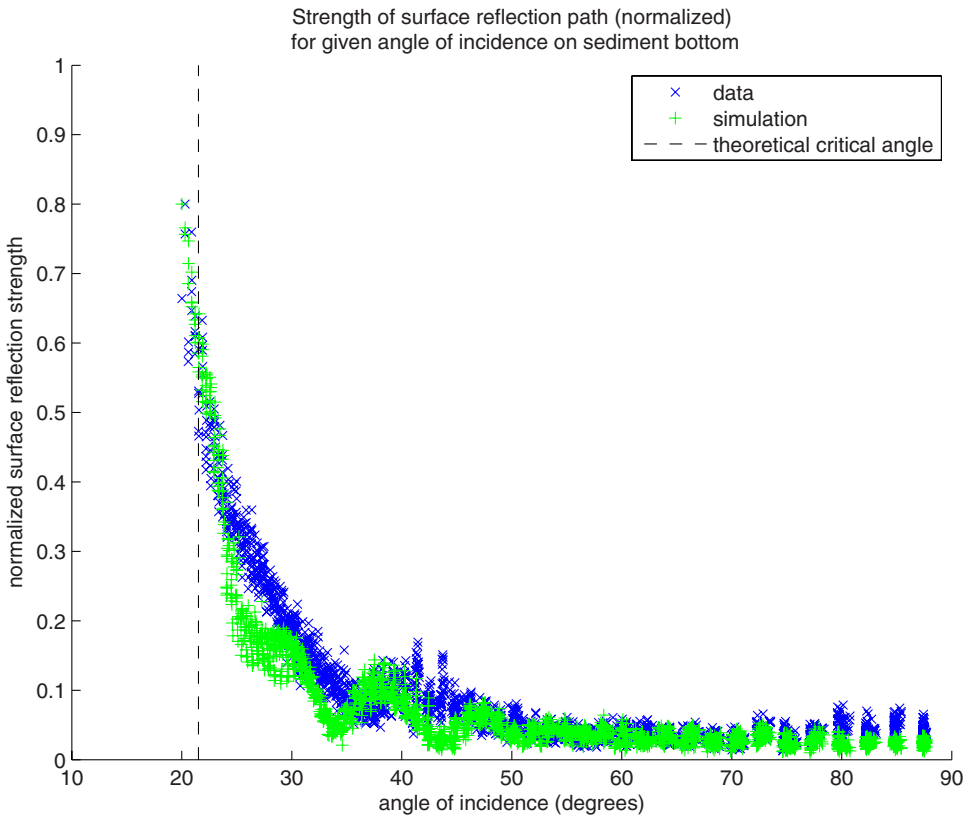


Fig. 3. (Color online) Strength of the surface reflection ray path return by theoretical angle of incidence on sediment floor (x axis) for both the processed data (x's) and the noise simulation (+'s). The strength of the return was normalized by the surface reflection pathlength (to negate 1/R spherical spreading term) and normalized from zero to one. Reflection strength curve indicates a critical angle of just over 20° which agrees with the theoretical critical angle (dotted black line) given the measured environment.

Figure 3 shows the normalized reflection strength for the NCF processed data and the Monte Carlo simulation. The resulting curve appears unfinished and is limited by the length of the array, but it points to a critical angle of incidence of around 20°.

This agrees with the expected critical angle from the previously calculated active geoacoustic inversions[10] of $\theta_c = \cos^{-1} c_{\text{water}}/c_{\text{sediment}} = \cos^{-1} 1495\ \text{m/s}/1607 \pm 45\ \text{m/s} \approx 21.5 \pm 4°$. This value is shown on the figure with a dotted black line.

4. Conclusion

Using only ambient noise on a horizontal array, a representation of the local TDGF was accurately extracted. By tracking the strength of the noise correlation function over time an optimum time to correlate can be chosen for a given noise environment. The resulting time series is shown to accurately approximate the local TDGF by comparison with simulations of the same environment. Analysis of the time series allowed us to accurately resolve environmental detail—in this case the critical angle for the ocean-sediment interface.

Acknowledgment

This work was supported by the Office of Navel Research under Award No. N00014-07-1-0223.

References and links

[1]R. L. Weaver and O. I. Lobkis, "Elastic wave thermal fluctuatons, ultra-sonic waveforms by correlation of thermal phonon," J. Acoust. Soc. Am. **113**, 2611–2621 (2003).

[2]M. Campillo and A. Paul, "Long-range correlations in diffuse seismic coda," Science **299**, 547–549 (2003).
[3]K. G. Sabra, P. Gersoft, P. Roux, W. A. Kuperman, and M. C. Fehler, "Surface wave tomography from microseism in southern California," Geophys. Res. Lett. **32**, L14311 (2005b).
[4]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**, 1995–2003 (2004).
[5]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).
[6]K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[7]P. Roux, K. G. Sabra, and W. A. Kuperman, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–94 (2005).
[8]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**, 3524–3531 (2005).
[9]G. L. D'Spain, L. Berger, W. S. Hodgkiss, W. A. Kuperman, and W. K. Melville, "Observations of the 1 hz–10 khz acoustic ambient noise field near the surf zone in the 1995 adaptive beach monitoring experiment," MPL Report No. TM=454, Marine Phys. Lab., Scripps Inst. of Oceanogr., San Diego, CA (1996).
[10]K. M. McArthur, "Geoacoustic inversions in a very shallow water environment," Ph.D. thesis, AOS Dept., University of California, San Diego (2002).
[11]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, AIP series in modern acoustics and signal processing (Springer-Verlag, Berlin, 2000).
[12]C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**, 1377–1389 (2002).

Effect of instrument timbre on melodic contour identification by cochlear implant users

John J. Galvin III, Qian-Jie Fu, and Sandra Oba
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057
jgalvin@hei.org, qfu@hei.org, soba@hei.org

Abstract: Because of interactions between the signal processing and acoustic input, cochlear implant (CI) users' melodic pitch perception may be influenced by instrument timbre. In the present study, CI listeners' melodic contour identification was measured for six instruments (organ, glockenspiel, trumpet, clarinet, violin, and piano). While performance was generally best with the organ and poorest with the piano, individual CI subjects exhibited different patterns of results across instruments. CI subjects with the most music experience were less susceptible to timbre effects, suggesting that music training may help less experienced CI users extract melodic pitch from a variety of sound sources.

PACS numbers: 43.66.Ts, 43.75.Cd, 43.66.Jh, 43.66.Hg [DD]
Date Received: May 2, 2008 **Date Accepted:** June 21, 2008

1. Introduction

Cochlear implant (CI) users generally have great difficulty with music perception and appreciation. The limited spectral and temporal resolution of the implant currently does not support complex pitch perception (e.g., Smith *et al.*, 2002), making melody recognition and timbre perception difficult. Timbre has been defined as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" (ANSI, 1960). While timbre may refer to sound qualities other than pitch and loudness, timbre has been shown to influence pitch perception in normal-hearing (NH) listeners (e.g., Poulin *et al.*, 2004; Wolpert, 1990). Perhaps the strongest attributes of timbre are the spectral and temporal envelope, which can vary across musical instruments and even within an instrument, depending on the pitch, loudness, and expression of a musical note. Perception of instrument timbre is important for segregating and streaming sound sources in a multi-instrument context, allowing different melodic components to be heard individually or in combination.

Previous timbre perception studies with CI users have largely focused on instrument identification and/or sound quality appraisals. Most studies have found CI users' instrument identification to be much poorer than that of NH listeners (e.g., Gfeller *et al.*, 2002a; McDermott, 2004; Nimmons *et al.*, 2008). In these studies, CI users typically were better able to identify percussive and/or "pitched-percussive" instruments (i.e., a strong dependence on temporal envelope cues), and often could identify different orchestral instrument families. A recent study by Looi *et al.* (2008) found that CI users' instrument identification was significantly poorer when stimuli were presented in the presence of different musical backgrounds than when presented in isolation. However, Pressnitzer *et al.* (2005) found that CI listeners were as sensitive as NH listeners to changes in attack time and spectral gravity, two components of instrument timbre (McAdams *et al.*, 1995), suggesting that CI users might be trained to properly label instrument sound sources. Indeed, Gfeller *et al.* (2002b) has shown improved instrument identification after training CI listeners with auditory and visual examples of instruments.

While these previous studies have provided information regarding instrument identification, it is unclear how timbre cues may interact with CI users' melodic pitch perception. Instrument timbre may interact with CI signal processing (e.g., dynamics processing, acoustic

Table 1. CI subject demographics.

Subject	Gender	Age	CI experience (years)	Device/ Strategy
S1	M	50	1	Freedom/ ACE
S2	F	63	4	N24/ ACE
S3	M	56	17	N22/ SPEAK
S4	M	49	15	N22/ SPEAK
S5	M	77	11	N22/ SPEAK
S6	M	74	8	CII/ Fidelity 120
S7	M	60	16	N22/ SPEAK
S8	M	42	8	N24/ ACE

frequency allocation, etc.) and/or individual CI users' spectral/temporal processing capabilities (e.g., electrode pitch range/resolution, modulation sensitivity), resulting in markedly different stimulation patterns for the same input melody. Once transmitted, these patterns may also be perceived differently by individual CI users, depending on the cues used to extract musical pitch. In most previous CI melody recognition studies, familiar melodies have been played by a single instrument (e.g., Kong *et al.*, 2004; Nimmons *et al.*, 2008), simple sine waves (e.g., Galvin *et al.*, 2007), or pulse trains delivered directly to selected electrodes (e.g., Pijl and Schwartz, 1995). As such, it is unclear whether the few melodies that were recognized in these studies would be correctly identified in another instrument context. It is also unclear from these studies how to optimize the acoustic input to support melodic pitch perception while retaining important timbre cues.

In the present study, melodic contour identification (MCI) was measured in CI and NH subjects for six instruments. The MCI paradigm was previously used by Galvin *et al.* (2007) to measure melodic pitch perception. One limitation to that study was the use of a 3-tone complex as the musical "instrument." Because CI users most likely attended to spectral envelope cues to perform the MCI task, changes in the spectral envelope associated with "real" instruments may influence melodic pitch perception. Different temporal properties may also influence MCI performance.

2. Methods

Eight CI patients and eight NH listeners participated in the present experiment. Six of the CI subjects participated in a similar previous experiment (Galvin *et al.*, 2007), and thus were familiar with the MCI task. CI subject demographics are shown in Table 1. Note that subjects S1 and S2 had greater music experience before and after implantation, compared with the other CI subjects. S1 was a guitarist and songwriter before implantation, and continued to play music after implantation; S2 received organ lessons as a child. Both S1 and S2 actively enjoy listening to music with their CI device. The remaining CI subjects exhibited a range of music listening habits before and after implantation; some enjoy listening to music with the CI, while others only occasionally listen to music after implantation. The mean age for NH subjects was 37.5 years (range: 20–52 years). Two of the NH subjects were active musicians, and five of the NH subjects had some musical instruction (e.g., piano, cello, organ, or drum lessons) during

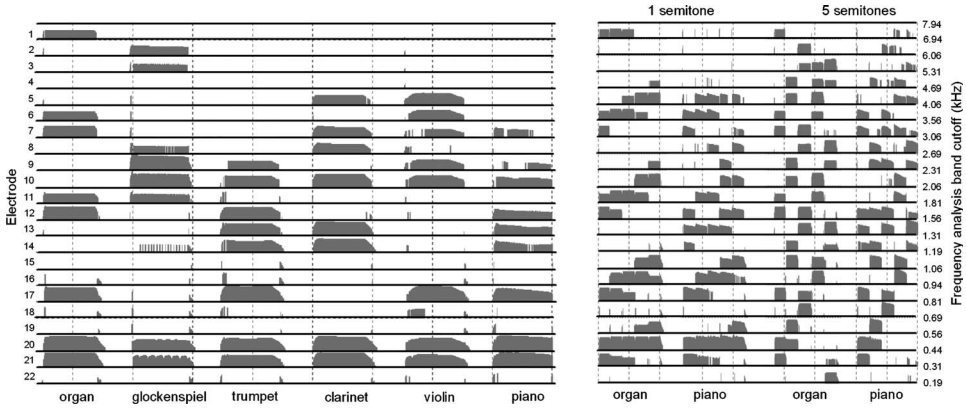


Fig. 1. Left panel: electrodogram for six instruments at A4 (440 Hz). Right panel: electrodogram (for organ and piano) for the Rising melodic contour, with 1-semitone or 5-semitone intonation. The left axes show electrode number and the right axes show the cutoff frequencies of the analysis bands.

childhood. NH subjects were largely recruited from the staff at the House Ear Institute and House Ear Clinic. None of the NH subjects had any previous experience with MCI task. All subjects were paid for their participation, and all provided informed consent before participating in the experiment.

Stimuli consisted of nine melodic contours (Rising, Flat, Falling, Flat-Rising, Falling-Rising, Rising-Flat, Falling-Flat, Rising-Falling, Flat-Falling), similar to those used in Galvin *et al.* (2007). Each contour consisted of five notes played in sequence. Melodic contours were generated in relation to a "root note," i.e., the lowest note in the contour (A4, or 440 Hz). All notes in the contours were generated according to: $f=2^{x/12}f_{\mathrm{ref}}$, where f is the frequency of the target note, x is the number of semitones relative to the root note, and f_{ref} is the frequency of the root note. To test the effects of intonation on MCI performance, the interval between successive notes in the contour was varied between 1, 2, 3, 4, and 5 semitones. Given the different contours and intonations, the fundamental frequency (F0) range was 440–1397 Hz. Each note was 250 ms in duration, and the interval between notes was 50 ms.

Melodic contours were generated for each instrument using MIDI sampling and synthesis. The six instruments were: organ, glockenspiel, trumpet, clarinet, violin, and piano. Instruments were selected to comprise a range of spectral and temporal properties, as well as to represent somewhat familiar musical instruments. For example, the piano and glockenspiel (similar to a xylophone) were chosen to represent pitched percussion instruments having similar temporal characteristics but very different spectra. Similarly, the organ and clarinet have somewhat similar spectra, but somewhat different temporal properties. While not explicitly controlled or balanced, instruments were chosen to represent orchestral families of instruments (i.e., brass, woodwind, string, pitched percussion). The left panel of Fig. 1 shows electrodograms for each of the six instruments for the A4 note. The electrodogram was generated using custom software that simulated the Advanced Combination Encoder (or ACE; Vandali *et al.*, 2000) strategy typically implemented in Nucleus-24 and Freedom implant devices; the default parameter settings were simulated (i.e., Frequency Allocation Table 6, 900 Hz/channel, 8 maxima, etc.). As seen in Fig. 1, the instruments produced markedly different stimulation patterns (i.e., spacing of electrodes, onset/offset). Note that the stimulation patterns could be quite different for alternative instrument samples, processor mappings (e.g., change in frequency allocation), or processor strategies. For example, with Continuously Interleaved Sampling (or CIS; Wilson *et al.*, 1991), all electrodes are stimulated, as compared to a "peak-picking" strategy like ACE, in which a subset of electrodes is stimulated in each frame.

The right panel of Fig. 1 shows the Rising contour for the organ and piano, for 1- and 5-semitone intonation. For both intonations, the stimulation pattern is more dense for the piano

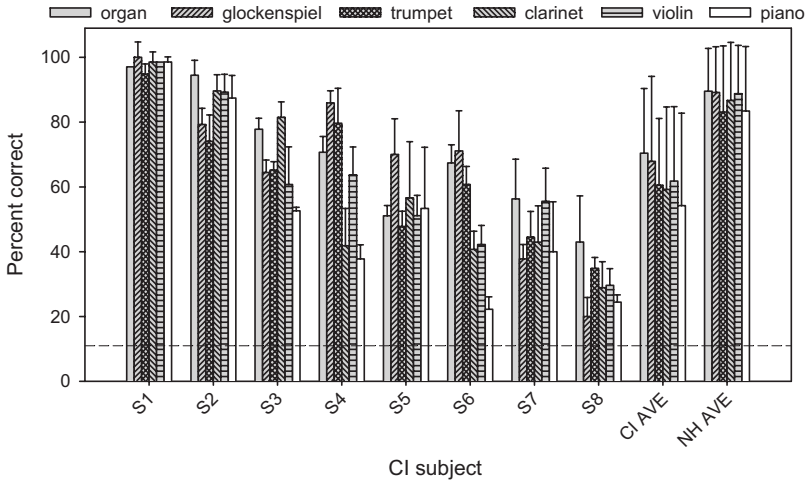


Fig. 2. MCI performance for individual CI subjects, for different instruments; mean CI and NH subject performance is shown at right. The dashed line shows chance performance level. The error bars show one standard deviation of the mean.

than for the organ. With 1-semitone intonation, the pattern changes over relatively few electrodes; with 5-semitone intonation, the pattern changes over nearly all electrodes. However, considering potential interactions between speech processor settings, instrument timbre, intonation, and contour direction, the stimulation pattern may contain conflicting cues, depending on the cues used to extract pitch information. If CI users mainly attend to low frequency components, timbre may not strongly affect MCI performance. If CI users attend to high frequency components, the acoustic input pattern may be distorted. As shown in Fig. 1, the highest frequency components are not represented throughout the Rising contour, especially for the 5-semitone intonation. In a peak-picking strategy such as ACE, lower frequency components would be selected to maintain the targeted 8 maxima per frame. If listeners attend mainly to the high frequency components, or to the "substitute" low frequency components, the input Rising contour might be perceived as Rising-falling, or Rising-flat.

For both NH and CI subjects, stimuli were presented acoustically via a single loudspeaker (Tannoy Reveal) at 70 dBA. Testing was conducted in a sound-treated booth (IAC) with subjects directly facing the loudspeaker. CI subjects were tested using their clinically assigned speech processors. Note that subjects S6 and S8 wore a hearing aid in the contralateral ear in everyday listening; the hearing aid was removed during testing (although the ear was not plugged). CI subjects were asked to set their sensitivity and volume settings as they would for comfortably loud speech; once set, subjects were asked not to change these settings. MCI performance was measured independently for each instrument; each test block per instrument was repeated a minimum of three times, and the test block order was randomized within and across subjects. During each test block, a stimulus was randomly selected (without replacement) from the 45 stimuli (nine contours $*$ five intonations). Subjects were asked to click on one of the nine response choices shown onscreen, after which the next stimulus was presented. Subjects were instructed to guess if they were unsure, but cautioned against choosing the same response for every stimulus. Subjects were allowed to repeat each stimulus up to three times; however, no preview or trial-by-trial feedback was provided.

3. Results

Figure 2 shows individual CI subjects' mean MCI performance (across intonation conditions), as well as overall mean CI and NH performance with the different instruments. Mean CI performance across subjects and instruments was 62.3% correct (range: 30.1% to 97.9% correct).

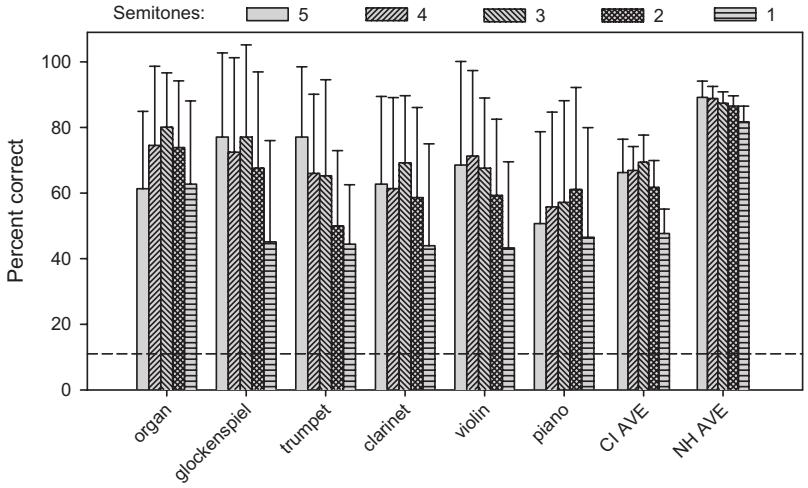


Fig. 3. Mean MCI performance (across CI subjects) for the different intonations as a function of instrument; mean CI and NH performance (across subjects and instruments) is shown at right. The dashed line shows chance performance level. The error bars show one standard deviation of the mean.

Overall mean CI performance was best with the organ (70.4% correct) and poorest with the piano (54.2% correct). A two-way repeated measures (RM) analysis of variance (ANOVA) showed that the instrument effect failed to achieve significance [$F(5,140)=2.24$, $p=0.07$], while intonation had a significant effect [$F(5,140)=14.79$, $p<0.001$]; there was a significant interaction between instrument and intonation [$F(20,140)=2.94$, $p<0.001$]. Because the pattern of results differed greatly across subjects, a one-way ANOVA was performed on individual subjects' data. Results showed that instrument timbre significantly affected MCI performance for subjects S2, S3, S4, S5, and S6 ($p<0.05$). Mean NH performance was 86.8% correct (range: 61.5%–100.0% correct). Similar to CI subjects, overall mean NH performance was best with the organ (89.5% correct) and poorest with the piano (83.4% correct). The pattern of results was more homogenous across NH subjects than across CI subjects. A two-way RM ANOVA showed that both instrument [$F(5,140)=3.74$, $p=0.008$] and intonation significantly affected performance [$F(5,140)=4.06$, $p=0.01$]; no significant interaction was observed [$F(20,140)=1.36$, $p=0.15$]. *Post hoc* Bonferroni pairwise comparisons showed significant differences only between the clarinet and piano with 5-semitone intonation, and between the trumpet and organ with 1-semitone intonation ($p<0.05$).

Figure 3 shows mean MCI performance (across subjects) for the different intonations as a function of instrument; mean CI and NH subject performance (across instruments) is shown at right. Mean CI performance ranged from 66.9% correct (4 semitones) to 47.7% correct (1 semitone). *Post hoc* Bonferroni pairwise comparisons from the above two-way RM ANOVA showed that performance with 1-semitone intonation was significantly poorer than with the other intonations ($p<0.05$). There was no significant difference between the 1- and 5-semitone intonations for the organ and piano; for the organ, performance with the 3-semitone intonation was significantly better than with the 1- and 5-semitone intonations ($p<0.05$). Mean NH performance ranged from 89.2% correct (5 semitones) to 81.7% correct (1 semitone). *Post hoc* Bonferroni pairwise comparisons from the above two-way RM ANOVA showed that performance with the 1-semitone intonation was significantly poorer than the 5- and 4-semitone intonation conditions ($p<0.05$).

4. Discussion

The results of the present study show that instrument timbre influenced most CI subjects' MCI performance. Mean CI performance was much poorer than that of NH listeners, consistent with

the previous MCI study (Galvin *et al.*, 2007) and many other CI music perception studies. With the A4 root note, mean CI performance with the different instruments (62.3% correct) was poorer than that with the 3-tone complex (81.7% correct) used in Galvin *et al.* (2007). Similarly, mean NH performance with the different instruments (86.8% correct) was poorer than that with the 3-tone complex (97.7% correct). For the stimuli used in these studies, it may have been more difficult to extract musical pitch from the relatively complex "real" instruments than from the simple 3-tone complex.

The pattern of results for the organ and piano suggest that some CI subjects' melodic pitch perception may have been influenced by high frequency components, which may have been truncated under certain test conditions by the CI signal processing (see Fig. 1). For these instruments, widely spaced notes in the contour proved as difficult as did closely spaced notes. This result differs from that in Galvin *et al.* (2007) with the 3-tone complex, where MCI performance generally improved as the intonation was increased. Changes in CI signal processing may help to remove some of the confusing frequency components and improve melodic pitch tracking. Other subjects (S1 and S2) may have relied more strongly on lower frequency components, suggesting that individual CI listeners may use different cues to extract melodic pitch. The best performers may be able to extract the "gestalt" of melodic pitch, despite potentially confusing frequency components.

While previous CI studies have found that percussive and "pitched-percussive" instruments were easier to identify, the piano (with a relatively sharp attack, compared to the other instruments) provided the poorest MCI performance in the present study. Note that the glockenspiel (with an even sharper attack) provided much better performance. The stimulation pattern differed greatly between these instrument samples (see Fig. 1), and the relatively complex pattern of the piano may have contributed to the poorer MCI performance. Despite the poorer MCI performance, many CI subjects commented that the piano sounded most "natural" and "musical" among the test instruments. While instrument identification may indicate listeners' ability to label sound sources, and quality appraisals may indicate the relative "pleasantness" of different instruments, these measures may not well predict CI listeners' ability to follow the same melody played on different instruments. While instrument timbre significantly affected most CI subjects' MCI performance, the pattern of results was not consistent across subjects. For example, the glockenspiel provided the worst performance for subject S8 and the best performance for S4. Unfortunately, the present data do not clearly reveal any acoustic properties (e.g., attack, density of spectral envelope, etc.) that might influence CI users' melodic pitch perception.

Many of the CI subjects in the present study participated in the previous Galvin *et al.* (2007) study. Mean CI performance was nearly 20 percentage points better with the 3-tone complex in the previous study, compared to performance with multiple instruments in the present study. This suggests that some "simplification" of the complex stimulation patterns produced by real instruments might improve CI users' melodic pitch perception. Some previous studies (e.g., Guerts and Wouters, 2004; Katsuri and Loizou, 2007) have offered alternative frequency allocations to improve reception of F0 cues. While these approaches may improve F0 perception, they typically exclude higher frequency components that are important for speech understanding and timbre perception. In the present study, some CI users were better able to extract melodic pitch cues from a variety of sources. For other CI subjects, high frequency and/or inharmonic cues may have interfered with melodic pitch perception. For these CI users, alternative signal processing approaches might improve pitch tracking. By simplifying the signal, these listeners might better attend to F0 cues, after which additional high frequency components could be included in the signal processing, thereby restoring some timbral differences between instruments.

One strong predictor of MCI performance seemed to be the degree of music experience, before and after implantation. The most musically experienced CI subjects (S1 and S2) were capable of high levels of performance (even higher than some NH subjects), and were largely unaffected by instrument timbre in the MCI task. Musical experience has shown to influence NH listeners' melody identification with alternative instrumentation (e.g., Poulin

et al., 2004; Wolpert, 1990). While many CI users may lack music experience, training may help to improve music perception. In Galvin *et al.* (2007), MCI performance with the 3-tone complex improved by as much as 45 percentage points after training with novel melodic contours; the improved MCI performance generalized to improved recognition of familiar melodies. Gfeller *et al.* (2002b) showed improved instrument identification (i.e., timbre perception) after moderate amounts of training. While such training efforts may improve CI users' music perception, it is unclear whether more intensive training (e.g., training with a musical instrument) is needed to allow CI users' to extract musical pitch from spectrally degraded representations of different sound sources.

Acknowledgments

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

References and links

ANSI (**1960**). American Standard Acoustical Terminology. American National Standards Institute, New York.
Galvin, J., Fu, Q.-J., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear Hear. **28**(3), 302–319.
Gfeller, K., Witt, S., Woodworth, G., Mehr, M., and Knutson, J. (**2002a**). "Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal," Ann. Otol. Rhinol. Laryngol. **111**(4), 349–356.
Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., and Ringgenberg, S. (**2002b**). "Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients," J. Am. Acad. Audiol **13**(3), 132–145.
Geurts, L., and Wouters, J. (**2004**). "Better place coding of the fundamental frequency in cochlear implants," J. Acoust. Soc. Am. **115**(2), 844–852.
Kasturi, K., and Loizou, P. (**2007**). "Effect of filter spacing on melody recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **122**, EL29.
Kong, Y.-Y., Cruz, R., Jones, J., and Zeng, F.-G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
Looi, V., McDermott, H., McKay, C., and Hickson, L. (**2008**). "Music perception of cochlear implant users compared with that of hearing aid users," Ear Hear. (in press).
McAdams, S., Winsberg, S., Donnadieu, D., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.
McDermott, H. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**(2), 49–82.
Nimmons, G., Kang, R., Drennan, W., Longnion, J., Ruffin, C., Worman, T., Yueh, B., and Rubinstein, J. (**2008**). "Clinical assessment of music perception in cochlear implant listeners," Otol. Neurotol. **29**, 149–155.
Pijl, S., and Schwarz, D. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J. Acoust. Soc. Am. **98**, 886–895.
Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., and McAdams, S. (**2004**). "Effects of a change in instrumentation on the recognition of musical materials," Music Percept. **22**, 239–263.
Pressnitzer, D., Bestel, J., and Frayssec, B. (**2005**). "Music to electric ears: Pitch and timbre perception by cochlear implant patients," Ann. N.Y. Acad. Sci. **1060**, 343–345.
Smith, Z., Delgutte, B., and Oxenham, A. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
Vandali, A., Whitford, L., Plant, K., Clark, G. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.
Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., and Rabinowitz, W. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Wolpert, R. (**1990**). "Recognition of melody, harmonic accompaniment, and instrumentations: Musicians vs. nonmusicians," Music Percept. **8**, 95–106.

Ultrasonic measurement of condensate film thickness

Jeramy T. Kimball
Department of Aeronautics and Astronautics, University of Washington, Box 352400, Seattle, Washington 98105
jeramyk@aa.washington.edu

Michael R. Bailey
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105
bailey@apl.washington.edu

James C. Hermanson
Department of Aeronautics and Astronautics, University of Washington, Box 352400, Seattle, Washington 98105
jherm@aa.washington.edu

Abstract: The current work describes a modified time-of-flight ultrasound signal processing technique applied to the study of a distal liquid layer with a free surface. The technique simulates multiple reflections analytically and determines the film thickness by comparison to the measured pulse echo signal. The technique is applied with 20 MHz transducers to an *n*-pentane film condensing on a copper plate. The technique proved capable of measuring liquid thickness from approximately 8 μm, 1/6 the acoustic wavelength in pentane, to greater than 5 mm. Near the lower thickness limit, echoes from the liquid/vapor interface overlap each other and the significantly larger echoes from the metal/liquid interface.

PACS numbers: 43.35.Yb [JM]
Date Received: March 24, 2008 **Date Accepted:** July 17, 2008

1. Introduction

Phase change involving liquid films is important to a variety of terrestrial and space applications. These include two-phase thermal control systems for environment control and life support, including humidity control, air revitalization, water purification, and heat rejection, as well as thermal management of spacecraft components, all of which depend upon condensation and evaporation.[1–4] Many more applications are dependent specifically upon evaporation, including fuel vaporization,[5] flame propagation over liquid fuels,[6] hardening of plastics, and numerous coating situations.[7] Phase change processes in space are a concern because buoyancy effects are important in terrestrial applications but are nearly absent in phase change in microgravity. The physical phenomena, including convection, operative in condensing and evaporating films have been studied extensively under a variety of conditions (see, e.g., Som *et al.*,[8] and references therein) but accurate, dynamic, nonintrusive measurements of film thickness are lacking.

Previous work involving ultrasound to investigate liquid films undergoing phase change separated the signal processing into two regimes.[9,10] Ultrasound signals corresponding to films where individual echoes overlapped were studied in the frequency domain, while thicker films were studied in time domain. As implemented, this technique was limited to studying films greater than 50 μm where standing waves in the film produce strong signals at the transducer.

The current work applies normal-incidence ultrasonic probing with an array of transducers to the study of the behavior of liquid films undergoing phase change. Described is a modified time-of-flight signal processing technique that is able to measure dynamically the thickness of liquid films on a metal substrate over a broad range (from less than 1/6 of the

acoustic wavelength, where significant ringing is present, to above several millimeters) with a precision of approximately ±1 μm and accuracy better than ±10% in trial experiments.

The technique uses scripts and functions in MATLAB to create a simulated signal matched using a least-squares technique to each acquired signal. Matching is accomplished by varying the film thickness in the simulation, which varies the time delay of each echo in the reverberant three-medium problem. The simulated signal is created by first calculating the amplitude of each wave based upon its ray path in a characteristics diagram and the material properties. The wave form of each wave is the same as, or an inverted form of, the wave form measured in normal reflection from the metal/vapor interface. The individual waves with appropriate time delays are then summed to create the simulated signal. The technique has the benefit of being able to process films from as little as 8 μm to more than 5 mm thick, even if the individual waves overlap in time.

2. Principles and analytical prediction

In the current experiment, the ultrasound transducer is separated from the fluid layer (the region of interest) by a copper plate approximately 3 mm thick on which a condensing *n*-pentane film resides. An initial pulse is sent from the transducer, and when it encounters the change of acoustic impedance at either the metal/liquid interface or the liquid/vapor interface, one wave is transmitted and another wave is reflected.

The amplitudes of the reflected and transmitted waves relative to the initial wave are predicted by Eqs. (1) and (2), where ρ and c represent the material density and sound speed, respectively, and subscripts 1 and 2 indicate material properties before and after the interface, respectively,

$$R = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1}, \tag{1}$$

$$T = \frac{2\rho_2 c_2}{\rho_2 c_2 + \rho_1 c_1}. \tag{2}$$

A simulated signal is created as part of the postprocessing of each ultrasound signal captured during the experiment. The time between successive echoes is varied in the simulated signal in order to minimize the sum of the squares of the residuals between the simulated signal and the acquired signal. Knowing the time between successive echoes in the simulated signal thus yields the film thickness. The details on the signal simulation method and the film thickness measurement are presented in the following section.

The simulated signal is created by summing many individual waves corresponding to distinct echoes received at the transducer from the metal/liquid and liquid/vapor interfaces. An individual wave received at the transducer is identified as wave(n,p), where n is the number of round-trip passes through the liquid layer and p is the number of round-trip passes through the metal layer, and has a corresponding amplitude coefficient, amplitude(n,p), and a time delay based on n times twice the fluid thickness plus p times twice the metal thickness. Several unique paths have the same delay and sum to form a single wave, wave(n,p), as seen in a propagation diagram (Fig. 1). All signals from the paths making up one wave have the same delay but have different magnitudes and may be amplitude inverted from one another. The number of different paths, N, as a function of n and p used in calculating the relative amplitude(n,p) corresponding to wave(n,p), is given by

$$N = \frac{(n+p-1)!}{(p-1)!n!}. \tag{3}$$

Figure 1 shows the paths for the wave where $n=2$, $p=3$, wave(*2,3*). In this case there are six paths, as detailed in Table 1. The vector corresponding to the first path, [2 0 0] (see Table 1), would indicate that the path includes two passes through the liquid layer in the first pass

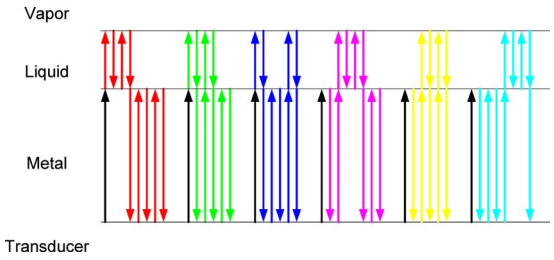


Fig. 1. (Color online) The six propagation paths associated with wave(*2,3*) resulting from a single excitation pulse. Corresponding path details can be found in Table I.

through the metal layer and zero passes through the next two passes through the metal layer. This path is represented as the red line in Fig. 1.

Given the path vector, the number of reflections and transmissions for each interface can be counted. Then, when coupled with Eqs. (1) and (2) and the known material properties, the amplitude for each path relative to the initial pulse amplitude can be calculated. The relative amplitude of the wave is the sum of the amplitudes of each path.

The following technique is used to account for the attenuation in the metal substrate material and to set the absolute amplitude and the wave form of the waves. The reflections are measured without the liquid film present. This is base$(n=0,p)$. The base wave form consists of a few rapidly decaying cycles at 20 MHz, the result of a very short electrical excitation of the resonant PZT source. After a correction to the base amplitude for the slight difference in R [Eq. (1)] with and without the film present, wave$(0,p)$ is set equal to base$(0,p)$ and more generally, wave(n,p) is defined by base$(0,p)\times$amplitude(n,p)/amplitude$(0,p)$. The frequency dependence of the attenuation in the metal is implicitly taken into account in that the measurement sets the amplitude and the wave form after propagation through the metal. The unknown reflection coefficient at the transducer/brass interface is similarly accounted for. The attenuation in the liquid layer, which is much smaller than the loss in the metal, is neglected.

The final step is to factor in the arrival times of the waves and sum them into one simulated signal. In other words, the wave$(n,p;t)$, where previously the time, t, has been suppressed, is delayed to wave$(n,p;t\text{-}n\Delta t_{\text{liquid}}-(p\text{-}1)\Delta t_{\text{metal}})$, where Δt_{metal} is the time between successive waves from the metal/liquid interface (i.e., twice the known metal thickness divided by sound speed of the metal) and Δt_{liquid} is the time between successive echoes from the liquid/vapor interface (i.e., twice the unknown thickness of the liquid divided by sound speed of the liquid). This delay sets the first reflection to arrive at time zero. Then all the waves are summed over p and n.

The free parameter Δt_{liquid} is directly proportional to the film thickness and matching the summed simulated signal to a signal acquired during the experiment by varying Δt_{liquid} allows for measurement of the film thickness.

Table 1. Path summary for example wave*(2,3)*. Colors indicated correspond to Fig. 1.

Path	Summary vector	Color in figure	Relative amplitude
1	[2 0 0]	Red	0.0378
2	[1 1 0]	Green	–0.0034
3	[1 0 1]	Blue	–0.0034
4	[0 2 0]	Magenta	0.0378
5	[0 1 1]	Yellow	–0.0034
6	[0 0 2]	Cyan	0.0378

3. Hardware

The ultrasound system includes a digital oscilloscope, a preamplifier, a multiplexing thickness gauge, a multiplexer, four transducers, and a desktop computer. The transducers have a 20 MHz center frequency and a 3-mm-diam element. The oscilloscope acquires the signals at 4 Gsamples/s. The thickness gauge acts as a voltage spike source and a transmit/receive switch. The transducers are excited with a single high voltage spike. Twenty megahertz sources, the highest frequency readily available, were chosen and a single electrical impulse about 20 ns in duration was used in order to maintain the shortest acoustic pulse possible, thus maximizing the chance of resolving two pulses. With the technique described here this was less important. The thickness of the copper plate was minimized in order to minimize attenuation, which allowed for the high, 20 MHz, acoustic frequency.

The gauge is set to switch between the multiplexer channels 20 times/s; therefore, a wave form from each of the four transducers is captured at 5 Hz. After switching channels, the gauge sends a pulse to the transducer. This pulse triggers the oscilloscope, which acquires the pulse and echo signals from the preamp. Once triggered, the oscilloscope digitizes and sends the wave form to the computer via a USB connection. The computer stores the wave form and instructs the scope to wait for the next trigger.

The experiment occurs in a pressure-controlled cylindrical test chamber approximately 12.5 cm in diameter by 15 cm tall. The base of the cylinder is the copper test block, the top is a glass viewing window, and the sides are Teflon™. *n*-pentane is supplied from a heated vapor generator, and the pressure is controlled using an electronic backpressure regulator. The more complicated problem of a downward-facing, unstable film was studied by Som *et al.*[8] in a test chamber similar to the current work.

The system was initially tested in simple control experiments where a known volume of fluid was applied to the copper surface and, therefore, the thickness was approximately known for comparison with the ultrasound measurements. For brevity and because it is difficult to obtain very thin films with that technique, those experiments are not reported here. Here we report measurements compared to calculations of film thickness based on known condensation rates.

4. Wave form postprocessing

Once the copper plate has reached the experimental temperature, 100 wave forms are acquired for each transducer with no liquid present on the metal substrate. These are averaged for each channel and are windowed to form the base wave forms discussed earlier. The signals in these wave forms are the echoes from the metal/vapor interface [see Fig. 2(a)]. The length of the acquired signal is approximately five times larger than Δt_{metal}.

During postprocessing, the base signals are shifted in time to set the center of the first metal/vapor echo at time equal to zero. As with the simulated signal, time signifies the time elapsed since the first echo from the metal/liquid interface. The temperature of the block does not change significantly (less than 0.5 °C) so that these offset times do not change significantly during the experiment.

The waves from the metal/liquid interface are removed from the experiment wave form by subtracting out the base signal, Fig. 2(a), from the measurement with a film, Fig. 2(b), to produce the signal in Fig. 2(c). Note the *y*-axis changes because the waves from the liquid/vapor interface are much smaller than those from the metal/liquid interface. Next, a least-squares fit between the simulated and measured signal is performed. The sum of the squares of the residuals between the measured and simulated signals is minimized using Δt_{liquid} as a parameter. Figure 2(d) shows a simulated signal matched to an acquired signal. The function finds the best local fit, so some care must be taken when choosing the start point. If the explained variation, i.e., quality of fit or R^2, is unacceptably low the user is prompted to enter a new set of initial parameters or accept the current fit. If the revised parameters fail to achieve an acceptable solution the measurement is rejected.

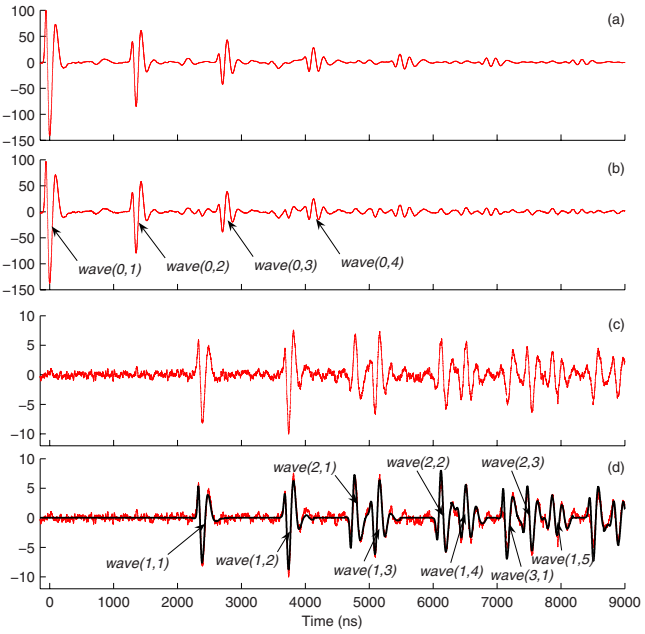


Fig. 2. (Color online) Acquired wave form processing steps: (a) Base wave form, no liquid present. (b) Wave form with 1104.2 μm *n*-pentane film present. (c) Wave form in (a) subtracted from the wave form in (b) to reveal reflections from the pentane film. Before subtraction, the amplitude of (a) is reduced commensurate with the change in impedance caused by adding a film to the metal surface. (d) Comparison of measured and simulated signals created by reflections in the pentane layer.

5. Results

The measured thickness of *n*-pentane films on the copper surface was in excellent agreement with the predicted condensation rate (Fig. 3). With the metal surface facing upward, the pressure of the *n*-pentane vapor in the surrounding test chamber was raised quickly. The corresponding increase in saturation temperature resulted in the metal surface becoming subcooled, causing pentane vapor in the chamber to condense on the surface. This experiment was chosen for two reasons. First, the film is flat and parallel to the copper and therefore is reasonably easy to measure with the ultrasound system. Second, it is possible to predict analytically the film thickness versus time. The heat flux per unit area at the film surface due to condensation is expressed as $q''=\delta_t \rho h_{fg}$; where δ_t is the derivative of thickness with respect to time, ρ is the density, and h_{fg} is the latent heat of fusion. Heat flux per unit area conducted through the film is expressed as $q''=-k(\partial T/\partial x)\approx -k(T_{\text{sub}}/\delta)$; where k is the thermal conductivity of the fluid, T_{sub} is the degree of subcooling, and δ is the film thickness. Equating these expressions and solving for δ yields Eq. (4), the established prediction for film thickness as a function of fluid parameters and time,

$$\delta = \sqrt{\frac{2T_{\text{sub}}kt}{\rho h_{fg}}}. \tag{4}$$

The fluid properties are taken from the NIST Chemistry Webbook.[11] The stated uncertainty in k, ρ, and h_{fg} are 1.0%, 0.2%, and 1.0%, respectively. The uncertainty in T_{sub} is 0.1 °C, yielding a total uncertainty in the thickness calculation of approximately 3.5%.

In Fig. 3, the ultrasonically measured film thickness versus time is compared to the thickness calculated with Eq. (4). The measured values are similar for all four channels and agree with the predicted values to within 10% [Fig. 3(b)] suggesting that the technique is accurate to at least this degree. Deviations appear around 15 s into the test interval when the film is 150 μm thick and continue through to the end of the experiment. These deviations are likely due

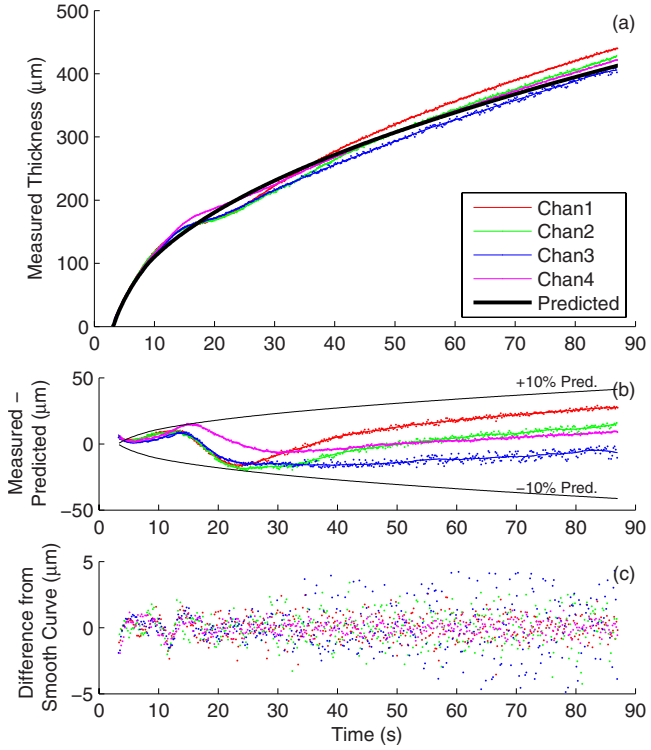


Fig. 3. (Color online) Measured film thickness for an upward facing condensing *n*-pentane film. (a) Measured thickness (points) and corresponding smoothed curves (thin lines) for each of the four transducers. Also plotted is the predicted thickness (thick line). (b) Difference between measured and predicted values from (a) (points) as well as the difference between smooth curve and prediction from (a) (lines). Also plotted are ±10% of predicted thickness. (c) Difference between measured values and corresponding smooth curve for each transducer. The variance of the differences is about 1 μm.

to the copper plate not being perfectly level to the ground resulting in fluid motion within the film. The level was off horizontal by less than 18 arcsec. The limitation in the accuracy appears to have been more a result of the difficulties associated with creating a perfect experimental configuration, specifically the lack of perfectly level test surface, than of the ultrasound technique itself. The measured thicknesses were also compared with smooth curves calculated for each channel using a 50-point Savitzky–Golay filter. That comparison indicated a deviation from the corresponding smoothed curves of less than 5 μm [Fig. 3(c)] with a variance of approximately 1 μm, suggesting that the precision of the measurements is approximately 1 μm.

6. Summary

The current work describes a technique for film thickness measurement that allows for processing ultrasound signals where echoes from the distal layer, the layer of interest, may overlap with echoes from the proximal layer and/or other echoes from the distal layer. The same technique can also handle signals with thicker distal layers where echoes from the distal layer are received after the second echo from the proximal layer. Although shear waves and possible oblique reflections from the film surface were likely present and were neglected, the system proved to be precise to within 1 μm and accurate to within 10%, with significantly greater accuracy likely possible with a more precisely leveled substrate material. The technique benefits from the requirement of carefully controlled temperature to conduct the condensation experiments, and thus, changes in temperature and sound speed are negligible.

In this work, the distal layer is an *n*-pentane liquid layer undergoing phase change. The proximal layer is a 3-mm-thick copper plate. Using the technique described it is possible to match simulated signals to acquired signals for layer thickness ranging from 8 μm to above 5 mm.

Acknowledgments

This research was supported by NASA Cooperative Agreement No. NNC04GA76G under the technical management of E.L. Golliher of NASA's Office of Biological and Physical Research as well as NSF Grant No. CBET-0651755. The authors thank Adam Maxwell, Brian MacConaghy, Jarred Swalwell, and Oleg Sapozhnikov of the UW Applied Physics Laboratory for their help in conducting preliminary experiments, and would also like to acknowledge many helpful suggestions from the Editor and reviewers.

References and links

[1]F. P. Chiaramonti and A. J. Jitendra eds., *Workshop on Critical Issues in Microgravity Fluids, Transport, and Reaction Processes in Advanced Human Support Technology*, Cleveland, OH, 2005, NASA/TM-2004–212940.

[2]E. Ungar, "Single phase vs. two-phase active thermal control systems for space application: A trade study," *AIAA 33rd Aerospace Science Meeting & Exhibit*, Reno, NV, 1995, AIAA 95–0634.

[3]Y. I. Grigoriev, E. I. Grigorov, V. M. Cykhotsky, Y. M. Prokhorov, G. A. Gorbenko, V. N. Blinkov, I. E. Teniakov, and C. A. Maluhkin, "Two phase heat transport loop of central thermal control system of the international space station 'Alpha' Russian segment," *Proceedings of the 31st Heat Transfer Conference*, Houston, 1996, pp. 9–18.

[4]L. Chow and R. Parish, "Condensation heat transfer in a microgravity environment," *AIAA 24th Aerospace Science Meeting & Exhibit*, Reno, NV, 1986, AIAA-0068.

[5]H. K. Cammenga, "Evaporation mechanisms of liquids," *Current Topics in Materials Science*, edited by K. Kaldis, North Holland Publishing Co., Amsterdam (1980).

[6]M. F. Schatz and G. P. Neitzel, "Experiments on thermocapillary instabilities," Annu. Rev. Fluid Mech. **33**, 93–127 (2001).

[7]R. O. Grigoriev, "Control of evaporatively driven instabilities of thin liquid films," Phys. Fluids **14**, 1895–1909 (2002).

[8]S. M. Som, J. T. Kimball, J. C. Hermanson, and J. S. Allen, "Stability and heat transfer characteristics of unsteady condensing and evaporating films," Int. J. Heat Mass Transfer **50**, 1927–1937 (2007).

[9]Z. Q. Chen, J. C. Hermanson, M. A. Shear, and P. C. Pedersen, "Ultrasonic monitoring of interfacial motion of condensing and non-condensing liquid films," Flow Meas. Instrum. **16**, 353–364 (2005).

[10]P. C. Pedersen, Z. Cakareski, and J. C. Hermanson, "Ultrasound monitoring of film condensation for applications in reduced gravity," Ultrasonics **38**, 468–490 (2000).

[11]E. W. Lemmon, M. O. McLinden, and D. G. Friend, "Thermophysical properties of fluid systems," *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, edited by P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Washington, D.C., 2005 (http://webbook.nist.gov).

Multichannel array diagnosis using noise cross-correlation

Laura A. Brooks[a] **and Peter Gerstoft**
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238
lbrooks@mecheng.adelaide.edu.au; gerstoft@ucse.edu

David P. Knobles
Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029
knobles@arlut.utexas.edu

Abstract: A practical application of noise cross-correlation for the diagnosis of a multichannel ocean hydrophone array is derived. Acoustic data were recorded on a horizontal line array on the New Jersey Shelf while Tropical Storm Ernesto passed through. Results obtained from active source measurements reveal that signals from several hydrophones, which were recorded on certain channels before the storm, are recorded on different channels after the storm. Noise cross-correlation of data recorded during the storm show when, and in what manner, these changes took place.

PACS numbers: 43.60.Ac, 43.60.Fg [NX]
Date Received: May 16, 2008 **Date Accepted:** July 5, 2008

1. Introduction

Lobkis and Weaver[1] showed, both theoretically and experimentally, that the Green's function between two points, and hence the acoustic travel time between them, can be extracted from their temporal cross-correlation within a diffuse ultrasonic field. Since then, noise cross-correlation techniques and their applications have been explored in numerous areas.[2–8]

In recent years, ocean noise temporal cross-correlation has been explained theoretically,[9] and demonstrated experimentally.[10] Siderius *et al.*[11] applied the concept to passive fathometry, and showed that it can be used to approximate sea floor structure. Sabra *et al.*[12] used noise cross-correlation for array localization and self-synchronization. These studies all concluded that accurate direct path acoustic travel times between hydrophones can be obtained from noise cross-correlation in the ocean.

During the Shallow Water 2006 experiments, which were conducted off the New Jersey shelf, Tropical Storm Ernesto passed through the experimental area. Large sea state and wind conditions started to develop during the evening of 1 September, resulting in all experimental activities ceasing until 3 September. The acoustic arrays remained operative throughout this period. Analysis of experimental data recorded on a 20 hydrophone horizontal line array (HLA) after the storm revealed that several channels in the array had switched.

For the purpose of this express letter, *channel switching* means that the signals from a given hydrophone which were previously recorded on a certain channel, are subsequently recorded on a different channel; *hydrophone* refers to the physical transducer in the array; and *channel* refers to the recording medium where the data from a hydrophone were stored. Hydrophone and channel numbers matched upon deployment of the array [see Fig. 1(b)], but not after the channel switching occurred [see Figs. 1(c) and 1(d)].

This express letter describes how ambient noise cross-correlation of array data from the storm are used to diagnose a problem of channel switching that occurred between hydrophone pairs. When channel switching occurs, a set of two given channels begin to record data

[a]Also at the School of Mechanical Engineering, University of Adelaide, Adelaide, Australia.

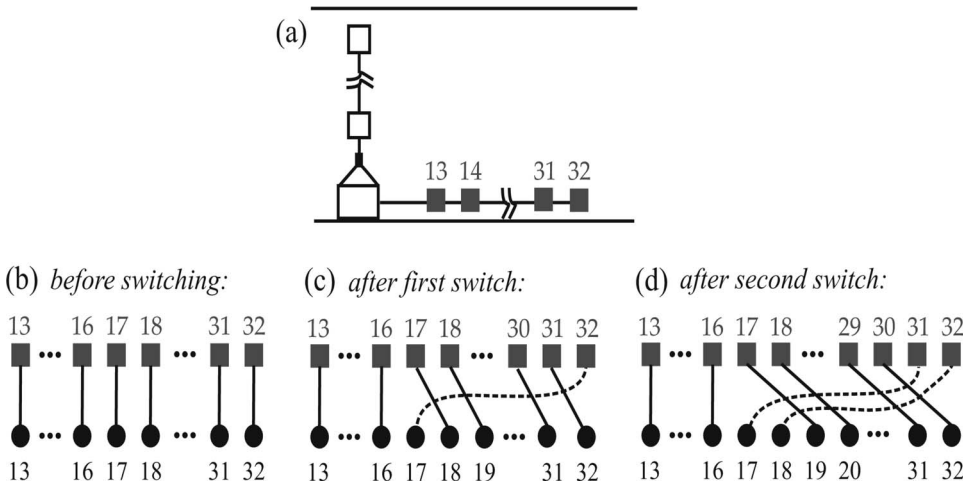


Fig. 1. (a) Array geometry. (b)–(d) Schematics showing the hydrophone-channel connections: (b) At deployment, (c) after the first switch, and (d) after the second switch.

from a different pair of hydrophones, and hence the interhydrophone travel time, determined from cross-correlation of the signals, will change. The estimated travel times, and specifically changes therein, are used to determine on which channel the data from each hydrophone are being recorded at any given time during the day. Consequently, the time and manner in which the channel switching occurs is ascertained.

2. Analysis of recorded data

The array [see Fig. 1(a)] is an L-shaped array with a 256-m-long horizontal portion (HLA) consisting of 20 hydrophones with tapered spacing.

Preliminary analysis of acoustic data (see Sec. 2.1) from active sources collected throughout the experiment shows that switching of channels occurred. Since no active source experiments were undertaken during the storm, further details of the switching could not be determined using traditional techniques; however, ambient noise cross-correlations (see Sec. 2.2) show when, and in what manner, the switching occurred.

2.1 Active sources

Data from both broadband pulses and combustive sound sources show that the signals recorded on each channel correspond to the correct hydrophones before the storm [see the *before switching* configuration of Fig. 1(b)]. The signal from a broadband 1100–2950 Hz energy pulse, recorded at 6:04 Z (Zulu) on 31 August, 21° from the HLA axis, and 1385 m from the closest hydrophone (hydrophone 32), was projected to the on-axis direction. The envelope of the projected signal received by each HLA channel (directly interchangeable with hydrophone numbers for this case) is shown in Fig. 2(a). The time of arrival is plotted relative to the first direct arrival. Several reflected arrivals are observable on each channel at times later than the direct. As expected, each arrival is received first by the channel corresponding to the closest hydrophone (channel 32), and the arrival times increase as the channel number decreases. The time interval between arrivals on each channel increases due to tapering of the array spacing.

After the storm, data from both linear frequency modulated (LFM) sweep sources and combustive sound sources indicated that some switching of channels occurred. The match filtered signal from a 1100 to 2900 Hz 1 s duration LFM source, held 10 m below the water surface, recorded at 14:40 Z on 3 September, on-axis with the HLA, and 150 m from the closest hydrophone, is shown in Fig. 2(b). The signals recorded on each channel no longer correspond to the correct hydrophones. The data from hydrophones 31 and 32 are recorded on channels 17

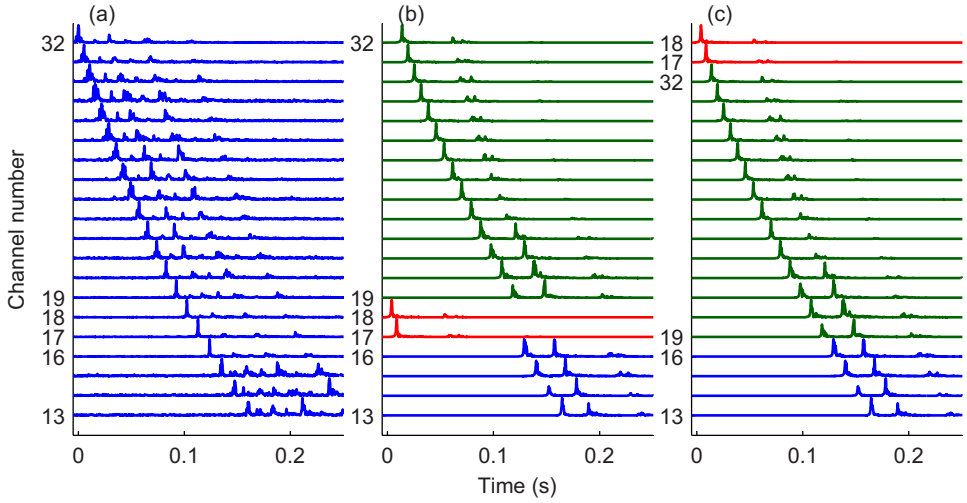


Fig. 2. (Color online) Envelope of the signal recorded on each channel from active source testing (a) before the storm [31 August, 6:04 Z], and (b) after the storm [3 September, 14:40 Z]. (c) Signal envelope from after the storm with channels resorted. Channels 14–15 and 20–31 are unnumbered, but are in order between 13 and 16, and between 19 and 32, respectively.

and 18, respectively, and the data from hydrophones 17 to 30 are all recorded two channels higher than expected [see Fig. 1(d)]. Figure 2(c) shows the data after re-sorting the channels so that the data from hydrophones 13 through 32 are in order. The arrival times and the manner in which the shape of the envelope evolves are consistent, which shows that data recorded after switching of the channels occurred are reliable (i.e., acoustic data from each hydrophone are still being recorded).

2.2 Ambient noise cross-correlations

The cross-correlation between two hydrophones, denoted A and B, as a function of time delay, t, is defined as

$$C_{AB}(t) = \int_{-\infty}^{\infty} p_A(t+\tau)p_B(\tau)d\tau, \tag{1}$$

where p is the pressure recorded at each hydrophone, and τ is time. It has been shown[9,13] that the arrival-time structure of the Green's function, between hydrophones A and B, G_{AB} defined as the signal which would be received at A given a unit impulsive source at B, can be extracted from the time derivative of the ocean noise cross-correlation function:

$$\frac{\partial C_{AB}(t)}{\partial t} \simeq -[G_{AB}(t) - G_{AB}(-t)]. \tag{2}$$

The raw cross-correlation, rather than its time derivative, is often used as an approximation to the Green's function,[10,14] and for a finite bandwidth signal this can be a good approximation, since the cross-correlation and its derivative closely resemble one another. However, if exact arrival times are desired, the cross-correlation time derivative should be employed, as this corrects for the $\pi/2$ phase difference between the raw cross-correlation arrival peak, and that of the Green's function.[9]

Ship-dominated 20–100 Hz noise was recorded throughout the day that Tropical Storm Ernesto passed through. After preprocessing the data, cross-correlation periods of 6 h 24 min, which corresponds to the length of a single data file, were employed. Since the cross-correlation wave forms are relatively narrow, it is sufficient to consider the raw cross-

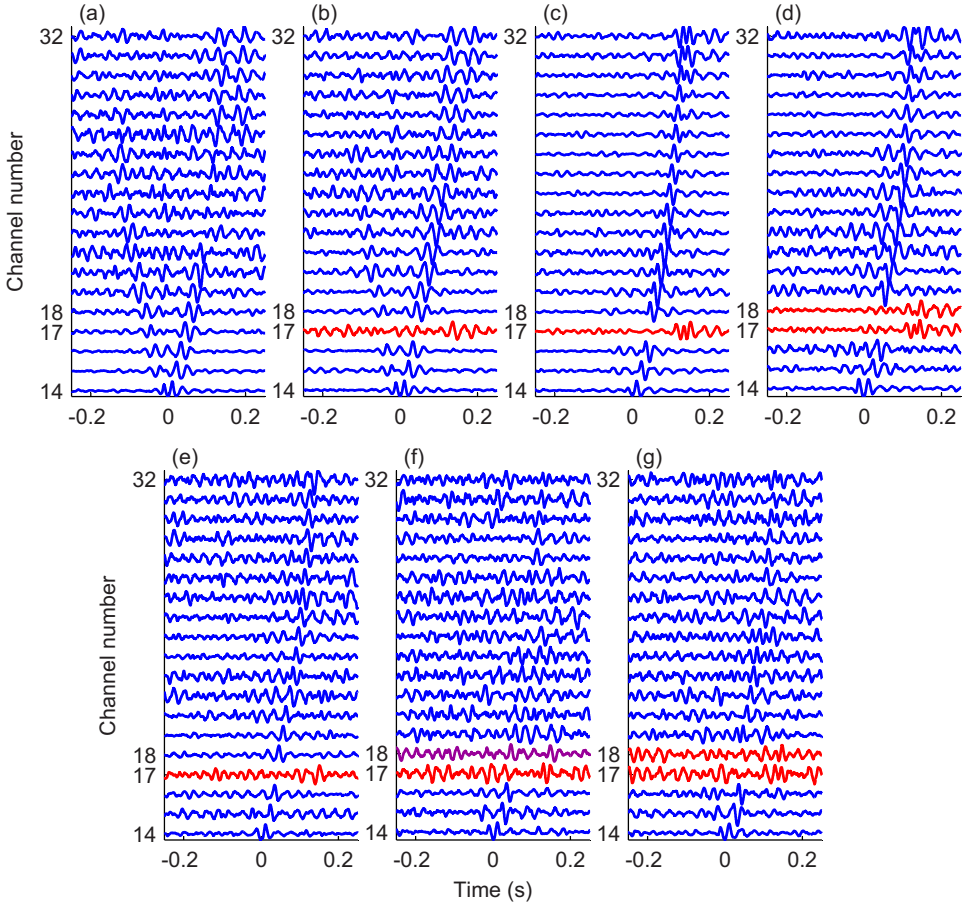


Fig. 3. (Color online) Short time cross-correlation (6 h 24 min) between channel 13 (first HLA hydrophone) and all other channels from the time period surrounding the first switch (a) 7:07:37 Z and (b) 7:14:01 Z, and from the time period surrounding the second switch (c) 12:08:40 Z and (d) 12:15:05 Z. (e)–(g) Cross-correlations 20 s long at 12:14 Z for times (e) 20–40 s, (f) 30–50 s, and (g) 40–60 s.

correlations for the purpose of determining which channels are recording data from which hydrophones. Due to the short cross-correlation times, peak amplitudes of the postprocessed data were prone to bias from high amplitude directional sources. However, even if the source field is not isotropic, the time of the cross-correlation peak between the hydrophones increases with distance. Examination of the raw cross-correlations over the entire day revealed three main findings: (a) the channel switching occurred in two steps; (b) an approximate time at which each switch occurred can be determined; and (c) the raw signals, and hence the signal cross-correlations, recorded for a significant time period before and after each switch, exhibit high levels of noise on the higher channels.

Cross-correlations from times prior to and after each switch occurred are shown in Figs. 3(a)–3(d). Figure 3(a), which is the cross-correlation starting at 7:07:37 Z, shows that the channels are in order, and the cross-correlation peak corresponds to a move-out velocity of approximately 1500 m/s, as expected. The cross-correlation of data starting at 7:14:01 Z [see Fig. 3(b)] shows an anomaly in the signal recorded by channel 17. The signal does show a small peak at the expected move-out velocity; however, the major peak occurs at a time of just under 0.2 s, which is the expected cross-correlation time for channel 32. The signals recorded on channels 18 through 32 exhibit major peaks corresponding to those expected for one channel

lower than their assigned values. These results suggest that channel 17 switched within the period 7:17 Z±3 min, the relative peak amplitudes indicating that the switch was closer to the start of this time period. Note that the cross-correlations for channels above 17 exhibit high levels of noise. High noise levels were, in fact, observed on channels 18–32 for a 2 h period surrounding this time (5:40–7:40 Z), suggesting that the channel switching and the increase in noise are linked. The hydrophone-channel connections after this first switch are shown in Fig. 1(c).

Results of cross-correlation of data starting at 12:08:40 Z are shown in Fig. 3(c). Channel 17 is still the only channel to have switched. Cross-correlation of data immediately following this, starting at 12:15:05 Z [see Fig. 3(d)], suggests that channel 18 has also switched. Now channels 17 and 18 are recording data from hydrophones 31 and 32, and channels 19–32 are recording data from two hydrophones less than their number. This configuration matches the "after switching" description of Sec. 2.1 [see Figs. 1(d) and 2(b)]. Cross-correlations with higher numbered channels exhibit less noise than during the first switch, and high noise levels were only observed for a few minutes around the time of the second switch.

Cross-correlations over shorter time periods were employed to narrow down the time window during which the second switch occurred. Cross-correlations of 20 s duration were calculated from 12:08:40 to 12:21:29 Z. Three results from the minute of 12:14 Z are shown in Figs. 3(e)–3(g): (e) 20–40 s, (f) 30–50 s, and (g) 40–60 s. Due to the shorter duration of the cross-correlation time, the cross-correlations exhibit high noise levels; however, the peak arrivals can still be observed. The first cross-correlation [Fig. 3(e)] suggests that channel 17 has switched but 18 has not. The second cross-correlation [Fig. 3(f)] is the least clear, but suggests that the switch occurs during this time period, since both the true and delayed arrival are seen. The third cross-correlation [Fig. 3(g)] also exhibits the true and delayed arrival; however, the delayed peak is more dominant, suggesting that the switch occurred closer to the start of the cross-correlation period. The likely time interval during which the second channel switching occurred is thus 12:14:45 Z±5 s.

Once the switch times had been determined, data from the entire day of 2 September were correlated in three time segments: (a) immediately before either switch, (b) after the first switch but before the second switch, and (c) immediately after the second switch. If the channels are re-sorted so that the data from hydrophones 13 through 32 are in order, and then plotted as a function of distance from hydrophone 13, the cross-correlation peaks are seen to increase linearly in time with distance along the array, in agreement with the results previously described [see Figs. 4(a)–4(c)]. The noise levels are lower than those in Fig. 3 due to the longer cross-correlation periods. The envelopes of the time derivatives of the cross-correlation functions, which, as previously mentioned, relate to the arrival times between hydrophones, are shown in Figs. 4(d)–4(f) [corresponding to data in Figs. 4(a)–4(c), respectively]. The simulated travel times between hydrophones (determined using OASES[15]) are also shown. The envelope peaks are in agreement with the direct arrival, and also with the surface reflected arrival at greater distances (the surface reflected arrival is not seen at closer distances due to the steeper grazing angles, which are accompanied by greater bottom loss). Although not shown, cross-correlations between all other channels in the HLA are in complete agreement with the findings presented here.

3. Conclusion

Results from active source experiments near the array prior to and after the day Tropical Storm Ernesto passed through the region (2 September, 2006) showed that some channels had switched during the storm.

Ambient noise cross-correlation of data from 2 September was successfully employed to determine more information about the nature of the channel switching. The cross-correlation analysis suggested that the switching occurred in two distinct stages. The mechanism which occurred at each stage was identical; the channel that was recording the data of hydrophone 17 started recording data from hydrophone 32, and the channels recording data from hydrophones 18 through 32 all moved down one hydrophone [see Figs. 1(c) and 1(d)]. The inferred time

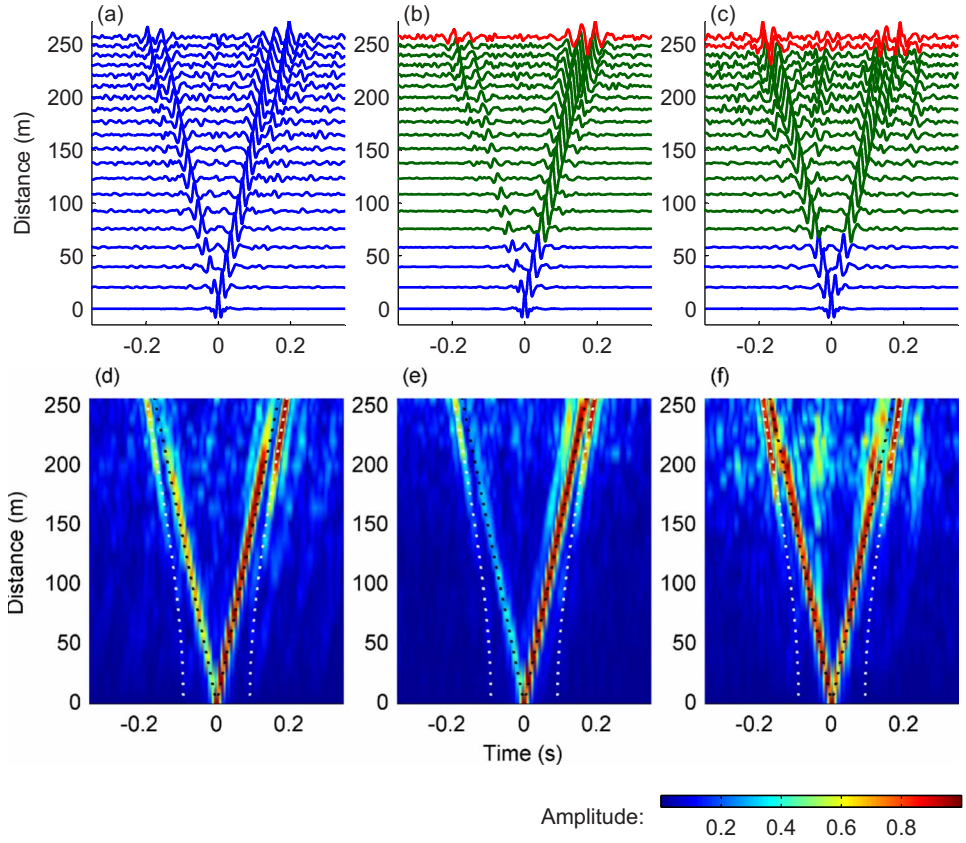


Fig. 4. (Color online) (a)–(c) Cross-correlations between channel 13 (first HLA hydrophone) and all other channels after re-sorting the channels, plotted as a function of distance from hydrophone 13 [(a) before, (b) between, and (c) after switches]. (d)–(f) Normalized envelopes of the cross-correlation function time derivatives, with simulated direct (black dotted lines) and surface reflected path travel times (white dotted lines) [(d) before, (e) between, and (f) after switches].

intervals during which each switch happened were narrowed down to the periods 7:17 Z ± 3 min and 12:14:45 Z ± 5 s. Elevated noise levels were observed on each channel for a significant time period surrounding each switch. The noise levels were especially high during the first switch, making it difficult to narrow down the time of the switch to as small a time window as was obtained for the second switch. Longer time period cross-correlations from before, between, and after the switches occurred, supported the findings. With the channels re-sorted to the correct hydrophones, the cross-correlation function time-derivative envelopes showed accurate arrival time structure.

Acknowledgments

Work supported by the Office of Naval Research under Grant No. N00014-05-1-0264, and by the Department of Energy National Energy Technology Laboratory via the Gulf of Mexico Hydrates Research Consortium, University of Mississippi. L.A.B. is appreciative of support from a Fulbright Postgraduate Award in Science and Engineering funded by Clough Engineering, and from the Defence Science and Technology Organisation, Australia.

References and links

[1]O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[2]K. van Wijk, "On estimating the impulse response between receivers in a controlled ultrasonic experiment,"

Geophysics **71**, SI79–SI84 (2006).
[3]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[4]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[5]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[6]K. Wapenaar and J. Fokkema, "Green's function representations for seismic interferometry," Geophysics **71**, SI33–SI46 (2006).
[7]E. Larose, A. Khan, Y. Nakamura, and M. Campillo, "Lunar subsurface investigated from correlation of seismic noise," Geophys. Res. Lett. **32**, L16201 (2005).
[8]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[9]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[10]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**, 1995–2003 (2004).
[11]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).
[12]K. G. Sabra, P. Roux, A. M. Thode, G. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[13]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[14]A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[15]H. Schmidt, *OASES Version 3.1 User Guide and Reference Manual*, (Massachusetts Institute of Technology, Cambridge, MA, 2004).

On the dynamic viscous permeability tensor symmetry

Camille Perrot, Fabien Chevillotte, and Raymond Panneton
Groupe d'Acoustique de l'Universite de Sherbrooke (GAUS), Department of Mechanical Engineering, Universite de Sherbrooke, Quebec J1K 2R1, Canada
camille.perrot@usherbrooke.ca, fabien.chevillotte@usherbrooke.ca, raymond.panneton@usherbrooke.ca

Jean-François Allard and Denis Lafarge
Laboratoire d'Acoustique de l'Université du Maine (LAUM), UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France
jean-francois.allard@univ-lemans.fr, denis.lafarge@univ-lemans.fr

Abstract: Based on a direct generalization of a proof given by Torquato for symmetry property in static regime, this express letter clarifies the reasons why the dynamic permeability tensor is symmetric for spatially periodic structures having symmetrical axes which do not coincide with orthogonal pairs being perpendicular to the axis of three-, four-, and sixfold symmetry. This somewhat nonintuitive property is illustrated by providing detailed numerical examples for a hexagonal lattice of solid cylinders in the asymptotic and frequency dependent regimes. It may be practically useful for numerical implementation validation and/or convergence assessment.

PACS numbers: 43.50.Gf, 43.55.Ev, 43.20.Wd, 43.58.Ta [MS]
Date Received: March 18, 2008 **Date Accepted:** July 8, 2008

1. Introduction

The viscous permeability tensor plays, for long-wavelength sound propagation in fluid-saturated rigid-framed porous structures, a role similar to the electric permittivity tensor in conducting media. In a comprehensive demonstration of Adler[1] in his classical textbook "Porous Media: Geometry and Transports," while examining the properties of the permeability tensor for flow in spatially periodic structures, it is clearly established that, essentially as a consequence of the linearity of the Stokes equations system for spatially periodic structures, the permeability is a positive definite and symmetric tensor, and can thus be diagonalized on an orthonormal basis. Wang[2] extends this property by showing that if the permeability is the same in two particular orthogonal transverse directions, the material is isotropic in all orthogonal transverse directions. There is an analog property in optics for the dielectric tensor of optically uniaxial cristals.[3]

In this express letter, emphasis is put on the—dynamic—permeability tensor (frequency dependent) for a hexagonal lattice of solid cylinders. Contrary to a simple cubic lattice, it is not intuitively clear this time that the orthogonal directions necessarily coincide with the principal permeability axes since these directions are geometrically distinguishable from one another. The demonstration is based on a direct extension in harmonic regime of a proof given by Torquato for symmetry property in static regime.[4] Consequently, the purpose of this express letter is first to clarify mathematically the reasons why the dynamic permeability tensor is symmetric for spatially periodic structures having symmetrical axes which do not coincide with orthogonal pairs being perpendicular to the axis of three- and sixfold symmetry. Second, it is proposed to illustrate this somewhat nonintuitive property by providing numerical examples for a hexagonal lattice of solid cylinders in the asymptotic and frequency dependent regimes. Third, it is finally mentioned that a systematic check of the diagonal and nondiagonal terms of the dynamic permeability tensor provides a fast and useful argument for convergence assessment of a numerical computation on this type of local periodic geometry.

The flow of a viscothermal fluid in a motionless homogeneous porous structure was described by D.L. based on ideas borrowed from electrodynamics as follows.[5] The angular frequency is ω and the time dependence is $\exp(-i\omega t)$. The statistical properties of the porous frame can be defined in homogenization volumes with dimensions much smaller than the wavelength of the acoustic waves that propagate in the saturating fluid. The microscopic quantities that describe the flow (pressure p, velocity $\mathbf{v}$) present variations at the microscopic scale in the homogenization volume. To smooth out these variations and leave only the macroscopic variations, angular brackets $\langle\ \rangle$, indicating a fluid-phase average, are introduced. By definition, it relates the macroscopic variable $\langle a\rangle$ to the microscopic variable a by

$$\langle a\rangle(\mathbf{r},t) = \int I(\mathbf{r}+\mathbf{x})f(\mathbf{x})a(\mathbf{r}+\mathbf{x},t)d^3\mathbf{x}, \tag{1}$$

where I is the characteristic function of the fluid phase (1 in the fluid Ω_f and 0 in the solid Ω_s), and f is a filtering smooth function nonzero only in some (homogenization) region surrounding $\mathbf{x}=\mathbf{0}$ and normalized to unity $\int I(\mathbf{r}+\mathbf{x})f(\mathbf{x})d^3\mathbf{x}=1$. $f(\mathbf{x})$ is typically a hat-shaped curve which is formed by a central plateau and two decreasing regions. The length L of the plateau and decreasing regions is chosen large compared to the material's local characteristic size ($L>a$), so that the averaging operation performed through Eq. (1) has the effect of decreasing the small microscopic fluctuations due to structure inhomogeneities. The so smoothed fields are thus varying at a much larger scale (wavelength $\lambda \gg L$) than the material's local characteristic size. The discontinuity of I at the contact surface $\partial\Omega$ between solid and fluid leads to the spatial average theorem

$$\langle \nabla a\rangle(\mathbf{r},t) = \nabla\langle a\rangle(\mathbf{r},t) + \int_{\mathbf{x}\in\partial\Omega} f(\mathbf{x}-\mathbf{r})\mathbf{n}(\mathbf{x})a(\mathbf{x},t)dS. \tag{2}$$

The vector $\mathbf{n}$ is the unit outward normal vector to the fluid domain. When the wavelengths become comparable to the local characteristic size, diffusion by the solid skeleton must be taken into account; and if the wavelengths are still reduced, all macroscopic propagating description becomes impossible. Here, we will consider only large wavelength situations where the diffusion theory can be ignored. Clearly, this is an attempt to directly model the geometry of the porous structure by defining a periodic unit cell that captures the intricate details of the porous material. As a result, this models an infinitely large matrix as if the porous structure within the unit cell were replicated in all other unit cells, without discontinuities in the porous structure between one unit cell and the adjacent one. Container effects are ignored, or at least negligible if the condition $L>a$ is well respected. At a given frequency, two equivalent relations linking the gradient of the macroscopic pressure to the macroscopic velocity are

$$\rho_0\alpha_{ij}(\omega)\frac{\partial\langle\mathbf{v}\rangle_j}{\partial t}(\mathbf{r}) = -\nabla_i\langle p\rangle(\mathbf{r}), \tag{3}$$

where ρ_0 is the equilibrium density of the fluid, and

$$\phi\langle\mathbf{v}\rangle_i(\mathbf{r}) = -\frac{1}{\eta}k_{ij}(\omega)\nabla_j\langle p\rangle(\mathbf{r}), \tag{4}$$

where ϕ is the porosity and η the viscosity of the fluid. Two second-order tensors, the dynamic tortuosity $\alpha_{ij}(\omega)$ and the dynamic permeability $k_{ij}(\omega)$, are defined by Eqs. (3) and (4). They only depend on frequency and on the geometry of the porous structure.

These two quantities, previously defined and studied by Johnson *et al.*,[6] not only characterize the response of the fluid to a spatially constant oscillating pressure gradient but also are relevant ingredients of the long-wavelength sound propagation theory. Indeed, for long-wavelength sound propagation in the material, the dynamic tortuosity tensor plays a role similar to the electric permittivity tensor.

When frequency decreases, the oscillatory flow locally becomes very similar to the static flow and Eq. (4) becomes

$$\phi\langle \mathbf{v}\rangle_i = -\frac{1}{\eta} k_{0ij} \nabla_j \langle p\rangle. \tag{5}$$

This relation is the Darcy law,[7] and the components k_{0ij} define the static viscous permeability tensor. When ω becomes very large, the effect of viscosity becomes negligible and $\alpha_{ij}(\omega)$ tends to the tortuosity tensor $\alpha_{\infty ij}$. The symmetry of k_{0ij} was demonstrated earlier for periodic structures by Ene and Sanchez-Palencia.[8] This was followed by a demonstration of the symmetry of the inverse Fourier transform of the dynamic permeability.[9] Some other properties of the inverse permeability tensor are also shown by Auriault.[10] Numerical determination of the dynamic permeability can be found for a simple periodic rib duct;[11] however, in this case, the symmetry property is practically useless since only one pressure gradient direction may be chosen.

Because the symmetry tensor property is relatively counterintuitive for apparently anisotropic porous structures, mathematical explanations are given in Sec. 2, and a straightforward consequence of this property for porous structures with trigonal and hexagonal symmetry is pointed out: that is, the invariance of the dynamic permeability tensor components in directions lying in the plane perpendicular to the axis of symmetry. Note that a similar property also exists for the permittivity tensor in uniaxial crystals.[3] In Sec. 3, steady and unsteady flow simulations in a porous structure with hexagonal symmetry are reported, with detailed systematic evaluations of the tortuosity, static, and dynamic permeability tensor components in the plane perpendicular to the axis of symmetry.

2. Symmetry of the dynamic viscous permeability tensor

As discussed in Ref. 5, describing the periodic oscillating flow created in a porous medium by an external unit harmonic pressure gradient $\mathbf{e}e^{-i\omega t}$, and using three individual unit solicitation vectors $\mathbf{e}^i$ in three perpendicular directions, with components $\mathbf{e}^i_j = \delta_{ij}$, one has to solve in the fluid volume Ω_f the following set of scaled equations (unsteady scaled Stokes problem):

$$\frac{-i\omega}{\nu}\mathbf{w}^j_i = -\nabla_i \pi^j + \Delta \mathbf{w}^j_j + \delta_{ij} \quad \text{in } \Omega_f, \qquad \nabla \cdot \mathbf{w}^j = 0 \quad \text{in } \Omega_f, \qquad \mathbf{w}^j = 0 \quad \text{on } \partial\Omega, \tag{6}$$

where $\nu = \eta/\rho_0$. The solution to the problem described by Eq. (6) is fixed by adding the condition that π^j is a spatially stationary or periodic field. This problem is relevant to sound propagation as long as the wavelength is large enough for the saturating fluid to behave as an incompressible fluid in volumes of the order of the homogenization volume (a period in the case of periodic structure).

From Eq. (4) and solution to Eq. (6), the viscous permeability components may be written in the form[12]

$$k_{ij}(\omega) = \phi\langle \mathbf{w}^j_i\rangle = \phi\langle \mathbf{w}^j_l \delta_{il}\rangle = \phi\langle \mathbf{w}^j_l \mathbf{e}^i_l\rangle = \phi\langle \mathbf{w}^j \cdot \mathbf{e}^i\rangle, \tag{7}$$

where, from the left-hand side of the first equation in Eq. (6), $\langle \mathbf{w}^j \cdot \mathbf{e}^i\rangle$ is given by

$$\langle \mathbf{w}^j \cdot \mathbf{e}^i\rangle = \frac{-i\omega}{\nu}\langle \mathbf{w}^j \cdot \mathbf{w}^i\rangle + \langle \mathbf{w}^j \cdot \nabla \pi^i\rangle - \langle \mathbf{w}^j \cdot \Delta \mathbf{w}^i\rangle. \tag{8}$$

The second term on the right-hand side of Eq. (8) is equal to 0. This can be shown in the following manner by rewriting

$$\langle \mathbf{w}^j \cdot \nabla \pi^i\rangle = \langle \nabla \cdot (\mathbf{w}^j \pi^i)\rangle - \langle \pi^i \nabla \cdot \mathbf{w}^j\rangle. \tag{9}$$

First, from the local incompressibility condition [the second equation in Eq. (6)], the last term in Eq. (9) is strictly equal to zero. Second, using the spatial average theorem, Eq. (2), the first term on the right-hand side in Eq. (9) is rewritten as

$$\langle \nabla \cdot (\mathbf{w}^j \pi^j) \rangle = \nabla \cdot \langle \mathbf{w}^j \pi^j \rangle + \int_{\mathbf{x} \in \partial\Omega} f(\mathbf{x} - \mathbf{r}) \mathbf{n}(\mathbf{x}) \cdot \mathbf{w}^j \pi^j d\mathbf{x}. \tag{10}$$

Since the fields are spatially stationary or periodic, the first term on the right-hand side in Eq. (10) is equal to zero, and from the no-slip condition [the third equation in Eq. (6)], the second term is also equal to 0. This proves that $\langle \mathbf{w}^j \cdot \nabla \pi^j \rangle = 0$. To further simplify Eq. (8), the third term on the right-hand side can be rewritten

$$\langle \mathbf{w}^j \cdot \Delta \mathbf{w}^i \rangle = \left\langle \frac{\partial}{\partial x_m} \left(\mathbf{w}^j \cdot \frac{\partial}{\partial x_m} \mathbf{w}^i \right) - \left(\frac{\partial}{\partial x_m} \mathbf{w}^j \right) \cdot \frac{\partial}{\partial x_m} \mathbf{w}^i \right\rangle. \tag{11}$$

It is shown by using the spatial average theorem that the first term on the right-hand side is equal to 0. Finally, from these simplifications, $k_{ij}(\omega)$ can be written in a symmetrical form in i and j,

$$k_{ij}(\omega) = -\frac{i\omega\phi}{\nu} \langle \mathbf{w}^j \cdot \mathbf{w}^i \rangle + \phi \left\langle \left(\frac{\partial}{\partial x_m} \mathbf{w}^j \right) \cdot \frac{\partial}{\partial x_m} \mathbf{w}^i \right\rangle. \tag{12}$$

It has been observed by Brown[13] that the same type of equation also describes the electrical current density in a conducting fluid saturating a nonconducting porous frame. Consequently, as for the permeability tensor, one can identify the tortuosity tensor elements from the resistivity tensor elements of the saturated medium, normalized by the resistivity of the saturating fluid, and multiplied by porosity.

These demonstrations clearly show that the viscous permeability tensor and tortuosity tensor are symmetric, $k_{ji}(\omega) = k_{ij}(\omega)$ and $\alpha_{ji}(\omega) = \alpha_{ij}(\omega)$. As a consequence, there exists a system of orthogonal axes, the principal axes, where the tensors have only diagonal elements different from zero (see, for example, Refs. 1–3). If the medium presents trigonal, tetragonal, or hexagonal symmetry, the axis of symmetry Z must coincide with one of the principal axes. Moreover, the invariance of the system through some discrete rotations along this axis necessarily means that the two transverse eigenvalues of the tensors are the same (transverse isotropy). In particular, the static permeability k_{0ij}, the tortuosity $\alpha_{\infty ij}$, and the dynamic permeability $k_{ij}(\omega)$ have values k_{0Z}, $\alpha_{\infty Z}$, and $k_Z(\omega)$ along axis Z, and different unique values k_{0P}, $\alpha_{\infty P}$, and $k_P(\omega)$ for all directions orthogonal to Z.

Parameters α_∞, and k_0, and $k(\omega)$ are important for the prediction of the acoustic properties of porous media and many numerical simulations have been performed to predict these parameters in periodic structures (see Ref. 12, and references therein). The constraint which exists when the structure has the uniaxial symmetry can provide a test for the precision of the simulations. An example is given in Sec. 3, where the tortuosity $\alpha_{\infty P}$, the static permeability k_{0P}, and the dynamic permeability $k_P(\omega)$ are evaluated in a hexagonal porous structure.

3. Flow simulation in a hexagonal porous structure

At $\omega = 0$, Eq. (6), with the condition that π^j is a stationary field, simply describe the viscous fluid motion in steady state regime. This is the steady Stokes problem for periodic structures, where $\mathbf{w}_0$ is the scaled static velocity field in the pore in square meters. Thus, the components k_{0ij} defining the static viscous permeability tensor are simply given by

$$k_{0ij} = \phi \langle \mathbf{w}_0^j \cdot \mathbf{e}^i \rangle. \tag{13}$$

At the opposite frequency range, when ω becomes very large, the viscous boundary layer becomes negligible and the fluid tends to behave as a perfect one, having no viscosity. Under these conditions, the perfect incompressible fluid formally behaves according to the electric problem.[6,13] The components $\alpha_{\infty ij}$ defining the tortuosity are now given by[5]

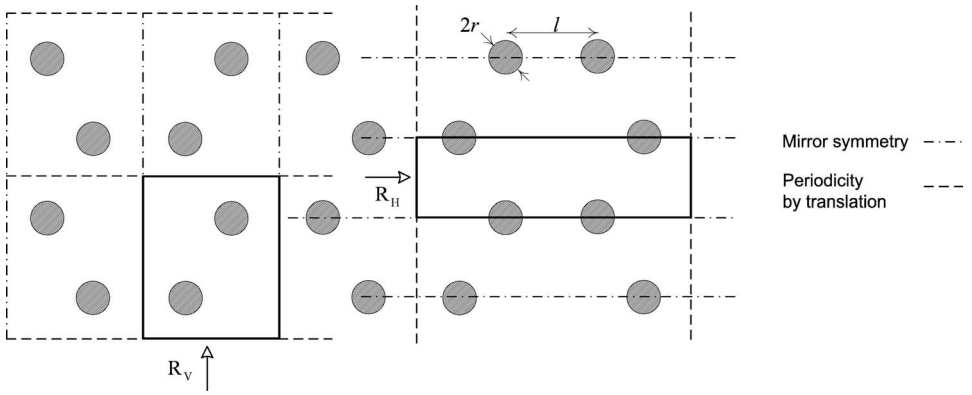


Fig. 1. Identification of two-dimensional rectangular periodic unit cells for vertical (R_V) and horizontal (R_H) wave propagations through a hexagonal lattice of solid cylinders.

$$\alpha_{\infty ij}^{-1} = \langle \mathbf{E}^j \cdot \mathbf{e}^i \rangle, \tag{14}$$

where $\alpha_{\infty ij}^{-1}$ denotes the inverse of the tortuosity tensor $\alpha_{\infty ij}$, and $\mathbf{E}$ is the scaled electric field that solves the corresponding electrical conduction problem for a porous medium filled with a conducting fluid and having an insulating solid phase, i.e.,

$$\mathbf{E} = -\nabla\varphi + \mathbf{e} \quad \text{in } \Omega_f, \quad \nabla \cdot \mathbf{E} = 0 \quad \text{in } \Omega_f, \quad \mathbf{E} \cdot \mathbf{n} = 0 \quad \text{on } \partial\Omega, \tag{15}$$

and φ is a spatially stationary or periodic scalar field representing the deviatoric part of the electric potential.

Taking the external macroscopic field $\mathbf{e}$ in the horizontal and vertical directions, numerical computations of the fields $\mathbf{E}$, $\mathbf{w}_0$, and $\mathbf{w}$ were performed on the periodic rectangles R_H and R_V of the hexagonal porous structure depicted in Fig. 1, where l is the length between two cylinders and r is the radius of a solid cylinder.

The radius of the cylinders is $r=(1-\phi)\Lambda'/\phi \approx 0.163$ mm, and $l = \sqrt{4\pi r(\Lambda'+r)/(3\sqrt{3})} \approx 0.894$ mm; corresponding to a porosity $\phi=0.92$ and a thermal characteristic length $\Lambda'=1.87$ mm (which is equal to twice the interconnected pore fluid volume to pore wet surface ratio). The corresponding scaled electric, steady, and unsteady Stokes problems were solved using a finite-element commercial code.[14] No-slip boundary conditions at the pore walls, and periodicity of π_0, $\mathbf{w}_0$, and $\mathbf{w}$ were prescribed. Neumann boundary conditions on the fluid–solid interface, and periodicity on the inlet–outlet surfaces were used for φ. Additional Neumann boundary conditions are set in the remaining lateral borders due to the symmetries of the problems.

The number of elements and their distribution in the fluid phase regions of R_H and R_V were varied, with attention paid especially to the throat and the near-wall areas, to examine the accuracy and convergence of the field solutions. As an illustration, it can be seen in Fig. 2 that once convergence is achieved, there is an almost perfect superposition of the diagonal terms of the dynamic permeability tensor values [i.e., $k_{11}(\omega)$ and $k_{22}(\omega)$] obtained for the external orthogonal solicitations. Furthermore, we show in Figs. 3 and 4 the two components of the static and high frequency scaled patterns obtained for excitations along the two principal directions in the plane perpendicular to the axis of symmetry of the periodic geometry. As previously noticed by several authors, such as Martys and Garboczi,[15] due to the nonslip condition, the fluid flow paths are more concentrated than the electric-current paths arising when effects of fluid viscosity are minimal. Once the flow fields are known, the static and dynamic viscous permeability and the tortuosity tensors are computed using Eqs. (13), (7), and (14), respectively. This yields the following tensors in the asymptotic regimes:

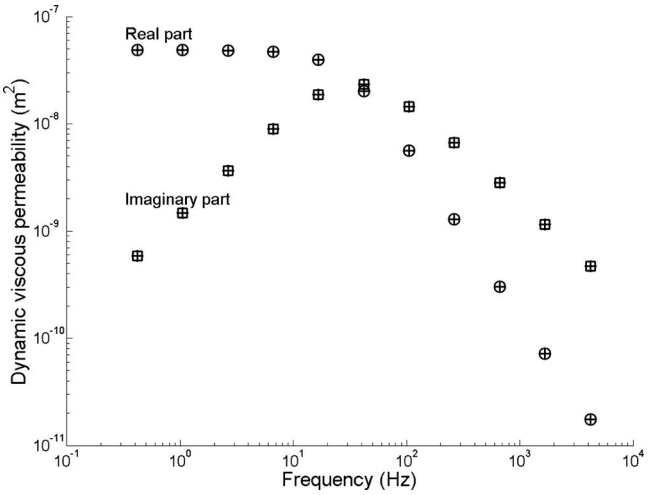


Fig. 2. Dynamic viscous permeability of a hexagonal lattice of solid cylinders: There is an almost perfect superposition of the permeability values computed from any orthogonal line pairs perpendicular to the sixfold axis of symmetry, suggesting that this property might be used for a convergence check.

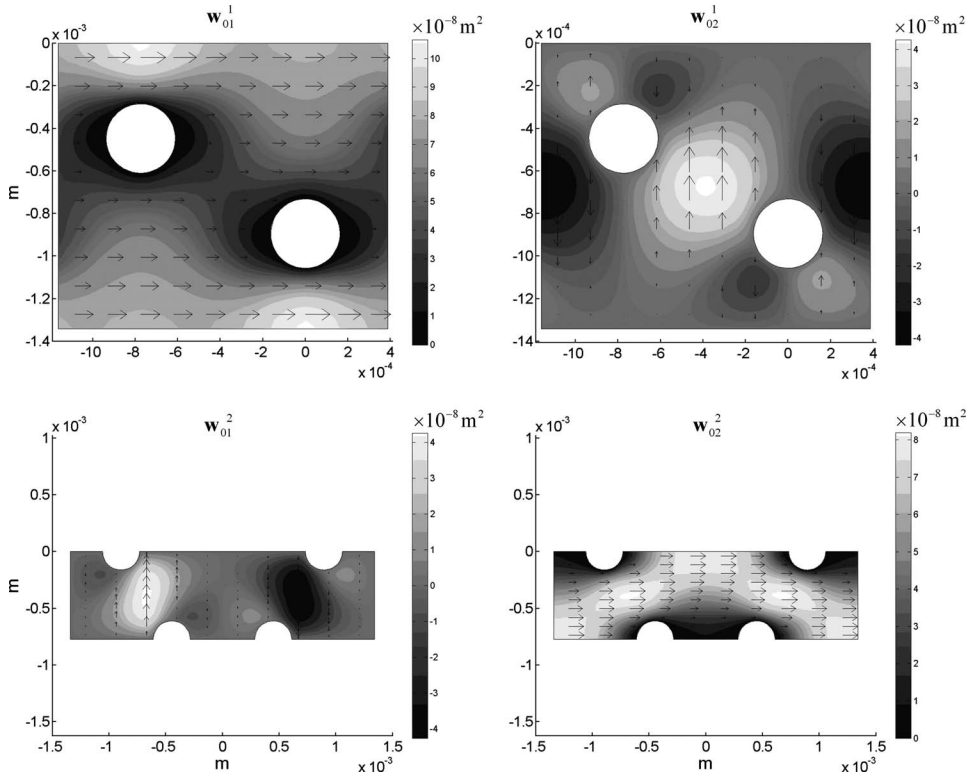


Fig. 3. Vertical (top) and horizontal (bottom) static scaled velocity fields obtained by solving the steady Stokes problem in principal directions of the periodic porous structure.

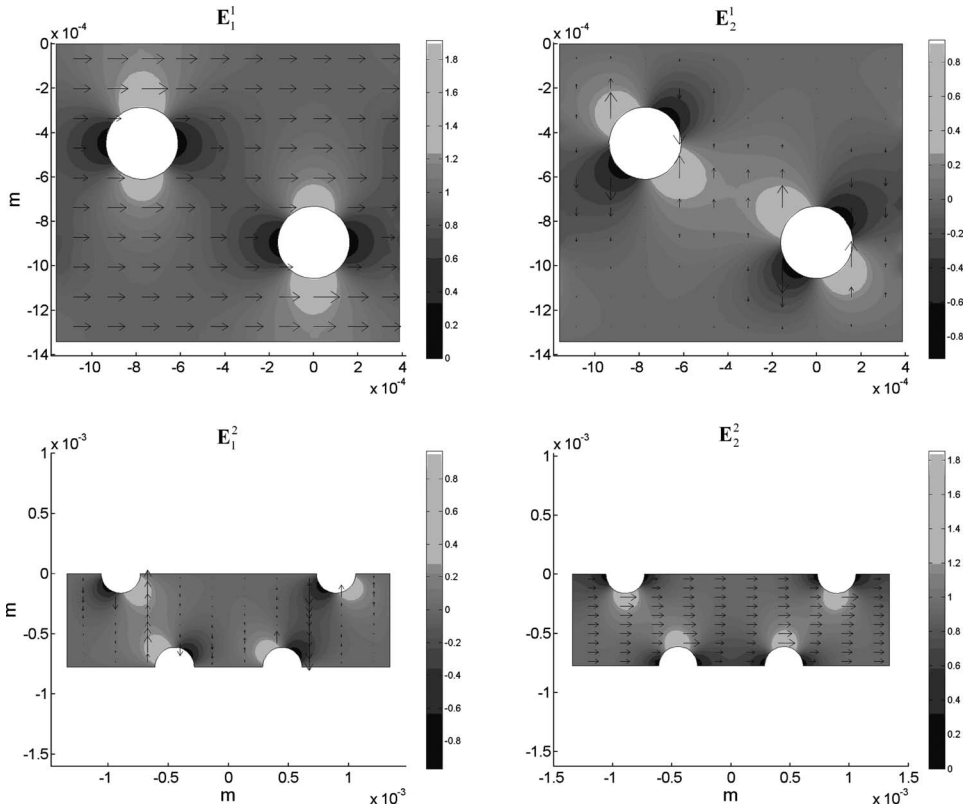


Fig. 4. Vertical (top) and horizontal (bottom) scaled electric fields obtained by solving the electric problem in principal directions of the periodic porous structure.

$$\begin{bmatrix} k_{011} & k_{012} \\ k_{021} & k_{022} \end{bmatrix} = \varphi \begin{bmatrix} \langle \mathbf{w}_{01}^1 \rangle & \langle \mathbf{w}_{01}^2 \rangle \\ \langle \mathbf{w}_{02}^1 \rangle & \langle \mathbf{w}_{02}^2 \rangle \end{bmatrix} = \begin{bmatrix} 4.9420 & 0 \\ 0 & 4.9353 \end{bmatrix} \times 10^{-8}\ \mathrm{m}^2, \tag{16}$$

and

$$\begin{bmatrix} \alpha_{\infty 11} & \alpha_{\infty 12} \\ \alpha_{\infty 21} & \alpha_{\infty 22} \end{bmatrix} = \frac{1}{\langle \mathbf{E}_1^1 \rangle \langle \mathbf{E}_2^2 \rangle - \langle \mathbf{E}_2^1 \rangle \langle \mathbf{E}_1^2 \rangle} \cdot \begin{bmatrix} \langle \mathbf{E}_2^2 \rangle & -\langle \mathbf{E}_2^1 \rangle \\ -\langle \mathbf{E}_1^2 \rangle & \langle \mathbf{E}_1^1 \rangle \end{bmatrix} = \begin{bmatrix} 1.0802 & 0 \\ 0 & 1.0801 \end{bmatrix}. \tag{17}$$

In both cases, the relative differences between the horizontal and vertical directions are less than 0.13%. Also, the nondiagonal terms are numerically equal to zero. These results are consistent with theoretical results presented in Sec. 2, and, as a consequence, prove the validity of the numerical implementation.

Finally, these results provide, in the limit of the precision of the finite-element method, a numerical illustration of the symmetry property of the dynamic viscous permeability tensor. In a plane perpendicular to the axis of a porous material with hexagonal symmetry, tortuosity, static, and dynamic viscous permeability tensors are reduced to a constant diagonal element.

4. Conclusion

A simple illustration of a somewhat nonintuitive property of the dynamic viscous permeability tensor in rigid-frame porous media was presented in the case of a wavelength much larger than the pore size. For periodic porous structures with hexagonal symmetry presenting different

local configurations to a wave propagating in different directions in the plane perpendicular to the axis of symmetry, the dynamic permeability tensor is diagonal and constant in the different directions. This property can notably be used for error estimation in numerical computations.

References and links

[1]P. M. Adler, "Flow in spatially periodic structures," in *Porous Media: Geometry and Transports*, edited by H. Brenner *et al.* (Butterworths, London, 1992), pp. 144–185.

[2]C. Y. Wang, "Stokes slip flow through square and triangular arrays of circular cylinders," Fluid Dyn. Res. **32**, 233–246 (2003).

[3]M. Born and E. Wolf, "Optics of crystals," in *Principles of Optics* (Cambridge University Press, Cambridge, 2002), pp. 790–849.

[4]S. Torquato, "Relationship between permeability and diffusion-controlled trapping constant of porous media," Phys. Rev. Lett. **64**, 2644–2646 (1990).

[5]D. Lafarge, "Modèles linéaires de propagation," in *Milieux poreux et poreux stratifiés (Linear Models of Propagation in Porous Media and Stratified Porous)*, Matériaux et Acoustique (Materials and Acoustics), edited by M. Bruneau and C. Potel (Lavoisier, Paris, 2006), pp. 143–187.

[6]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).

[7]H. Darcy, *Les Fontaines Publiques de la Ville de Dijon* (Victor Dalmont, Paris, 1856).

[8]H. I. Ene and E. Sanchez-Palencia, "Équations et phénomènes de surface pour l'écoulement dans un modéle de milieu poreux (Equations and surface phenomena for flow through a porous media model)," J. Mec. **14**, 73–108 (1975).

[9]E. Sanchez-Alencia, "Fluid flow in porous media," in *Non-Homogeneous Media and Vibration Theory* (Springer, Berlin, 1980), pp. 149–154.

[10]J.-L. Auriault, "Dynamic behaviour of a porous medium saturated by a Newtonian fluid," Int. J. Eng. Sci. **18**, 775–785 (1980).

[11]J.-L. Auriault, L. Borne, and R. Chambon, "Dynamics of porous saturated media, checking of the generalized law of Darcy," J. Acoust. Soc. Am. **77**, 1641–1650 (1985).

[12]C. Perrot, F. Chevillotte, and R. Panneton, "Dynamic viscous permeability of an open-cell aluminium foam: Computations vs experiments," J. Appl. Phys. **103**, 024909–8 (2008).

[13]R. J. S. Brown, "Connection between formation factor for electrical-resistivity and fluid-solid coupling factor in Biot equations for acoustic waves in fluid-filled porous media," Geophysics **45**, 1269–1275 (1980).

[14]Comsol 3.4, WTC-5 pl. Robert Schuman, 38000 Grenoble, France.

[15]N. Martys and E. J. Garboczi, "Length scales relating the fluid permeability and electrical conductivity in random two-dimensional model porous media," Phys. Rev. B **46**, 6080–6090 (1992).

A mode-based technique for estimating uncertainty in range-averaged transmission loss results from underwater acoustic calculations

R. A. Zingarelli
Code 7182, Naval Research Laboratory, Stennis Space Center, Mississippi 39529
zingarelli@nrlssc.navy.mil

Abstract: The equivalence of range and frequency averaging of acoustic propagation model results, based on the similarity of their analytic forms in mode calculations, was shown by Harrison [J. Acoust. Soc. Am. **97**, 1314–1317 (1995)]. Here it is shown how oceanographic measurement errors and receiver bandwidth can be mapped into uncertainty in the number of modes being propagated. This can be mapped into range boundaries for averaging calculations, thereby giving upper and lower confidence boundaries for frequency-averaged transmission loss calculations. Examples of the application of this technique to synthetic data, where the measurement uncertainties are known and deliberately included, are shown.

PACS numbers: 43.30.Zk [WS]
Date Received: May 9, 2008 **Date Accepted:** July 5, 2008

1. Introduction

Errors are inherent in any physical measurement. In practice, they are divided into systematic and random errors, the latter of which is frequently called uncertainty. In experimental work, great pains are taken to eliminate systematic errors. Random errors, however, can at best only be minimized.[1] Similarly, computational models have associated errors. Systematic errors occur when a model is either performing a calculation that is not closely matched with the physical reality that it is intended to mimic, or when physical approximations inherent in a model are overly simplistic. Random errors arise through either machine precision limitations, or when the limitations of mathematical approximations being used are exceeded.[2]

A more immediate and practical type of error pertains to the correspondence between experimental measurements and computational results. That is, if an investigator makes a computational prediction using a model believed to encompass the relevant physical processes and not having any significant numerical problems, then proceeds to make a corresponding physical measurement (again, believed to be free of systematic errors, and with relatively small random uncertainties), how closely will the two results match? Note that there is an additional component to this question that is not addressed in the preceding discussions of experimental and computational errors: that of the characteristics of the experimental measurement systems used.

To bring this discussion concretely into the context of underwater acoustics, the question becomes: If acoustic transmission loss (TL) as a function of range is predicted using some computational model, and these results are subsequently processed through an algorithm to translate them into the same units as physical measurements, how closely can one expect the prediction to match actual at-sea measurements. To address this question, we must first understand the properties of the experimental data processing chain, and of the algorithm used to translate the pressure values of the computational model into quantities that can be directly compared to experimental data. One mapping between computational and experimental results is to equate range-averaged TL values generated by a model at a single center frequency to frequency-averaged TL from an at-sea measurement. Early efforts at this simply tuned the range-averaging interval until the smoothness of the computational result seemed similar to that of the data.[3] A later approach makes use of the similarity of mathematical forms between fre-

quency and range averaging in a simple mode representation.[4] Because this method makes explicit use of the number of available modes and requires only TL output from an acoustic model as its input, and because estimation of mode count variation as a function of some environmental uncertainties is fairly straightforward, the error prediction method presented in this work builds on this second approach.

Having examined the mapping between computational and experimental results, the correspondence question presented earlier can be restated more precisely: Having some knowledge of the modal properties of the acoustic field from a computational model, how susceptible to random variations in the environment and the measurement process will a given set of experimental data be. The result of such an error estimate may equivalently be expressed as error bars on the data, or confidence bounds on the computational result.

2. Method

The equivalence given in Ref. 4 between the mode representations of range- and frequency-averaged acoustic fields is

$$I_f = 2\pi \int \left| \sum_{n=1}^{\gamma N_0} \phi_n^0(z_s)\phi_n^0(z_r)[\gamma K_n^0 r_0]^{-1/2} \exp[i\gamma K_n^0 r_0] \right|^2 \exp[-(\gamma-1)^2/\alpha^2]d\gamma \Bigg/ \int \exp[-(\gamma-1)^2/\alpha^2]d\gamma, \quad (1)$$

$$I_r = 2\pi \int \left| \sum_{n=1}^{N_0} \phi_n^0(z_s)\phi_n^0(z_r)[\gamma K_n^0 r_0]^{-1/2} \exp[i\gamma K_n^0 r_0] \right|^2 \exp[-(\gamma-1)^2/\alpha^2]d\gamma \Bigg/ \int \exp[-(\gamma-1)^2/\alpha^2]d\gamma, \quad (2)$$

where I_f and I_r are, respectively, the frequency- and range-averaged intensities, K_n^0 is the wave number for the band center frequency, α is fractional bandwidth, and γ is the respective frequency or range integration variable as either f/f_0 or r/r_0.

Equating the two expressions gives a simple relationship, $R_{\text{avg}} = \alpha r$, for the range-averaging interval as a function of range and fractional bandwidth.

The equivalence between the expressions in Eqs. (1) and (2) is strictly valid only in a range-independent environment. However, the long use of such averaging methods in practical acoustic oceanographic work has shown that it is a good approximation for third-octave bandwidth measurements in weakly range-dependent environments.[5] (While not referenced in Ref. 5, a copy of the inversion model source code supplied by the authors of that work shows that the range-frequency averaging equivalence from Ref. 4 is used.) Further limitations on this approximation are discussed in Ref. 4, but chief among these are that the bandwidth is relatively narrow (i.e., that the range over which γ varies is always close to 1) and that there are no strong foci, caustics, or shadow boundaries. Although computational speed improvements since publication of Ref. 4 have made explicit frequency-averaging practical for some uses and thus reduced the need for this approximation, it is still useful for reducing computational requirements for very intensive calculations, such as the genetic algorithm bottom parameter inversions discussed in Ref. 5.

Regardless of the improvements that have been made in computation time, the reason here for using this approach is that random errors can be incorporated into the above-presented expressions as an uncertainty in the number of modes being propagated by modifying Eq. (2):

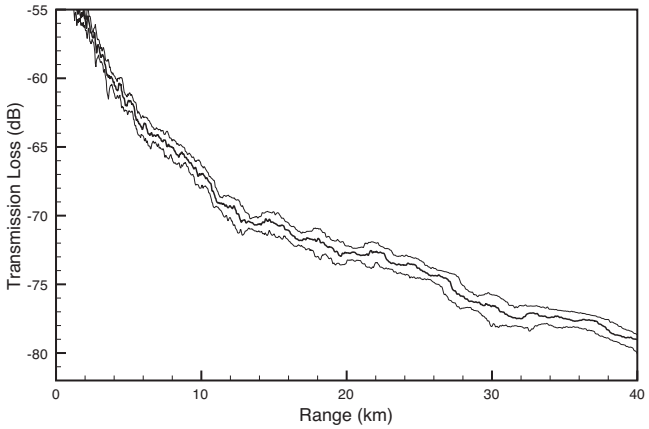


Fig. 1. Third-octave equivalent range average transmission loss curve and uncertainty bounds. The mode solution described in the text was used to calculate the transmission loss. Bandwidth (i.e., frequency uncertainty) was the source of uncertainty used to generate the bounds.

$$I_r = 2\pi \int \left| \sum_{n=1}^{\beta N_0} \phi_n^0(z_s)\phi_n^0(z_r)[\gamma K_n^0 r_0]^{-1/2} \exp[i\gamma K_n^0 r_0] \right|^2 \exp[-(\gamma-1)^2/\alpha^2] d\gamma \Bigg/ \int \exp[-(\gamma-1)^2/\alpha^2] d\gamma, \quad (3)$$

where $\beta = (1 \pm \Delta N_0/N_0)$, and ΔN_0 is the total uncertainty in the number of propagating modes.

The noteworthy feature of Eq. (3) is that only the numerator is modified by the variation in the number of modes. Thus, in a practical implementation as an error estimation program, the summation range in the averaging routine is modified by the factor β, while the bin count (i.e., the denominator) remains unchanged. Being used in the same way as γ in Eq. (1), much the same set of limitations discussed previously should apply for β.

The mode count used in the error estimation procedure should be the actual number of modes with significant energy being propagated. If we assume the waveguide is nearly isotropic, the number of propagating modes may be estimated as $(2Z_b/\lambda)\sin\theta + 1/2$, where Z_b is the local water depth, λ is the acoustic wavelength, and θ is the maximum propagating angle. While an estimate for the maximum propagating angle may be made from the bottom critical angle, due to bottom inhomogeneities and range dependencies it is preferable to glean this information directly from the propagating field. For ray, mode, or split-step Fourier parabolic equation based acoustic propagation models, this angle is available within the calculation. For other solution methods, this angle information is readily available through Fourier analysis.

The uncertainty of the number of modes, ΔN_0, can arise from the bandwidth of the source or receiver, from uncertainty in the total depth of the waveguide, from oceanographic profile uncertainties, and (with somewhat more difficulty) from uncertainties in source and receiver placement. The first two uncertainty sources are readily utilized by estimating the total mode count differences when their respective parameters are varied. The mode count uncertainties arising from oceanographic profile and positional uncertainties are not so easily defined, and are the subject of ongoing research. For the examples presented here, we will concentrate on the TL uncertainty arising from bandwidth and bathymetry. The remaining uncertainties, being from independent sources, can later be incorporated in the usual sum-of-squares fashion.

3. Comparison with benchmark mode calculations

An example result of this uncertainty estimation method is shown in Fig. 1. The TL calculation was done using an isotropic mode solution[6] for a Pekeris waveguide. This allowed for explicit

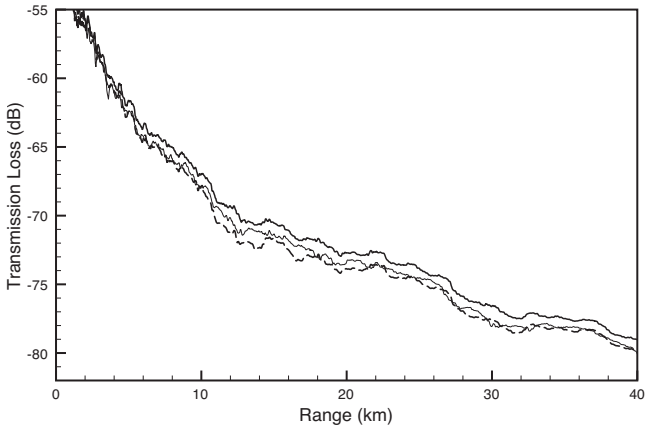


Fig. 2. Transmission loss, lower uncertainty estimate, and decimated mode calculation results (dashed line). The transmission loss and lower uncertainty estimate curves are from Fig. 1. The decimated mode result was generated using a modified version of the mode representation discussed in the text which simulated the effects of mode removal due to frequency uncertainties.

mode truncation in testing. A shallow water environment described in Ref. 7 was used for all test cases, in which the water was isovelocity at 1491 m/s, and was 150 m deep. Bottom properties were 1565.55 m/s sound speed, 1.268 g/cm^3 density, and 0.6 dB/λ attenuation. The source was 1 kHz; source and receiver depths were both 30.48 m. The simulated receiver bandwidth was set to one-third octave. Upper and lower uncertainty estimates based on this bandwidth are shown with the TL plot in Fig. 1.

The lower boundary estimate generated using this method was tested by direct comparison with a modified version of the mode model in which the number of contributing modes has been truncated by the factor β. These results are shown in Fig. 2, and are in agreement with the lower bound estimated using the present method. Unfortunately, there is no analogous direct test for the upper bound.

4. Further testing and examples

The only source of uncertainty discussed thus far has been frequency uncertainty, introduced by the receiver's bandwidth. With this source of uncertainty, this method may be further tested by running a propagation model across an ensemble of frequencies and averaging. Figure 3 shows the results of one such calculation, where TL predictions were made using 51 frequencies uniformly spread across the third-octave interval surrounding 1 kHz. These were averaged, and sampled at 1/3 nautical mile intervals in order to simulate ocean acoustic survey data collection and processing methods.[5] The predicted error bounds are consistent with the spread of the simulated data points. In a more manageable ensemble of ten frequencies over this same band, 72% of all 2500 computed TL points fell within the uncertainty predictions, indicating that 1 s.d. in frequency spread maps accurately into a meaningful error bound via this method. Simply counting the synthetic data points in Fig. 3 shows that 77±11% fall within the uncertainty band.

In another test, the previously used bandwidth uncertainty was replaced by a bathymetry uncertainty of 20 m as the source of mode count uncertainty. With the mode count estimate depending upon the water depth as well as source frequency, an ensemble average of 51 different depths gave a set of synthetic data points similar to that previously shown in Fig. 3. When compared with the uncertainty calculation, 69±10% of the synthetic data points fell between the uncertainty bounds. This demonstrates that depth variations can also be mapped into usable TL uncertainties, and that the results are consistent with an ensemble average made using the same depth variability.

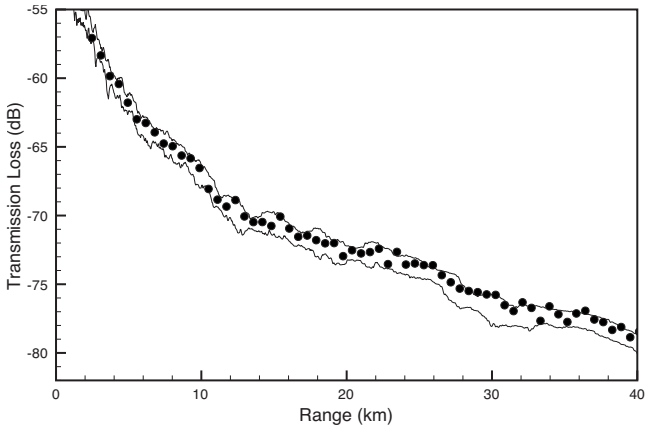


Fig. 3. Transmission loss bounds from Fig. 1 overlaid with simulated third-octave data at one-third nautical mile intervals. Again, receiver bandwidth is the only source of uncertainty included in these calculations.

5. Summary and conclusions

A method for mapping uncertainties in ocean environmental parameters and frequency bandwidth into a TL error bound prediction has been developed. The method makes conceptual use of the similarities between frequency- and range-averaging expressions in mode theory, bandwidth properties of acoustic receiver systems, and simple relations between environmental measurements and the number of available modes.

Direct comparison with isotropic mode calculations indicates that this method correctly maps uncertainty in mode count into TL lower-bound error predictions. Comparisons with synthetic data, in which frequency-averaged acoustic model results having a known bandwidth (i.e., frequency uncertainty) were used, show that the error bounds generated using this method are in agreement with the variability found in the synthetic data. Finally, uncertainties in bathymetry have been demonstrated to map into uncertainties in TL using the same method.

Because this method relies both upon the analytic aspects of acoustic propagation as well as the bandwidth properties of sensor systems, the results from this method can be applied either as confidence bounds on TL model predictions, or as error estimates on data. These are, in turn, useful in either displaying computational or experimental results, or in weighting data used in inversion computations.

Acknowledgment

This work was supported by Battlespace Awareness and Information Operations Program Office (PMW-120), Navy PEO C4I.

References and links

[1]P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969), pp. 2–8.
[2]P. Henrici, *Essentials of Numerical Analysis* (Wiley, New York, 1982), pp. 12–39.
[3]J. S. Hanna and P. V. Rost, "Parabolic equation calculations versus North Pacific measurements data," J. Acoust. Soc. Am. **70**, 504–515 (1981).
[4]C. H. Harrison and J. A. Harrison, "A simple relationship between frequency and range averages for broadband sonar," J. Acoust. Soc. Am. **97**, 1314–1317 (1995).
[5]D. W. Harvey, A. Lowrie, and R. D. Filipczyk, "The generation of *N*-layered geoacoustic descriptions of the subbottom using NAVCEANO-collected experimental data," in *Proceedings of the MTS/IEEE Oceans '02 Conference*, Biloxi, Mississippi, October 2002, pp. 358–362.
[6]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000), pp. 264–269.
[7]R. E. Keenan and H. Weinberg, "Gaussian ray bundle (GRAB) model shallow water acoustic workshop implementation," J. Comput. Acoust. **9**, 133–148 (2001).

Informational masking increases the costs of monitoring multiple channels

Gerald Kidd, Jr.
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University, Boston, Massachusetts 02215
gkidd@bu.edu

Virginia M. Richards
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104
richards@psych.uppen.edu

Christine R. Mason
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University, Boston, Massachusetts 02215
cmason@bu.edu

Frederick J. Gallun
National Center for Rehabilitative Auditory Research, Portland VA Medical Center, Portland, Oregon 97239
frederick.gallun@va.gov

Rong Huang
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104
rongh@sas.upenn.edu

Abstract: This study examined the costs of simultaneously monitoring two frequency regions. Listeners detected low- and high-frequency tones in a 2I4AFC procedure. On every trial, each signal was presented in either the first or second interval independently. Comparison of thresholds in single- and dual-signal conditions provided an estimate of the costs. Thresholds were obtained in quiet, in notched-filtered noise, and in randomized multitone maskers. No cost was found in quiet, whereas large costs were found for the masked conditions, especially for the multitone masker. These results suggest that costs of dividing attention in frequency depend on both signal and non-signal channels.

PACS numbers: 43.66.Dc, 43.66.Ba, 43.66.Lj [QJF]
Date Received: May 5, 2008 **Date Accepted:** July 17, 2008

1. Introduction

If the task of a listener is to detect the presence of a pure-tone signal in a background of noise, uncertainty about the frequency of the signal elevates detection threshold relative to the threshold that is found when the frequency of the signal is known. However, the elevation in threshold due to signal frequency uncertainty is relatively small, only about 3 dB for a wide range of possible frequencies (e.g., Green, 1961). In fact, attempting to explain the small effect of signal frequency uncertainty was one of the early problems addressed by researchers applying signal detection theory to auditory psychophysics (e.g., Tanner *et al.*, 1956; Veniar, 1958; Creelman, 1960; Green, 1961; Sorkin *et al.*, 1973). The basic problem was that the variance of the decision statistic computed from monitoring multiple frequency channels simultaneously, or from shifting the center frequency of a single tunable filter across a wide range of frequencies, appeared to be too great to account for the empirical results. Later work (Buus *et al.*, 1986) demonstrated that the apparent discrepancy was less than originally considered. However, taken as a whole,

the conclusion that is drawn from all of these studies is that the *cost*[1] of monitoring multiple frequency channels is relatively low.

The various attempts at modeling the effects of signal frequency uncertainty have in common the underlying assumption that the costs associated with monitoring different frequency regions arise solely from the channels containing, or potentially containing, signal energy. This assumption makes intuitive sense because frequency channels that never include the signal contain no information useful in solving the task and, as such, an ideal observer would not consider them in forming a decision rule. Human listeners, however, often use decision strategies that are less than ideal. The phenomenon of informational masking, in which masker energy remote in frequency from the signal energy exerts a significant effect on performance, is one such example. Neff and Green (1987) were the first to demonstrate that a masker composed of randomly drawn sinusoids having a different composition on every presentation produced large amounts of masking that could not be attributed to peripheral overlap of excitation. Since that initial report many investigators have used the multitone masking paradigm to examine central influences in auditory masking (see the review by Kidd *et al.*, 2008).

Although it is clear that large amounts of informational masking may be produced by highly uncertain multitone maskers, it is not known whether there is an interaction between informational masking and the costs of signal frequency uncertainty. Such an interaction seems plausible because the decrease in performance associated with signal frequency uncertainty, and the informational masking caused by masker uncertainty and/or similarity, both presumably reflect limitations on higher-level processes. This is in contrast to energetic masking, which is assumed to be due primarily to overlap of excitation in the auditory periphery. In the present study, we attempt to address this issue by assessing the costs of signal frequency uncertainty under three levels of masker uncertainty: no masker (quiet, no masker uncertainty), Gaussian noise masker (low masker uncertainty; intended to produce primarily energetic masking), and randomized multitone masker varying in composition on every stimulus presentation (high uncertainty; intended to produce large amounts of informational masking).

2. Methods

2.1 Listeners

A total of six young adult college students served as paid participants in these experiments. The subjects had normal hearing as determined by routine audiometry.

2.2 Stimuli

The stimuli were computer-generated tone and noise bursts that were played through a 16 bit Digital-to-Analog Converter (DAC) and programmable analog attenuators, mixers, low-pass filters (7.5 kHz), and headphone buffers (Tucker-Davis Technologies, Aluchua, FL).

The frequencies of the signal tones were 558 and 1791 Hz. The duration of the masker was 205 ms. In all cases except one the signal was also 205 ms and presented synchronously with the masker. In the one exception described in the following, the onset of a 195 ms signal was delayed by 10 ms relative to that of the masker. All stimuli were gated by applying 5 ms cosine-squared ramps and the durations specified are total durations including ramps and steady-state portions.

There were two types of maskers: double-notched-filtered Gaussian noise and random-frequency multitone complexes. The Gaussian noise masker was generated digitally from 200 to 5000 Hz and had spectral notches centered geometrically on the two signal frequencies each spanning a width of 28% of the center frequency. The noise was first generated in the frequency domain and then inverse Fourier transformed producing extremely steep filter slopes. The level of the noise was held constant at 50 dB SPL (about 14 dB/Hz within the passbands). The random-frequency multitone maskers were drawn from the same frequency range as the noise bands with "protected regions" (i.e., in which masker components were not allowed) equal in width to the notches in the noise masker. The number of masker tones varied across intervals such that the total number of tones—signal tone(s) plus masker tones—was

constant at eight. Thus, for example, if both signal tones were in interval 1, six randomly chosen masker tones were presented in interval 1, while eight randomly chosen masker tones were presented in interval 2. If there was one signal tone in each interval, seven masker tones were presented in each interval. The purpose was to restrict judgments based purely on the number of tones present. There was always a different random draw of masker frequencies on each interval. The level per masker component was held constant at 41 dB SPL. There were two presentation conditions for the multitone masker. In one case, the 205 ms signal and masker were gated on and off together and that condition is referred to as "synchronous." The other condition, in which the 195 ms signal was delayed by 10 ms from the 205 ms multitone masker onset, is designated "asynchronous." The asynchronous condition was intended to be intermediate between the notched-noise and the synchronous multitone masker (i.e., the asynchrony was intended to release some of the expected informational masking).

2.3 Procedures

Listening was monaural with stimuli presented to the subject's preferred ear through a calibrated TDH-50 earphone. The listeners were tested individually in double-walled sound attenuating booths.

Detection thresholds were obtained using a 3-down 1-up adaptive procedure that estimates the 79.8% correct point on the psychometric function (Levitt, 1971). The main variables in the experiment were whether the listener's task was to detect one or two signals on each trial and the presence/type of masker that was used. In dual-signal conditions the listener made two separate detection judgments, one corresponding to each signal and two separate tracks provided a threshold estimate for each. The four alternatives (signal frequency by presentation interval) were represented on the response keys so only a single response was required to indicate the listener's choice. For all listeners, masked thresholds were obtained for both single-signal and dual-signal conditions in quiet (no masker) and in the presence of double-notched-filtered Gaussian noise and random-frequency multitone maskers. A minimum of 15 blocks of trials was obtained in each condition (with an average of 20 blocks obtained across listeners and conditions) in mixed and pseudorandom order with threshold estimates computed from the last 10 blocks in each condition. Each block consisted of 60 trials, with a starting step size of 4 dB, reduced to 2 dB after the third reversal. The average of the reversals obtained was computed after discarding the initial three or four reversals depending on whichever produced an even number. The mean number of reversals obtained across all conditions was 10.2.

3. Results

The group mean results from all conditions are plotted in Fig. 1. First, all of the group-mean quiet signal thresholds—for both frequencies in single- and dual-signal conditions—fell within a very narrow range from 5.6 to 7 dB SPL. As may be seen in Fig. 1, the various maskers raised thresholds considerably for the single-tone conditions (compare open symbols in the left-hand panel with corresponding symbols in the right three panels). There was more masking (the difference in threshold between the quiet single-tone threshold and the masked single-tone threshold at each frequency) for the multitone conditions than for the corresponding noise masked conditions despite the fact that both were presented at roughly the same SPL, have similar "protected regions" that exclude masker components, and have the same distribution of expected level per masker component outside of the protected regions. As is apparent from the error bars in Fig. 1 the intersubject variability in the amount of masking is substantially greater for the informational masker than for the energetic masker, a finding that is common in the masking literature (cf. Kidd *et al.*, 2008). Note that the asynchronous multitone masker produced only a few decibels less masking than the synchronous masker. This will be discussed further in the following. Across masker conditions, the low- and high-frequency thresholds are very similar for the quiet and noise conditions, but are somewhat higher for the high-frequency tone than the low-frequency tone in the two multitone maskers.

A three-way within-subjects analysis of variance was conducted on the results with the main factors being masker condition (none, noise, asynchronous multitone, and synchronous

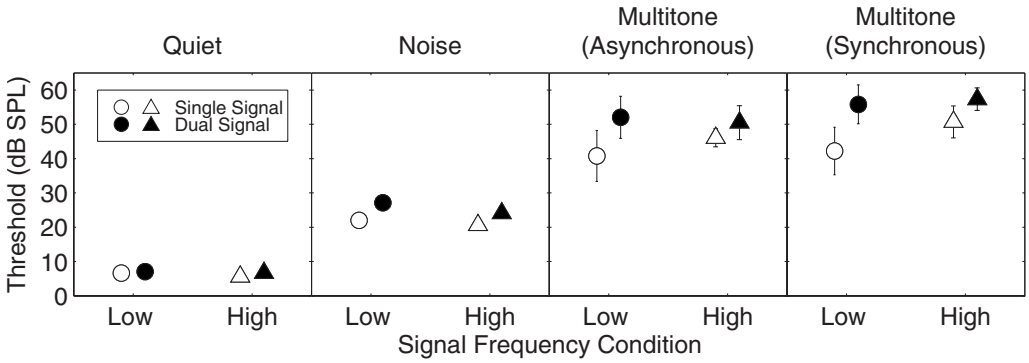


Fig. 1. Group mean thresholds (dB SPL), and intersubject standard errors of the means, for all conditions. The different panels are for the various quiet and masked conditions (quiet, noise, asynchronous and synchronous, left to right). Within each panel, thresholds for low- and high-frequency signals are plotted as circles and triangles, respectively. Open symbols represent results from single-signal conditions and closed symbols represent results from dual-signal conditions.

multitone), number of signals (single or dual), and signal frequency (low or high). Masker condition [$F(3,15)=50.8, p<0.001$] and number of signals [$F(1,5)=34.4, p<0.002$] were both significant, whereas signal frequency [$F(1,5)=0.2, p<0.66$] was not. The interaction of masker and number of signals was significant [$F(3,15)=4.1$, $p<0.03$] as was the number of signals with signal frequency [$F(3,15)=4.1$, $p<0.03$], whereas masker condition by signal frequency was not. The three-way interaction was also not significant.

The costs associated with the performance in dual-signal conditions may be estimated directly by taking the differences between single-signal and dual-signal thresholds. Those cost estimates are displayed in Fig. 2 as group mean differences and standard errors of the mean differences. For the quiet condition, the costs of dual-signal presentation were negligible, on average 0.4 and 1.1 dB for low and high signal frequencies, respectively. The dual-signal costs increased for the noise masker, with group mean differences of 5.1 dB for the low-frequency signal and 3.5 dB for the high-frequency signal. These values are quite consistent with those in the literature, especially considering that an identification judgment was required (i.e., high versus low signal frequency; cf. Green, 1961). For the multitone masker, the dual-signal costs were 11.2 and 4.4 dB for the low- and high-frequency tones, respectively, in the asynchronous condition and 13.6 and 6.6 dB in the synchronous condition. The apparently greater costs for the low-frequency signal in masked conditions have not, to our knowledge, been reported previously. There is no obvious reason for this trend, which deserves further study.

4. Discussion

The results of the current experiment demonstrate that substantial costs of monitoring multiple frequency channels may be observed under some conditions. These large effects—up to about

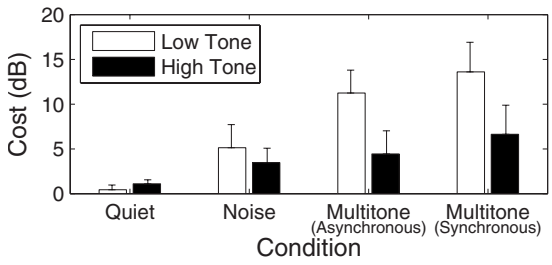


Fig. 2. Group mean differences in thresholds between single-signal and dual-signal tasks (dual-signal costs). The error bars are standard errors of the mean differences.

14 dB—are consistent with related work using uncertain frequency targets embedded in randomized multitone maskers reported by Richards and Neff (2004), although estimates of costs were not directly obtained in that study. In contrast to the relatively small elevation in threshold typically reported for detecting a tone in Gaussian noise when there is trial-by-trial a uncertainty about the signal frequency (roughly 3–5 dB), the same signal frequency uncertainty often produced significantly greater elevations in threshold for a random-frequency multitone masker. And, somewhat unexpectedly, this same degree of signal frequency uncertainty produced almost no elevation in threshold when there was no masker present. Although the asynchronous multitone masker condition was included with the expectation that it would produce substantially less informational masking, this was not the case. The small effect of delaying the signal relative to the masker was somewhat surprising given our informal listening while deciding on signal parameters and past studies using onset asynchrony to reduce informational masking. Other studies (e.g., Neff, 1995; Durlach *et al.*, 2005; Leibold and Neff, 2007) reported release from informational masking due to onset asynchrony that was much larger than that found here. The smaller effect in the current study is probably due to the briefer delay used (10 ms), which apparently was not sufficient to perceptually segregate the signal from the masker very effectively for these listeners.

One alternative explanation for the current findings is that the dual-signal costs increase simply as the amount of masking increases without regard to the presence/absence or type of masker. This seems to be highly unlikely given the fairly large body of past work on uncertain signal frequency detection, which has typically used noise levels producing much more masking than our double-notched noise did yet the costs were about the same. For example, Green (1961) used a 40 dB spectrum level noise with threshold values around E/N_0 of 10–20 dB (signal duration was a variable), which, as noted in the text, compare favorably to our results albeit for significantly lesser amounts of masking. Furthermore, Creelman (1960) measured the effects of uncertain frequency signal detection for several levels of the background noise. His results, which were given in percent correct performance, also fail to show any systematic change in costs with noise level (and therefore amount of masking). Thus, this does not appear to be a concern in the context of the current study.

The current results pose significant problems for models that attempt to account for the detrimental effects of detecting a tone of uncertain, as compared to certain, frequency. First of all, as discussed in Sec. 1, a common characteristic of these models is that the variance in the decision statistic, which accounts for the elevation in threshold due to monitoring multiple channels, only takes into account the channels that could potentially contain the target. The findings of this study are incompatible with this fundamental assumption underlying such models. In all three cases tested here: no masker, double-notch-filtered noise, and multitone masker, the increase in the number of potential signal channels upon which costs were computed was the same; i.e., from one channel to two. However, the range of costs actually observed across these conditions was more than 13 dB and was closely tied to the specific condition tested. Very small costs (1 dB or less) were observed when no masker was present. This finding was unexpected based on the common assumption that quiet threshold is determined by internal noise that is considered to be largely independent across channels. If this were true the costs of increasing the number of channels should be very much the same as occurs for actual (externally applied) noise. Historically, and also in the noise condition tested here, the costs are about 3–5 dB. Furthermore, and even more difficult to explain, the very large costs (up to about 14 dB) found for the multitone masker means that there is an interaction between the number of signal channels that must be monitored and the type of masking (i.e., energetic versus informational) produced by the masker.

Informational masking, at least for the task of detecting tones in a multitone masker, usually is thought of as reflecting a failure of the listener to selectively attend to a single frequency channel and ignore or suppress the outputs of other irrelevant (masker) channels (cf. Kidd *et al.*, 2008). The finding here of much greater costs for the informational masker than for the energetic masker suggests that the ability to ignore irrelevant channels becomes much less effective when attention must be divided across two signal channels.[2] It is not clear how to

modify the existing models of signal frequency uncertainty to account for the variation in costs across masker conditions found in this study. Presumably, however, the decision rule must take into account events occurring in nonsignal channels as well as potential signal channels. The multitone masker obviously produces large fluctuations in the outputs of nonsignal frequency channels when considered across stimulus presentations. These fluctuations, or the varying perceptual salience of the individual masker tones or timbre of the complex as a whole, not only produce informational masking but may interact with signal frequency uncertainty to increase the costs of dividing attention across locations in frequency. Buus *et al.* (1986) summarized their findings for uncertain tone detection in noise by stating "… that listeners may not make an optimally weighted sum of independent stochastic variables from each channel, but may make only an unweighted sum across an optimally chosen set of channels." The current results may indicate that in some highly uncertain masking conditions listeners are prevented from choosing those channels optimally.

Another difficulty for models of uncertain signal frequency effects is that, in the design of this experiment, a "switching" strategy is available to the listener. In that case the listener could simply attend to one frequency channel in the first interval, make a single YES–NO detection decision, and then switch attention to the other frequency channel for the second interval and make a second YES–NO detection decision and then report the two decisions. This strategy should greatly reduce any two-channel monitoring costs. This is further evidence for less-than-ideal performance in these tasks. The listeners were fully informed about the structure of the experiment and the nature of the stimulus. Whether further training specifically intended to promote a channel-switching strategy would reduce the costs observed here is not currently known.

5. Summary

To summarize, when listeners were required to monitor and respond to events in two widely spaced frequency channels within each trial, detection thresholds were significantly higher than corresponding single-signal thresholds. However, the costs associated with dual-signal presentation were only observed in masked conditions and not in quiet listening. The magnitude of the effect depended upon the type of masker that was present, and consequently the type of masking that was produced, and was significantly higher when the masker was a random-frequency multitone complex than a notched-filtered noise.

Acknowledgments

This work was supported by AFOSR Award No. FA9550-05-1-2005 and by Grant No. DC004545 and DC02012 from NIH/NIDCD. F.G. was supported by F32 DC006526 from NIDCD.

References and links

[1]In the present context, we are using the word "cost" to reflect an increase in detection threshold due to signal frequency uncertainty. The word cost sometimes is used in different ways such as in experiments in which costs and benefits are manipulated directly as controlled experimental variables.

[2]An alternative possibility is that the difference in costs observed for the noise versus multitone maskers reflects differences in the slopes of the underlying psychometric functions for a constant change in d'. Psychometric functions were not obtained in this study so the issue must await future work for clarification.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). "Decision rules in detection of simple and complex tones," J. Acoust. Soc. Am. **80**, 1646–1657.

Creelman, C. D. (**1960**). "Detection of signals of uncertain frequency," J. Acoust. Soc. Am. **32**, 805–810.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G., Jr. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.

Green, D. M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008**). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science

+Business Media, New York), pp. 143–190.
Leibold, L., and Neff, D. L. (**2007**). "Effects of masker-spectral variability and masker fringes in children and adults," J. Acoust. Soc. Am. **121**, 3666–3676.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.
Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
Sorkin, R. D., Pohlmann, L. D., and Gilliom, J. D. (**1973**). "Simultaneous two channel signal detection. III. 630 and 1400 Hz signals," J. Acoust. Soc. Am. **53**, 1045–1050.
Tanner, W. P., Swets, J. A., and Green, D. M. (**1956**). "Some general properties of the hearing mechanism," Technical Rep. No. 30, Electronic Defense Group, University of Michigan, Ann Arbor, MI [cited in *Signal Detection Theory and Psychophysics*, D. M. Green and J. A. Swets (Kreiger, New York, **1974**)].
Veniar, F. A. (**1958**). "Signal detection as a function of frequency ensemble. I," J. Acoust. Soc. Am. **30**, 1020–1024.

Modeling of axisymmetric wave modes in a poroelastic cylinder using spectral method

Florian Karpfinger and Boris Gurevich[a)]
Department of Exploration Geophysics, Curtin University of Technology, GPO Box U1987, Perth, Western Australia 6845, Australia
florian.karpfinger@postgrad.curtin.edu.au, B.Gurevich@curtin.edu.au

Andrey Bakulin[b)]
WesternGeco, 10001 Richmond Ave., Houston, Texas 77042
abakulin@slb.com

Abstract: Algorithm and code are presented which solve the dispersion equation for cylindrical poroelastic structures. The algorithm is based on the spectral method, which discretizes the underlying wave equations with the help of spectral differentiation matrices and solves the corresponding equations as a generalized eigenvalue problem. The results are illustrated for the case of a fluid-saturated free cylinder with open- and closed-pore boundary conditions on its surface. The computed dispersion curves are in good agreement with analytical results, which confirms the accuracy of the method.

PACS numbers: 43.20.Mv, 43.20.Bi, 43.20.Jr, 43.20.Gp [RW]
Date Received: June 8, 2008 **Date Accepted:** July 17, 2008

1. Introduction

Modeling wave propagation modes in cylindrical structures is important for the understanding and quantitative interpretation of acoustic and seismic measurements in hydrocarbon wells, pipelines, as well as laboratory measurements. For porous materials these waves are affected by material permeability. These effects can be analyzed using Biot's equations of poroelasticity (Biot 1956a, b, 1962).

For a porous cylinder with open-pore boundary conditions on its surface this was first done by Gardner (1962), who derived the dispersion equation for extensional waves at low frequencies. For the full frequency range and closed boundary conditions the dispersion in a fluid-saturated cylinder was studied by Berryman (1983).

This was done for the first few modes because conventional *root finding* becomes challenging for poroelastic media. An alternative approach to modeling wave propagation in circular structures was recently introduced by Adamou and Craster (2004) based on the spectral method.

Karpfinger *et al.* (2008) extended the spectral method to axisymmetric waves for arbitrary fluid and solid layers. In this letter we extend the spectral method to cylindrical poroelastic structures.

2. Biot's thory of poroelasticity

Waves propagating in a poroelastic medium satisfy the following equations (Biot, 1962):

$$\nabla \cdot \boldsymbol{\sigma} = -\omega^2(\rho \mathbf{u} + \rho_f \mathbf{w}), \tag{1}$$

[a)]Also at CSIRO Petroleum, ARRC, 26 Dick Perry Ave., Kensington, Perth, Western Australia 6151, Australia.
[b)]Present address: Shell International Exploration & Production Inc., Houston, Texas.

$$\nabla p = \omega^2(\rho_f \mathbf{u} + q\mathbf{w}), \tag{2}$$

where the field variables are the solid $\mathbf{u}$ and the average solid–fluid displacement $\mathbf{w}=\phi(\mathbf{U}-\mathbf{u})$, ϕ is the porosity, and $\mathbf{U}$ is the fluid displacement. ρ and ρ_f are the densities of the porous material and the fluid, respectively. The frequency-dependent density term $q(\omega)$ is responsible for the viscous and inertial coupling between the solid and fluid phases (Biot, 1956b). The total stress tensor $\boldsymbol{\sigma}$ and the fluid pressure p are related to the field vectors by the constitutive equations

$$\boldsymbol{\sigma} = [(H-2\mu)\,\nabla\cdot\mathbf{u} + \alpha M\,\nabla\cdot\mathbf{w}]\mathbf{I} + \mu[\nabla\mathbf{u} + \nabla\mathbf{u}^T], \tag{3}$$

$$p = -M\,\nabla\cdot\mathbf{w} - \alpha M\,\nabla\cdot\mathbf{u}. \tag{4}$$

The stress–strain relations, Eqs. (3) and (4), are defined in terms of the shear modulus of the skeleton μ and two poroelastic material constants M and H which are related to the bulk moduli of the fluid K_f, the grain material K_g, and the skeleton K. This relation is given by the Gassmann equations (Gassmann, 1951).

Substituting the constitutive relations, Eqs. (3) and (4), into Eqs. (1) and (2), we obtain two coupled wave equations. In order to decouple these equations into separate equations for the three bulk waves the displacements can be written in terms of potentials.

For time-harmonic oscillations this yields Helmholtz equations for the shear waves

$$(\nabla^2 + k_s^2)\boldsymbol{\beta} = 0, \quad \boldsymbol{\chi} = -\frac{\rho_f}{q}\boldsymbol{\beta}, \tag{5}$$

and for the compressional waves

$$(\nabla^2 + k_\pm^2)A_\pm = 0. \tag{6}$$

Here $A_\pm$ is a linear combination of the scalar potentials Υ and ψ. $\boldsymbol{\beta}$ and $\boldsymbol{\chi}$ are vector potentials, k_s is the wave number of the shear wave, and k_+ and k_- are the wave numbers of the fast and slow longitudinal waves, respectively. These wave numbers give the dispersion in an infinite poroelastic medium.

In order to distinguish between the torsional and extensional components of the wave field, the potential $\boldsymbol{\beta}$ is further decomposed as $\boldsymbol{\beta}=\hat{\mathbf{z}}\beta_1+\nabla\times(\hat{\mathbf{z}}\beta_2)$, where $\hat{\mathbf{z}}$ is the unit vector in z direction. For axisymmetric modes only β_2 is nonzero. Due to the symmetry of the system, Eqs. (3)–(6), expressed in cylindrical coordinates, describe the axisymmetric wave propagation in a poroelastic cylinder.

3. The Poroelastic eigenvalue problem

With the help of the *root-finding* technique, the velocity dispersion for a porous cylinder can be obtained, but can be very difficult to compute for a poroelastic medium (as it requires finding numerous complex roots of a cumbersome dispersion equation involving Bessel functions).

The spectral method bypasses these difficulties and solves the underlying Helmholtz equations numerically. For elastic wave propagation this was first implemented by Adamou and Craster (2004), who investigated circumferential waves in an elastic annulus. The problem is solved by numerical interpolation using spectral differentiation matrices (DM's). This approach is explained in detail for axisymmetric waves in cylindrical structures in Karpfinger *et al.* (2008).

To apply the spectral method to the cylindrical poroelastic problem, Eqs. (5) and (6) can be presented as the following eigenvalue problem

$$\underbrace{\begin{pmatrix} L_+ & 0 & 0 \\ 0 & L_{v_s} & 0 \\ 0 & 0 & L_- \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} A_+ \\ \beta_2 \\ A_- \end{pmatrix}}_{\Theta} = k_z^2 \begin{pmatrix} A_+ \\ \beta_2 \\ A_- \end{pmatrix}. \tag{7}$$

The differential operators $L_\pm$ and L_s are discretized using DMs (Karpfinger *et al.*, 2008) and combined in a $3N \times 3N$ matrix. N is the number of Chebyshev points discretizing the cylinder along its radius.

In order to solve Eq. (7), appropriate boundary conditions have to be incorporated (Deresiewicz and Skalak, 1963). For a porous-vacuum interface two scenarios can be considered: open- or closed-pore boundary conditions. Open-pore boundary conditions on the free surface of the cylinder $r=a$ are

$$\sigma_{rr}|_{r=a} = \sigma_{rz}|_{r=a} = -p_f|_{r=a} = 0. \tag{8}$$

In addition to the zero stress, the fluid pressure is zero on the surface and fluid can thus flow across the interface. For a closed surface, the boundary conditions become

$$\sigma_{rr}|_{r=a} = \sigma_{rz}|_{r=a} = 0, \quad w_r|_{r=a} = 0. \tag{9}$$

The pore pressure condition is replaced by the condition that relative motion of fluid with respect to solid is zero on the surface of the cylinder.

Such boundary conditions thus have to be set within Eq. (7). In order to achieve this, the stress components σ_{rr}, σ_{rz}, and $-p$ are grouped in a matrix of the same size

$$\begin{pmatrix} \sigma_{rr} \\ \sigma_{rz} \\ -p \end{pmatrix} = \underbrace{\begin{pmatrix} S_{rr}^+ & S_{rr}^s & S_{rr}^- \\ S_{rz}^+ & S_{rz}^s & S_{rz}^- \\ S_p^+ & S_p^s & S_p^- \end{pmatrix}}_{S} \cdot \begin{pmatrix} A_+ \\ \beta_2 \\ A_- \end{pmatrix}, \tag{10}$$

where the components of **S** are separated in terms of fast longitudinal wave, shear wave, and slow longitudinal wave.

A similar matrix can be introduced for the components of displacement

$$\begin{pmatrix} u_r \\ u_z \\ w_r \end{pmatrix} = \underbrace{\begin{pmatrix} T_{rr}^+ & T_{rr}^s & T_{rr}^- \\ T_{rz}^+ & T_{rz}^s & T_{rz}^- \\ T_p^+ & T_p^s & T_p^- \end{pmatrix}}_{T} \cdot \begin{pmatrix} A_+ \\ \beta_2 \\ A_- \end{pmatrix}, \tag{11}$$

where the components of the $3N \times 3N$ matrix **T** are arranged in analogy to **S**.

The last step is to combine these displacements and stress components with the equations of motion, in an appropriate way, in order to solve the problem for such boundary conditions. This means that the rows of **S** and/or **T** representing the points at the outer boundary (N, $2N$, and $3N$) replace the corresponding rows in the **L** matrix. The modified matrix is now denoted by $\tilde{\mathbf{L}}$. Boundary conditions are introduced by setting the corresponding rows to zero on the right-hand side. The eigenvalue problem can now be formulated in the form

$$\tilde{L}\Theta = k_z^2 Q \Theta, \tag{12}$$

where Q is a $3N \times 3N$ diagonal matrix with zeros elements in rows corresponding to the points on the surface and ones in all other diagonal elements. Equation (12) is a generalized eigenvalue

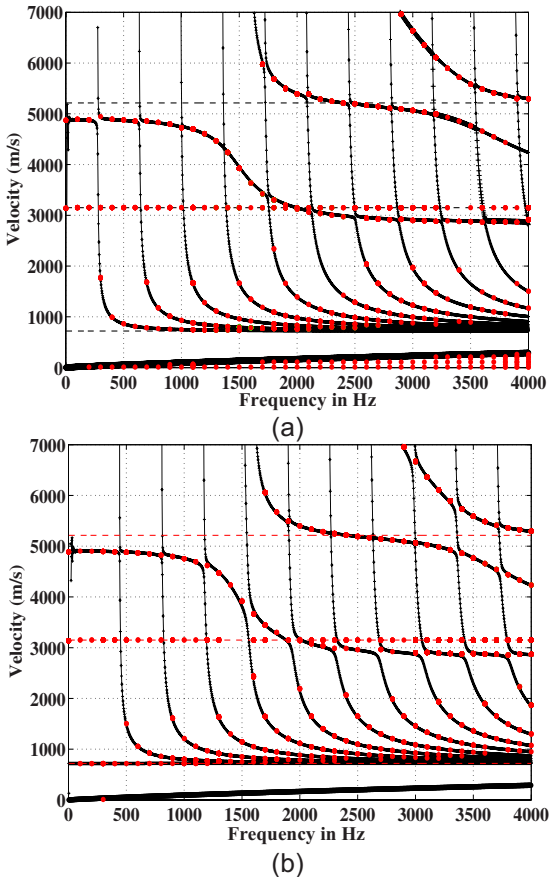


Fig. 1. (Color online) Dispersion curves for open-pore (a) and closed (b) boundary conditions; spectral method (positive signs); *root finding* (dots); and the dashes show explicitly the fast and slow *P* waves and the shear wave velocities.

problem, which means that we cannot find the inverse Q^{-1} as $\det(Q)=0$, but it can be solved with numerical eigenvalue routines.

4. Dispersion curves

Solution of Eq. (12) yields wave numbers k_z for a given frequency. Repeating this for various frequencies allows the computation of dispersion curves. To illustrate the results, in Figs. 1(a) and 1(b) dispersion curves are computed of a porous cylinder with open and closed boundary conditions, respectively. This is done for the case of high permeability which corresponds to the high frequency limit of Biot's theory where the slow *P* wave is propagatory. The parameters chosen are the same as in Berryman (1983).The permeability used in this study is $k_0=24.4$ darcy and the radius of the cylinder is 1 m. The positive sign (+) indicates the results computed with the spectral method. They are overlayed with the analytical solution obtained with *root-finding* method (dots). The dashed lines show the velocity of the fast *P* wave, the shear wave, and the slow *P* wave, respectively.

In both cases the dispersion curves show a very good agreement between the results for spectral method (+) and the *root-finding* approach. Due to a very high computational effort of *root finding*, only a few points were computed. The spectral method is about 1000 times more efficient.

In the low and high frequency limit the extensional mode velocity agrees with the values calculated by Berryman (1983) for both types of boundary conditions. The extensional mode propagates for low frequencies with 4910 m/s, while for high frequencies the limit is the Rayleigh velocity 2890 m/s.

Figure 1(a) shows singularities in the main extensional mode. Berryman (1983) observed these singularities and discussed their connection to Biot's slow wave. Figure 1(a) shows that these singularities are connected to higher modes of Biot's slow wave. These higher modes propagate, for frequencies higher than the first cut-off frequency (~270 Hz), with the Biot slow wave velocity. At the cut-off frequencies they go to infinity and cross over with the extensional mode where the interaction between the two modes causes the distortion of the dispersion curve of the extensional mode. The nature of the higher modes can be explained using the analogy with waves along a water column. In the high frequency limit of Biot's theory, the fluid phase can be considered as an independent fluid column. Since a fluid has zero Young's modulus, at zero frequency no slow extensional mode exists (Rubinow and Keller, 1971).

Overall, the result for closed-pore boundary conditions [Fig. 1(b)] looks similar to that of open-pore boundary conditions. The major difference is that now the slow extensional mode can be observed down to zero frequency. This is in agreement with Feng and Johnson (1983), who showed that the extensional Biot slow wave becomes a propagating mode for a closed interface. The fast extensional mode is again distorted by the higher modes of the Biot slow wave at the cut-off frequencies. The cut-off frequencies are shifted toward higher frequencies. Such modes can be explained in the same way as for open boundary conditions. Due to the closed boundary conditions, the fluid phase is analogous to the fluid in a closed solid tube (Rubinow and Keller, 1971). Thus a tube wave exists down to zero frequency. For both boundary conditions the shear wave velocity is a root of the dispersion equation in agreement with Berryman (1983). This is not the case for the spectral method.

5. Conclusions

We conclude that the spectral method provides an efficient and precise tool for the analysis of wave dispersion in poroelastic cylindrical structures. The results of the spectral method are in very good agreement with the analytical results. The advantage of this approach is that, in contrast to traditional methods, it is easier to implement and faster by a factor of approximately 1000. For multilayered structures the advantages of the spectral method will be even more significant. The results for a single poroelastic cylinder show that the singularities of the dispersion curve for the fast extensional mode correspond to intersections with the high order slow extensional modes.

Acknowledgments

The authors are grateful to Boris Kashtan (St. Petersburg State University, Russia), who suggested the idea of applying the spectral method to the problem at hand, Richard Craster (Imperial College, London) for helpful advice at the initial stage of the project, and Jim Berryman (Lawrence Berkeley Laboratory, CA) for helpful comments. The authors thank Shell Int. E & P for financial support.

References and links

Adamou, A. T. I., and Craster, R., V. (**2004**). "Spectral methods for modelling guided waves in elastic media," J. Acoust. Soc. Am. **116**, 1524–1535.

Berryman, J. G. (**1983**). "Dispersion of extensional waves in fluid-saturated porous cylinders at ultrasonic frequencies," J. Acoust. Soc. Am. **74**, 1805–1812.

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

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

Biot, M. A. (**1962**). "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 1254–1264.

Deresiewicz, H., and Skalak, R. (**1963**). "On uniqueness in dynamic poroelasicity," Bull. Seismol. Soc. Am. **53**, 783–788.

Feng, S., and Johnson, D. L. (**1983**). "High-frequency acoustic properties of a fluid/porous solid interface. i. New surface mode," J. Acoust. Soc. Am. **74**, 906–914.
Gardner, G. H. F. (**1962**). "Extensional waves in fluid-saturated porous cylinders," J. Acoust. Soc. Am. **34**, 36–39.
Gassmann, F. (**1951**). "Über die elastizität poröser medien," Vierteljahrsschr. Natforsch. Ges. Zur. **96**, 1–23.
Karpfinger, F., Gurevich, B., and Bakulin, A. (**2008**). "Computation of wave propagation along cylindrical structures using the spectral method," J. Acoust. Soc. Am. **124**(2), 859–865.
Rubinow, S. I., and Keller, J. B. (**1971**). "Wave propagation in a fluid-filled tube," J. Acoust. Soc. Am. **50**, 198–223.

Detection in reverberation using space time adaptive prewhiteners

Wei Li,[1,2] Xiaochuan Ma,[1] Yun Zhu,[1] Jun Yang,[1,2] and Chaohuan Hou[1]
[1]*Institute of Acoustics, Chinese Academy of Sciences*
[2]*Graduate University of Chinese Academy of Sciences, NO.21, Bei-Si-huan-Xi Road, Beijing, China*
liwei@mail.ioa.ac.cn, maxc@mail.ioa.ac.cn, zhuyun@mail.ioa.ac.cn, yangjun@mail.ioa.ac.cn, hch@mail.ioa.ac.cn

Abstract: A major problem in moving platform active sonar systems is the detection of targets in spatially distributed and Doppler-spread reverberation. This paper presents a novel space time adaptive prewhitener for reverberation based on a two-dimensional autoregressive model. The space time adaptive prewhitener jointly processes received data in angle and Doppler to improve the separation of a target from reverberation. The detector using the space time adaptive prewhitener is shown to yield better detection performance than previously known schemes when operating in a reverberation background containing target echoes.

PACS numbers: 43.30.Vh, 43.60.Gk, 43.60.Jn [WC]
Date Received: May 1, 2008 **Date Accepted:** June 9, 2008

1. Introduction

Detecting moving targets by active sonar is heavily degraded by reverberation. Reverberation is strongly correlated with the signal, and it is doubly spread[1] as well as nonstationary. Conventional methods such as matched filtering are inefficient because of these difficulties. A suboptimal solution given by Kay *et al.* is the autoregressive (AR) prewhitener[2,3] which is based on the assumption of temporally local stationarity. The reverberation in the current block is whitened via modeling the reverberation in the previous block. They have managed to deal with Doppler-spread character of reverberation; however, the spatial-spread character of reverberation is neglected. To cope with this problem, space time processing is necessitated. In this paper, we propose a novel space time adaptive prewhitener (STAPW) based on a two-dimensional autoregressive (2D-AR) algorithm and the detector via STAPW is derived. The model of STAPW we present is similar to the space time adaptive process (STAP) signal model, which makes the problem simple and easy for further development with STAP. The STAPW is also based on the assumption of temporally local stationarity of reverberation. The spatially local stationarity is not considered since the number of sensors in a sonar system is limited thus data partition only in time domain is more efficient. The performance of the new scheme is evaluated and compared with one-dimensional (1D)-AR detectors by simulating a set of reverberation data containing targets and shows better results.

2. Signals and noise models

Consider a linear array of M sensors with equally inter-element spacing d. The signal emitted by the active sonar is assumed to be a continuous wave (cw). If we adopt a STAP vector notation, the detection problem is written as

$$\begin{cases} H_0: \mathbf{x} = \mathbf{n} + \mathbf{r}, \\ H_1: \mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} + \mathbf{r}. \end{cases} \tag{1}$$

The received data $\mathbf{x}$ under H_1 is composed of a target component $\mathbf{s}$, a complex amplitude A, a reverberation component $\mathbf{r}$, and an uncorrelated noise component $\mathbf{n}$. The target is characterized by normalized Doppler frequency f_d, inter-element spacing d, direction β, and wavelength λ. We obtain $\mathbf{s} = \mathbf{d}_f(f_d) \otimes \mathbf{d}_s(\beta) = \mathbf{d}_{f,s}(f_d, \beta)$. In this formula, $\otimes$ denotes the Kronecker product, $\mathbf{d}_s(\beta)$

is the M-dimensional steering vector with elements $\exp(-j2\pi md\cos\beta/\lambda)$, $m=0,\ldots,M-1$, and $\mathbf{d}_f(f_d)$ is the N-dimensional Doppler vector with elements $\exp(-2j\pi f_d n)$, $n=0,\ldots,N-1$.

3. Proposed detector

Consider we have $N^lM\times 1$ received data from a linear array, which is divided into several blocks in time domain. Assume there is $NM\times 1$ data in each block.

3.1 Whitening using 2D-AR modeling

The received data under H_0 is modeled as a two dimensional zero-mean complex Gaussian AR process:

$$\mathbf{y}=-\mathbf{Y}\mathbf{a}+\mathbf{e}, \tag{2}$$

where the $(pq-1)\times 1$ parameter vector $\mathbf{a}$ is written as

$$\mathbf{a}=[a_{01}\ldots a_{0,q-1}a_{10}\ldots a_{1,q-1}\ldots a_{p-l,0}\ldots a_{p-l,q-1}]'. \tag{3}$$

To obtain the expression of $\mathbf{y}$ and $\mathbf{Y}$ we rearrange the $NM\times 1$ data into a vector in a similar but reverse way as $\mathbf{x}$. Let $\mathbf{y}_t(m)=[x_t(m),x_t(m-1),\ldots,x_t(m-Q+1)]'$ where $x_t(m)$ is the received data from mth sensor at tth time. Then the $PQ\times 1$ vector $\mathbf{y}_t^m$ is defined as $\mathbf{y}_t^m=[\mathbf{y}_t'(m),\mathbf{y}_{t-1}'(m),\ldots,\mathbf{y}_{t-P+1}'(m)]'$ where $P=N-p+1$, and $Q=M-q+1$. The $PQ\times 1$ vector $\mathbf{y}$, is given by $\mathbf{y}=\mathbf{y}_t^m$. And the $PQ\times(pq-1)$ data matrix $\mathbf{Y}$ is written as

$$\mathbf{Y}=[\mathbf{y}_t^{m-1}\cdots\mathbf{y}_t^{m-q+1}\mathbf{y}_{t-1}^m\cdots\mathbf{y}_{t-1}^{m-q-1}\cdots\mathbf{y}_{t-p+1}^m\cdots\mathbf{y}_{t-p+1}^{m-q+1}], \tag{4}$$

in each $NM\times 1$ block, $t=N-1$, $m=M-1$.

The least squares solution of Eq. (2) results in $\mathbf{Y}^*\mathbf{Y}\mathbf{a}=-\mathbf{Y}^*\mathbf{y}$ or $\mathbf{R}\mathbf{a}=-\mathbf{r}$. Alternatively, we can write it in the block matrix form as follows:

$$\begin{bmatrix}\tilde{\mathbf{R}}_{00} & \tilde{\mathbf{R}}_{01} & \cdots & \tilde{\mathbf{R}}_{0,p-1}\\ \tilde{\mathbf{R}}_{01}' & \mathbf{R}_{11} & \cdots & \mathbf{R}_{1,p-1}\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{\mathbf{R}}_{0,p-1}' & \mathbf{R}_{1,p-1}' & \cdots & \mathbf{R}_{p-1,p-1}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{a}}_0\\ \mathbf{a}_1\\ \vdots\\ \mathbf{a}_{p-1}\end{bmatrix}=-\begin{bmatrix}\tilde{\mathbf{r}}_0\\ \mathbf{r}_1\\ \vdots\\ \mathbf{r}_{p-1}\end{bmatrix}. \tag{5}$$

Eq. (5) can be solved by the algorithm proposed by Tummala[4] which is able to get a good result even when p and q are low. The notation $\tilde{}$ for matrices $\tilde{\mathbf{R}}_{0i}$, where $i=0,1,\ldots,p-1$ and for $\tilde{\mathbf{a}}_0$ and $\tilde{\mathbf{r}}_0$ indicates a reduction in the matrix and vector sizes. From Eq. (5), we get 2D-AR parameters $\mathbf{a}$ and σ^2. The prediction equation is given by $\hat{\mathbf{y}}=-\mathbf{Y}\mathbf{a}$. The STAPW result is written as

$$\mathbf{z}=\mathbf{y}_{i+1}-\hat{\mathbf{y}}_i, \tag{6}$$

where $\mathbf{y}_{i+1}$ indicates the data rearranged in block $i+1$, while $\hat{\mathbf{y}}_i$ indicates the prediction for block $i+1$ using block i. To whiten reverberation using the above-noted methods, temporally local stationarity is assumed. In this paper, we compute the covariance coefficients of the spectrum obtained by 2D-AR parameters in block 1 and the spectrum by fast Fourier transform in all blocks in Fig. 1(a) using real data. The covariance coefficients are high at the beginning, then decrease gradually except a lowest point in block 18 where a target echo presents. Results show temporally local stationarity assumption is a reasonable theory. The data are from the signal emission to the end of receiving with a target echo embeded, thus can be taken as a whole detection duration for active sonar.

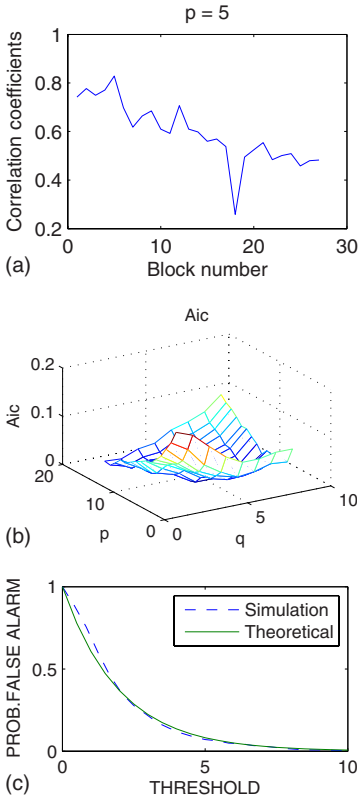


Fig. 1. (Color online) (a) Correlation of estimated spectrum and real spectrum. (b) Aic for 2D-AR model order selection. (c) False-alarm probability of proposed detectors using STAPW.8

3.2 Optimal detector

Then the detection problem after using STAPW is replaced by

$$\begin{cases} H_0: \mathbf{z} = \boldsymbol{\epsilon}, \\ H_1: \mathbf{z} = \mathbf{A}\mathbf{u}_\Theta + \boldsymbol{\epsilon}, \end{cases} \tag{7}$$

where $\boldsymbol{\epsilon}$ is complex white Gaussian noise of variance σ^2, and the subscript Θ stands for the unknown parameters that rule signal $\mathbf{u}_\Theta$: these are Doppler frequency f_d and the direction of the target β. For cw ping and a linear array, $\mathbf{u}_\Theta$ is written as

$$\mathbf{u}_\Theta = \mathbf{S}\mathbf{a}, \tag{8}$$

where $\mathbf{S}$ is defined by rearranging signal $\mathbf{s}$ in the same way as $\mathbf{Y}$. Let

$$\mathbf{d}_t^m = \tilde{\mathbf{d}}_{fP} \otimes \tilde{\mathbf{d}}_{sQ}, \tag{9}$$

where

$$\tilde{\mathbf{d}}_{sQ} = [e^{-j2\pi md \cos\beta/\lambda}, e^{-j2\pi(m-1)d\cos\beta/\lambda}, \ldots, e^{-j2\pi(m-Q+1)d\cos\beta/\lambda}]', \tag{10}$$

$$\tilde{\mathbf{d}}_{fP} = [e^{-2j\pi f_d n}, e^{-2j\pi f_d(n-1)}, \ldots, e^{-2j\pi f_d(n-P+1)}]'. \tag{11}$$

Hence $\mathbf{S}$ is given by

$$\mathbf{S} = [\mathbf{d}_t^{m-1} \ldots \mathbf{d}_t^{m-q+1} \mathbf{d}_{t-1}^{m} \ldots \cdot \mathbf{d}_{t-1}^{m-q-1} \ldots \mathbf{d}_{t-p+1}^{m} \ldots \mathbf{d}_{t-p+1}^{m-q+1}], \tag{12}$$

where $\mathbf{S}$ is also written as

$$\mathbf{S} = \mathbf{d}_t^m \mathbf{l}_{pq}, \tag{13}$$

in which $\mathbf{d}_t^m$ is a $PQ \times 1$ vector contained Θ and $\mathbf{l}_{pq}$ is a known $1 \times (pq-1)$ vector with complex unit elements.

Since σ^2 and A will need to be estimated, we base our detector on the generalized likelihood ratio test (GLRT) which can be shown to be a uniformly most powerful invariant test for large data records. Then for Eq. (7) the test statistic reads

$$L(\mathbf{z}) = \frac{\max_{A,\theta,\sigma^2} P_{\mathbf{z}/H_1}(\mathbf{z})}{\max_{\sigma^2} P_{\mathbf{z}/H_0}(\mathbf{z})}. \tag{14}$$

The noise conforms Gaussian distribution so that

$$P_{\mathbf{z}/H_0} = \frac{1}{\pi^{PQ}\sigma^{2PQ}} \exp\left(-\frac{\mathbf{z}^*\mathbf{z}}{\sigma^2}\right), \quad P_{\mathbf{z}/H_1} = \frac{1}{\pi^{PQ}\sigma^{2PQ}} \exp\left(-\frac{(\mathbf{z}-\mathbf{A}\mathbf{u}_\Theta)^*(\mathbf{z}-\mathbf{A}\mathbf{u}_\Theta)}{\sigma^2}\right). \tag{15}$$

The test statistic is $\mathcal{L} = \max_\Theta \mathcal{L}(\Theta)$, where

$$\mathcal{L}(\Theta) = 2PQ \log\left[\left(1 - \frac{|\mathbf{u}_\theta^*\mathbf{z}|^2}{(\mathbf{u}_\theta^*\mathbf{u}_\theta)(\mathbf{z}^*\mathbf{z})}\right)^{-1}\right]. \tag{16}$$

Substituting Eqs. (8)–(11) and (13) to Eq. (16) we get

$$\begin{aligned}\mathcal{L}(\Theta) &= 2PQ \log\left[\left(1 - \frac{|\mathbf{a}^*\mathbf{S}^*\mathbf{z}|^2}{(\mathbf{a}^*\mathbf{S}^*\mathbf{S}\mathbf{a})(\mathbf{z}^*\mathbf{z})}\right)^{-1}\right] \\ &= 2PQ \log\left[\left(1 - \frac{|\mathbf{a}^*\mathbf{l}_{qp}^*|^2|\mathbf{d}_{PQ}^*\mathbf{z}|^2}{(\mathbf{a}^*\mathbf{l}_{qp}^*\mathbf{l}_{qp}\mathbf{a})(\mathbf{d}_{PQ}^*\mathbf{d}_{PQ})(\mathbf{z}^*\mathbf{z})}\right)^{-1}\right] \\ &= 2PQ \log\left[\left(1 - \frac{|\mathbf{d}_{PQ}^*\mathbf{z}|^2}{PQ\mathbf{z}^*\mathbf{z}}\right)^{-1}\right].\end{aligned} \tag{17}$$

The test statistic is equivalent to $\mathcal{L}' = \max_\Theta \mathcal{L}'(\Theta)$. Thus, the GLRT decides H_1 if

$$\max_\Theta \mathcal{L}'(\Theta) = \max_\Theta \frac{2|\mathbf{d}_{PQ}^*\mathbf{z}|^2}{\mathbf{z}^*\mathbf{z}} > \eta. \tag{18}$$

We can easily get the theoretical probability false alarm $P_{\mathrm{FA}} = P(\mathcal{L}' > \eta | H_0) = \exp(-\eta/2)$. The curve by simulation is based on 500 Monte Carlo realizations. The results are shown in Fig. 1(c) for the number of sensors $M=10$, the number of time samples $N=10$ in each block, and $p=2$, $q=2$. In this simulation, we only use block 1 to calculate the 2D-AR parameters and block 2 to perform the detection. Theoretical and simulation results show good agreement.

The proposed detector takes the following steps to detect the targets: (1) Section the data into nonoverlapping blocks in time domain, and assume no target echo in block i; (2) Rearrange the data in block i, and estimate the 2D-AR filter parameters using block i; (3) Prewhiten the data in block $i+1$; (4) $\mathcal{L}' \leqslant \eta$: No target echo detects. Let $i=i+1$, and return to (3); (5) $\mathcal{L}' > \eta$: A target is detected. Stop.

4. Simulation results

The simulation of reverberation is described as follows. The complex envelope of the Doppler shifted reverberation data received at the mth sensor located at (x_m, y_m). Consider a narrowband, M element linear sonar array with a constant inter-sensor spacing d placed along the x direction with a velocity v. The delayed and Doppler shifted data received at the mth sensor at $(x_m, y_m)=((m-1)d, 0)$, $(x_0, y_0)=(0, 0)$ at time $t_n=\tau_0+nT$, is written as[5] $r_{mn}=\Sigma_{\theta_i\phi_l}\alpha(\theta_i, \phi_l)\exp(j2\pi \cos \phi_l(\sin \theta_i(m-1)d+2v \sin \theta_i nT)/\lambda)$, where $T=1/f_s$ is the sampling interval, azimuth $-\pi \leqslant \theta_i < \pi$, $1 \leqslant i \leqslant M_\theta$ and elevation angle $|\phi_l| \leqslant \phi_{\max}$, where $\phi_{\max}$ is the multipath elevation angle spread defined by the critical angle of the ocean acoustic channel. $\alpha(\theta_i, \phi_l)$ is the complex scatter amplitude from a reverberation patch at range $c\tau_0/2$, where c is the propagation speed of sound in water. The total number of reverberation patches $M_\theta N_\phi \gg MN$.

The 2D-AR model order is determined by Akaike information criterion (AIC) which as an extension of 1D-AR model case is given by[6] $AIC(k_1, k_2)=log(\hat{\sigma}^2)+2k_1k_2/(MN)$ in which $\hat{\sigma}^2$ is the prediction error variance. The results are shown in Fig. 1(b) for the number of sensors $M=64$, the number of time samples $N=20$ in each block. We choose the argument of minimum value of the AIC $p=5$, $q=3$ to correspond to the true order by this criterion. Suppose that there is an echo of a target in block 4. Let the Doppler frequency of the echo be $f_d=0.2$, and the direction of arrival (DOA) be $\beta=\pi/4$. The signal-to-noise ratio (SNR) in the simulation is -12.4 dB. Assume that the number of sensors $M=20$ and the number of time samples $N=64$ in each block. Detection is not performed in block 1 because there is no preknowledge for whitening.

The results of the proposed detector used in block 2,3,4 are shown in Figs. 2(a)–2(c).

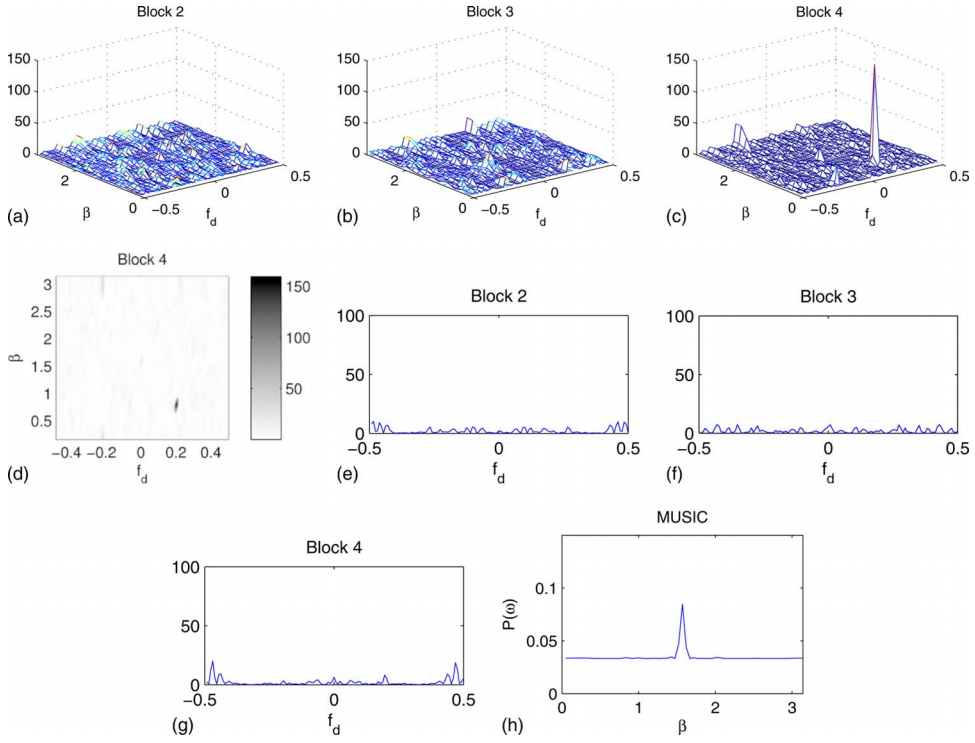


Fig. 2. (Color online) (a)–(c) Proposed detectors using STAPW. (d) Pseudocolor plot of the proposed detectors for Block 4. (e)–(g) Detectors using 1D-AR prewhiteners. (h) DOA estimation in Block 4 using conventional MUSIC algorithm.9

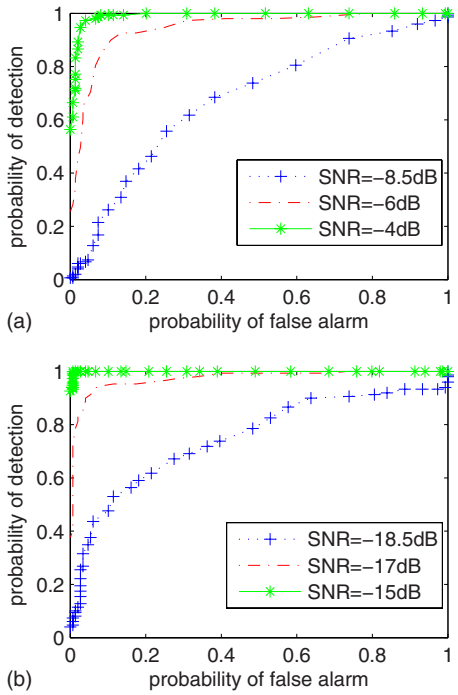


Fig. 3. (Color online) Experimental ROC curves (a) for the detectors using 1D-AR prewhiteners, (b) for the proposed detectors using STAPW.10

The target is easily detected in block 4. To see the estimation more clearly, we plot the proposed detectors in block 4 using pseudocolor plot. The estimated parameters at the maximum of the detector are $\hat{f}_d$=0.2 and $\hat{\beta}$=0.785 and are thus good estimates of the true parameters. Figs. 2(e)–2(g) are the detectors processed by 1D-AR prewhiteners using the same data but no target echo is detected. Since the detectors using 1D-AR prewhiteners cannot estimate the direction of the target, we use the conventional MUltiple SIgnal Classification (MUSIC) algorithm to estimate the direction of the target in block 4. Figure 2(h) shows the conventional MUSIC results which give false estimation $\hat{\beta}$=1.6. In this experiment, by prewhitening jointly in space and time domain via STAPW, good estimation of DOA and Doppler frequency is obtained when conventional methods fail due to the doubly spread reverberation.

The superiority of the proposed detection scheme is demonstrated from the experiments. However, to make this claim more precise, we evaluate the experimental performance of the detectors by receiver operating characteristic (ROC) curve where the detection rate is plotted versus the false alarm probability. Monte Carlo simulations were performed comprising 150 realizations with the signal present and equally many with reverberation and noise only. The ROC curves are shown in Fig. 3(a) for the detectors using 1D-AR prewhiteners and in Fig. 3(b) for the proposed detectors using STAPW. We vary the magnitude of the echo so that the input SNR is changed. Comparing Figs. 3(a) and 3(b), we see that for a fix couple (probability of detection, probability of false alarm) we have a gain of 10–11 dB by passing from 1D-AR to STAPW detectors.

5. Conclusion

In this paper, the conventional 1D-AR prewhitener for reverberation has been extended to 2D-AR as STAPW, which whitens not only the temporal spread but also spatial-spread reverberation. The detector is based on GLRT. Good results are achieved even with low SNR when

the 1D-AR detector fails in detection and when conventional DOA estimation loses efficiency in reverberation. The method would be tested for real data in further work.

References and links

[1]A. Jaffer, "Constrained partially adaptive space-time processing for clutter suppression," in *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, and Systems and Computer,* **1**, (Pacific Grove, California, October 31,-November 2), 671–676 (1994).

[2]S. Kay and J. Salisbury, "Improved active sonar detection using autoregressive prewhiteners," J. Acoust. Soc. Am. **87**, 1603–1611 (1990).

[3]V. rie Carmillet, P.-O. Amblard, and G. Jourdain, "Detection of phase- or frequency-modulated signals in reverberation noise," J. Acoust. Soc. Am. **105**, 3375–3389 (1999).

[4]M. Tummala, "New algorithm for solving block matrix equations with applications in 2-d ar spectral estimation," IEEE Trans. Signal Process. **39**, 759–764 (1991).

[5]V. Varadarajan and J. Krolic, "Array shape estimation and tracking using active sonar reverberation," IEEE Trans. Aerosp. Electron. Syst. **40**, 1073–1086 (2004).

[6]B. Aksasse and L. Radouane, "Two-dimensional autoregressive (2-d ar) model order estimation," IEEE Trans. Signal Process. **47**, 2072–2077 (1999).

Allis shad (*Alosa* alosa) exhibit an intensity-graded behavioral response when exposed to ultrasound

Maria Wilson
Department of Biological Sciences, University of Aarhus, Building 1131, C. F. Moellers Alle, 8000 Aarhus C, Denmark
maria.wilson@biology.au.dk

Marie-Laure Acolas
UMR INRA-Agrocampus Ecobiologie et Qualité des Hydrosystèmes Continentaux, 65 rue de Saint Brieuc, CS 84215, F-35042 Rennes Cedex, France
marie-laure.acolas@bordeaux.cemagref.fr

Marie-Laure Bégout
UMR 6217 CNRS, Ifremer, University de La Rochelle Place du Séminaire F-17137 L'Houmeau, France
marie-laure.begout@ifremer.fr

Peter T. Madsen
Department of Biological Sciences, University of Aarhus, Building 1131, C. F. Moellers Alle, 8000 Aarhus C, Denmark and Biology Department Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
peter.madsen@biology.au.dk

Magnus Wahlberg
Fjord & Bælt and University of Southern Denmark, Margrethes Plads 1, DK-5300 Kerteminde, Denmark
magnus@fjord-baelt.dk

Abstract: Most fish cannot hear frequencies above 3 kHz, but a few species belonging to the subfamily *Alosinae* (family Clupeidae) can detect intense ultrasound. The response of adult specimens of the European allis shad (*Alosa alosa*) to sinusoidal ultrasonic pulses at 70 and 120 kHz is tested. The fish showed an intensity-graded response to the ultrasonic pulses with a response threshold between 161 and 167 dB re 1 μPa (pp) for both frequencies. These response thresholds are similar to thresholds derived from juvenile American shad (*Alosa sapidissima*) in previous studies, supporting the suggestion that these members of *Alosinae* have evolved a dedicated ultrasound detector adapted to detect and respond to approaching echolocating toothed whales.

PACS numbers: 43.80.Lb [CM]
Date Received: April 5, 2008 **Date Accepted:** May 18, 2008

1. Introduction

Predator-prey interactions in the aquatic environment involve several sensory modalities such as vision, hearing, olfaction detection of water displacement with the lateral line system. These stimuli are used by predators to detect and track prey, and by the prey to detect and evade the approaching predators (Collin and Marshall, 2003). Toothed whales produce directional, ultrasonic clicks with sound pressure levels of more than 220 dB re 1 μPa (pp) to echolocate prey (Au 1993). The predation pressure from toothed whales can be intense (Santos *et al.*, 2001), and it is therefore conceivable that some prey species may have evolved sensory means to detect the powerful echolocation signals of toothed whales (Mann *et al.*, 2001), similarly to how some moths have evolved ultrasound hearing to detect echolocating bats (Miller and Surlykke, 2001).

Most fish cannot hear frequencies above a few kHz (Hawkins, 1981). However, some members of the subfamily *Alosinae* (family Clupeidae) have been shown to detect, respond to, and process intense ultrasonic signals (Nestler *et al.*, 1992; Mann *et al.*, 1997; Mann *et al.*, 2001; Plachta *et al.*, 2004; Gregory *et al.*, 2007). It has been suggested that this capability could be a counter move against echolocating toothed whales (Mann *et al.*, 2001). Studies of ultrasound detection abilities in Alosinae so far have focused on juvenile American species such as American shad (*Alosa sapidissima*) and gulf menhaden (*Brevoortia patronus*) (Mann *et al.* 2001), but it is not known if the capability to detect and respond to intense ultrasound is found across the entire subfamily.

One of the European members of the Alosinae, the allis shad, spawn in the rivers of the northern part of France and spend most of their life in the Bay of Biscay (Baglinière *et al.* 2003; Acolas *et al.* 2004), where a range of piscivorous toothed whales are found. However, stomach contents from the common dolphin (*Delphinus delphis*) show that even though the dolphins feed on a wide range of fish species, allis shad has not been identified as prey, despite temporal and spatial overlap of these two species (Pusineri *et al.*, 2007). Here we test the behavioral response of the anadromous allis shad (*Alosa alosa*) when exposed to ultrasonic signals and discuss implications for avoidance of echolocating toothed whales.

2. Materials and metals

Allis shad were caught in a live fish trap at INRA Station de piégeage, Le Moulin des Princes, Pont Scorff in Le Scorff River (Brittany, France) in May 2006. Twenty six adults of mixed sex with body lengths (nose to tail) between 45 and 55 cm were used. All fish were released back into the river after the experiment. Experiments were conducted at the place of capture in an outdoor test tank measuring $2.1 \times 2.1 \times 0.37$ m (length $\times$ width $\times$ depth) filled with water from the river at 13 °C.

Six groups consisting of two to five fish were exposed in two different playback sequences. Having several fish in each test group facilitated more natural shoaling behavior, but only data from the first responding fish in each group were used to avoid the possible bias of one fish evoking a change in the swimming behavior of the other fish in the tank. In the first sequence a 70 kHz signal was used (four groups) and in the second sequence a 120 kHz signal (two groups). The frequencies used are within the range of the centroid frequencies of echolocation clicks of toothed whales (Au 1993). Each playback sequence consisted of 12 stimulations at the two frequencies using received levels (±4 dB) at the fish of 157, 161, 167, 173, 179, 185 dB re 1 μPa (pp) with 5 min in between each exposure. The fish were exposed to the same intensity level twice. Half of the fish groups were exposed to an increasing followed by a decreasing series of intensities, and the other half to intensity steps in reversed order. This procedure made it possible to investigate if the thresholds for incrementing and decrementing exposure levels were different.

Ultrasonic pulses consisting of 50 000 cycles of sine waves were transmitted from an omnidirectional HS70 transducer (transmitting efficiency of 145 dB re 1 μPa/V at 1 m) for the 70 kHz pulse (pulse duration of 0.7 s) and an omnidirectional Brüel & Kjær 8105 transducer (transmitting efficiency of 145 dB re 1 μPa/V at 1 m) for the 120 kHz (pulse duration of 0.4 s). The transducer was placed in the middle of the tank and connected to a tone generator Agilent 33220A via a 46 dB custom-built power amplifier.

Measurement of the sound field in the test tank were performed with a calibrated Brüel & Kjær 8105 hydrophone. Signal analysis of the pulses using MATLAB 6.1 (*Mathworks*) showed that all significant energy was contained within 100 Hz around the center frequency. Furthermore, to test for the effect of low-frequency by-products of the high-frequency pulse each fish group was also exposed to a control sound stimuli consisting of a pure tone pulse at 2 kHz with 0.5 s duration and a sound pressure level of 160 dB re 1 μPa (pp) played with a UW30 transducer (transmitting efficiency of 110 dB re 1 μPa/V at 1 m). This stimulus should be 10 dB above the hearing threshold of allis shad, as estimated from hearing measurements made on

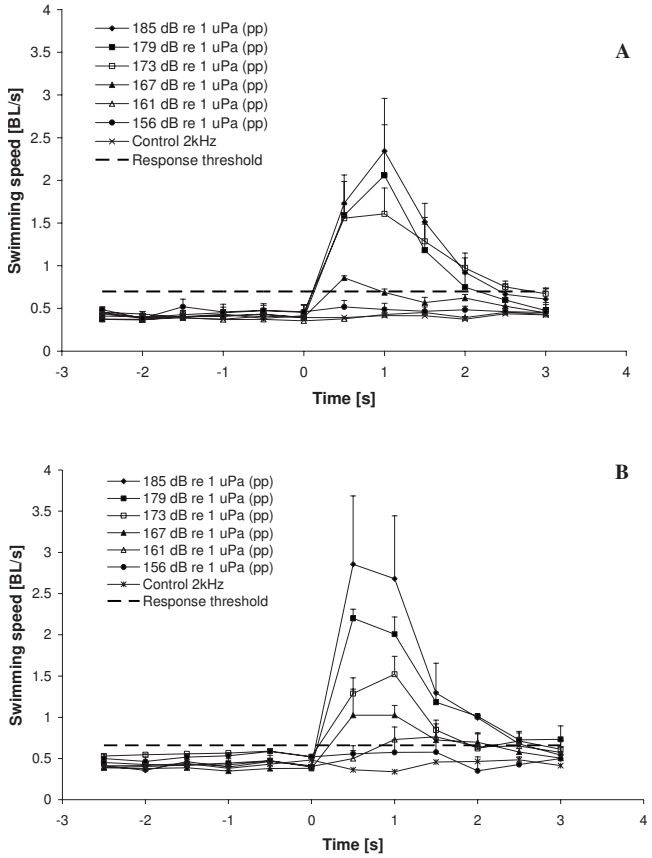


Fig. 1. Allis shad swimming speed (mean +S.E.) before, during, and after stimulation with ultrasound at two frequencies played at six different sound pressure levels. After each experiment the allis shads were exposed to a control sound at 2 kHz and 160 dB re 1 μPa pp. The allis shad are stimulated at the time 0 s. Swimming speed was measured with 30 s intervals. (A): Mean swimming speed of four fish when exposed to a 70 kHz tone. (B): Mean swimming speed of two fish when exposed to a 120 kHz tone.

other Alosa species (Popper *et al.*, 2004), and well below the spectral background noise in the tank ($<$135 dB re 1 μPa^2/Hz). The signal should therefore be audible to the fish.

The swimming behavior of the fish was recorded with a Profiline CTV7040 video camera (25 frames s^{-1}) mounted 1.5 m above the water surface of the tank. The camera images were digitized to a laptop via a Grabster 400 video card using the software Ulead VideoStudio7 (Ulead Systems Inc.). Each sound stimulation was accompanied by a cue given by the operator recorded by a microphone connected to the audio input of the video card. The synchronization between video and audio was estimated to be within a few 100 ms. Single video frames were analyzed using Pinnacle studio Plus 9.3 (Pinnacle System Inc.) and MB-ruler 3.0. The swimming speed (in body lengths s^{-1}) was estimated in intervals of 0.5 s in a 6 s window starting 3 s before and 3 s after stimulation.

The 97.5% confidence interval of the mean swimming speed for the 3 s interval before exposure was computed by pooling the data from all fish for each exposure. A behavioral response was considered present if the swimming speed after exposure was twice this value (Fig. 1).

3. Results

Allis shad showed a change in swimming speed when exposed to ultrasound played at 70 and a 120 kHz (Fig. 1). Each swimming velocity is the mean of the response obtained during the

Table 1. The result of student's *t*-test made on the correlation coefficient at 70 and 120 kHz for the fish responding first in each test group when exposed to an increasing and a decreasing sound pressure level. G =fish group, coefficient $t=t$ value, significance level (marked with asterisks) $p<0.05$.

		70 kHz				120 kHz	
		G1	G2	G3	G4	G5	G6
Increase	r^2	0.76	0.24	0.94	0.44	0.96	0.83
	t	2.34*	0.50	5.53	0.99	6.50*	300*
Decrease	r^2	0.98	0.83	0.95	0.84	0.59	0.90
	t	9.32*	2.92*	6.17*	3.14*	1.50	4.10*

increasing and decreasing sound level exposure series for each group (Fig. 1). In all but three of the 12 stimulation sequences, a significant correlation was seen between the received level and the swimming speed measured 1 s after stimulation (Student's *t*-test on the correlation coefficient, $p<0.05$, see Table 1). The change in swimming speed gradually declined as the sound pressure level decreased. The response threshold defined as two times the 97.5% confidence interval (threshold at 70 kHz: 0.70 BL/s and at 120 kHz: 0.66 BL/s) was between 161 and 167 dB re 1 μPa (pp) at 70 kHz and 161 and 167 dB re 1 μPa (pp) at 120 kHz (Fig. 1). When exposed to the 2 kHz control sound, none of the fish exceeded the defined response threshold (Fig. 1).

Half of the fish groups were exposed to an increasing followed by a decreasing series of intensities, and the other half to the opposite with no difference in the derived thresholds.

To make sure that the fish responded to the ultrasonic output of the transducer and not to any omnidirectional low-frequency by-products or electric noise, a directional ultrasonic transducer (Reason 2116) was used on three fish (*sensu* Nestler *et al.*, 1992). When the transducer was directed toward the fish, the fish was exposed to both the directional ultrasonic pulse and the weak, omnidirectional low-frequency by-product. Sound exposure elicited a strong response in all of the three fish when the transducer was pointed at them. When the transducer was turned 90° with respect to the fish, exposing them to the weak low-frequency by-product only, no response could be detected. This shows that the fish did in fact respond to the ultrasonic stimuli and not the low-frequency by-product (*sensu* Nestler *et al.*, 1992).

4. Discussion

Allis shad respond to ultrasonic signals in the frequency range where toothed whales echolocate. The response thresholds between 161 and 167 dB re 1 μPa pp at 70 and 120 kHz are comparable to the behavioral thresholds obtained from juvenile American shad, which showed a very weak or no behavioral response below 160 dB re 1 μPa at frequencies between 20 and 160 kHz (Mann *et al.*, 1997). The similar thresholds of European and American species suggest that members of the *Alosinae* have evolved a dedicated ultrasound detector, possibly unique among all fish species.

The duration of 70 and 120 kHz pulses of 0.7 and 0.4 s, respectively, is three orders of magnitude longer than the duration of clicks from toothed whales (20–250 μs). We used such long pulses to make the study comparable with previous behavioral studies on ultrasound detection in other shad species (Plachta and Popper, 2003). The drawback of using long-duration signals in small tanks is a varying received level caused by interference patterns and that the energy carried in the sound pulses is much larger than for a toothed whale's click at the same peak intensity.

The behavioral response thresholds measured here are at least 20 dB above the hearing threshold found for American shad using acoustic brainstem response techniques (Mann *et al.*, 2001). Even though the fish can actually detect weaker sounds, the sound intensity apparently needs to be considerably higher before the fish responds behaviorally to the stimulus. Increased sound intensity leads to stronger behavioral responses indicating that allis shad have an

intensity-graded response to the output of its ultrasound detector as indicated for American shad (Plachta and Popper, 2003). The response threshold to ultrasound may reflect a trade-off between being caught and the costs associated with futile escapes (energy expenditure and lost opportunity to engage in other activities). Therefore, the response should depend on the animal's perception of the risk (Ydenberg and Dill, 1986). Allis shad may therefore use the intensity of the echolocating signals as a cue to estimate the proximity of an echolocating toothed whale: A distant toothed whale will cause lower received levels and hence only require a mild response of turning away from the predator, while high received levels of echolocation clicks would signify a close-by predator necessitating a strong and forceful escape.

Allis shad do not appear in the stomach content of the common dolphins (Pusineri *et al.* 2007), even though the habitats of the allis shad and common dolphins overlap in the study area. This observation, in combination with the present demonstration of a clear behavioral response when exposed to ultrasonic signals, indicates that allis shad may benefit from their ability to detect ultrasound to successfully minimize predation from echolocating toothed whales.

Acknowledgments

The authors thank N. Jeannot and the rest of the staff and students at the INRA Station de piégeage, Le Moulin des Princes, Pont Scorff, France, for hospitality and logistical support and L. Miller for lending us equipment. M.Wi. was funded by SOAS and NF, University of Aarhus and M.Wa. by the Carlsberg Foundation. This study was conducted under the approval of the Animal Care Committee of France under the official licence of M. L. Bégout (17-010).

References and links

Acolas, M. L., Bégout Anras, M. L., Véron, V., Jourdan, H., Sabatié, M. R., and Baglinière, J. L. (**2004**). "An assessment of the upstream migration and reproductive behaviour of allis shad (*Alosa alosa* L.) using acoustic tracking," J. Marine Sci. **61**, 1291–1304.

Au, W. W. L. (**1993**). *Sonar of Dolphins* (Springer, New York).

Baglinière, J. L., Sabatié, M. R., Rochard, E., Alexandrino, P., and Aprahamian, M. W. (**2003**). "The allis shad (*Alosa alosa*): Biology, ecology, range, and status of populations," in *Biodiversity, Status and Conservation of the World's Shad*, American Fisheries Society Symposium, edited by K. E. Limburg and J. R. Waldman, Vol. **35**, pp. 85–102.

Collin, S. P., and Marshall, N. J. (**2003**). *Sensory Processing in Aquatic Environments* (Springer, New York).

Gregory, J., Lewis, M., and Hateley, J. (**2007**). "Are twaite shad able to detect sound at a higher frequency than any other fish? Results from a high resolution imaging sonar," Proceedings of the Institute of Acoustics, Loughborough University, UK, p. 29, Part 3.

Hawkins, A. D. (**1981**). "The hearing abilities of fish," *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 109–138.

Mann, D. A., Higgs, D. M., Tavolga, W. N., Souza, M. J., and Popper, A. N. (**2001**). "Ultrasound detection by clupeiform fishes," J. Acoust. Soc. Am. **109**, 3048–3054.

Mann, D. A., Zhongmin, L., and Popper, A. N. (**1997**). "A clupeid fish can detect ultrasound," Nature (London) **389**, 341.

Miller, L. A., and Surlykke, A. (**2001**). "How some insects detect and avoid being eaten by bats: Tactics and countertactics of prey and predator," BioScience **51**(7), 570–581.

Nestler, J. M., Ploskey, G. R., Pickens, J., Menezes, J., and Schilt, C. (**1992**). "Responses of blueback herring to high-frequency sound and implications for reducing entrainment at hydropower dams," North Am. J. Fish. Manage. **12**, 667–683.

Plachta, D. T. T., and Popper, A. N. (**2003**). "Evasive responses of American shad (*Alosa sapidissima*) to ultrasonic stimuli," ARLO **4**(2), 25–30.

Plachta, D. T. T., Song, J., Halvorsen, M. B., and Popper, A. N. (**2004**). "Neuronal encoding of ultrasonic sound by a fish," J. Neurophysiol. **91**, 2590–2597.

Popper, A. N., Plachta, D. T. T., Mann, D. A., and Higgs, D. (**2004**). "Responses of clupeid fish to ultrasound: A review," ICES J. Mar. Sci. **61**, 1057–1061.

Pusineri, C., Magnin, V., Meynier, L., Spitz, J., Hassani, S., and Ridoux, V. (**2007**). "Food and feeding ecology of the common dolphin (*Delphinus delphis*) in the oceanic northeast Atlantic and comparison with its diet in neritic areas," Marine Mammal Sci. **23**(1), 30–47.

Santos, M. B., Pierce, G. J., Reid, R. J., Patterson, A. P., Ross, H. M., and Mente, E. (**2001**). "Stomach contents of bottlenose dolphins (*Tursiops truncatus*) in Scottish waters," J. Mar. Biol. Assoc. U.K. **81**, 873–878.

Ydenberg, R. C., and Dill, L. M. (**1986**). "The economics of fleeing from predators," Adv. Study Behav. **16**, 229–249.

Finding the onset of a room impulse response: Straightforward?

Guillaume Defrance, Laurent Daudet, and Jean-Dominique Polack
UPMC University Paris 06, Institut Jean Le Rond d'Alembert, LAM CNRS UMR 7190, Ministère de la Culture et de la Communication, 11 rue de Lourmel, 75015, Paris, France
defrance@lam.jussieu.fr, daudet@lam.jussieu.fr, polack@ccr.jussieu.fr

Abstract: This letter deals with precision issues in the determination of the timing of the room impulse responses (RIRs) onset. First, it is shown that while errors of onset timing estimation do not have that much effect on temporal indices, an erroneous onset estimation leads to significant differences in energetic and statistic acoustical indices. Twelve automatic onset detection methods are compared, in terms of precision, robustness, and complexity. Experimental validation made on a set of 100 RIRs provides good evidence in favor of spectral and/or energetic methods, according to the type of sound source.

PACS numbers: 43.55.Br, 43.55.Mc, 43.58.Gn [NX]
Date Received: April 22, 2008 **Date Accepted:** June 13, 2008

1. Introduction

A standard way to document the acoustics of a room is to measure a set of room impulse responses (RIRs). A RIR should ideally be recorded within absolute silence, a condition which of course is never met in practice. It then belongs to the acoustician to identify the edges of the RIR. Acousticians have for a long time designed various methods, as attested in the ISO 3382[1] standard, for determining the last point of the reverberation tail, which is mixed with background noise. However, nothing is said about the onset of the RIR, implicitly assuming that this is a straightforward task: Each acoustician, or software, can have its own method. In order to find the onset time, the simplest ways that come first to mind would be (1) to determine it manually, i.e., visually on the waveform; or (2) to consider that the RIR always starts by direct sound, which often has the largest amplitude of the signal. Measurements in a concert hall typically lead to 100 RIRs, so that the first method becomes easily cumbersome. The second method, as we shall see, can in some cases provide poor estimates of the onset time. For instance, when strong scattering attenuates the direct sound, it is the first reflection that presents the largest amplitude, and the RIR maximum and its beginning can differ from less than 0.1 ms up to more than 50 ms [Fig. 3(b)]. Therefore, direct comparison between different authors and/or methodologies may become unreliable.

The first goal of this letter is to show that, for some of the most commonly used room acoustical indices, the above methods lead to large errors. The ISO 3382 standard lists some of these indices. Energetic and statistic, such as Clarity (C_{80} in dB, at 80 ms) and Central Time (T_C in ms), are derived from a ratio of integration of the energy of the RIR; while temporal indices, such as Early Decay Time (EDT_{10}) and the Reverberation Times (RT_{20} and RT_{30}) are obtained from a linear regression made on the integration of the energy of the RIR. We do not attempt to discuss the relevance of such indices, be they orthogonal or not. This has already been done, for instance, by Pelorson *et al.*[2] We do attempt, however, to propose and evaluate a number of different automatic onset detection methods for their precision, robustness, and complexity. This is the second goal of this letter.

The letter is constructed as follows. In Sec. 2, we investigate how the precision for the localization of RIR onsets influences the result for five commonly used acoustic indices. Section 3 describes 12 methods for the automatic determination of RIR onsets. The performance of

Table 1. Variations of acoustical indices as a function of errors in the onset time Δt_0: $\Delta EDT_{10}(\%)$, $\Delta RT_{20}(\%)$, $\Delta RT_{30}(\%)$, $\Delta C_{80}(dB)$, $\Delta T_C(\%)$

Δt_0 (ms)	$\Delta EDT_{10}(\%)$	$\Delta RT_{20}(\%)$	$\Delta RT_{30}(\%)$	$\Delta C_{80}(dB)$	$\Delta T_C(\%)$
0.05	0.002	0.002	0.002	0.25	3.6
0.1	0.005	0.006	0.006	0.31	4.5
0.2	0.01	0.01	0.01	0.37	5.3
1.0	0.05	0.06	0.06	1.29	17.6
2.0	0.11	0.11	0.12	1.42	18.7
4.0	0.23	0.23	0.024	1.35	16.5
10.0	0.57	0.59	0.6	0.92	9.9

these methods is empirically compared in Sec. 4. The last section (Sec. 5) discusses guidelines for choosing the right method according to the application at hand, and concludes on the need to present reliable results with documented uncertainty.

2. Precision of onset timing versus precision of the indices

Assuming that t_0 is the reference time index corresponding to the maximum of the RIR, we compute on one full-band typical RIR (from Salle Pleyel, Paris[3]) the EDT_{10}, RTs, C_{80}, and T_C indices using different onset times $t' = t_0 - \Delta t_0$, with Δt_0 varying from 0.05 to 10 ms. Table 1 shows how acoustic indices can be affected differently by differences on onset timing. Even with $\Delta t_0 = 10$ ms, errors on EDT_{10} and RTs still remain weak (around 0.6%). The authors would like to point out that calculating the EDT according to Jordan's definition[4] leads automatically to large errors, since it looks for the time that the total energy has decayed by 10 dB.

On the contrary, $\Delta t_0 > 2$ ms leads to a difference in clarity of 1.45 dB (i.e., 55%). Variations of T_C remain inferior to 20% for the RIR tested here, with similar values for other RIRs. This way, even if the method used to determine the onset index is the maximum of the RIR, the user should prefer to calculate the central time than the clarity, because they both refer to the density repartition of the energy in the signal. This simple example shows the importance of exact onset detection of impulse responses in room acoustics. As errors on temporal indices are low, and would be considered as insignificant by most experts, as documented in Ref. 5, they will not be discussed in this letter. Nevertheless, attention is paid to energetic and statistic indices, since they have a strong dependence on the precision of the onset timing.

3. Proposed onset detection methods

This letter presents 12 different methods for onset detection. As the study of the acoustics of a hall can lead to typically hundreds of measurements (hence 100 RIRs), this article is focused on methods that can be computed automatically. The methods presented below rely on the idea that the onset is linked to abrupt changes in one or more properties of the audio signal,[6] and can therefore be detected by detecting the changes.

3.1 Temporally based methods

When observing the temporal structure of a RIR, it is noticeable that the occurrence of an onset is accompanied by a sudden increase of amplitude. The first methods of onset detection were based on this property by using a detection function which follows the envelope of the signal.[6]

Four simple functions are presented here: (1) Maximum (M): As seen in introduction, the first idea is to consider the onset t_0 to be the maximum of the absolute value of the RIR. (2) Maximum minus 5 ms (M_5): A few softwares, such as the MIDAS package,[7] consider that the onset can be defined 5 ms before the maximum of the RIR. (3) Mean over time (D_E): An envelope follower can easily be constructed by low-pass filtering the local energy.[6] The maximum of $E(n)$ is detected, and the signal analyzed from its beginning to its maximum. The ratio of two

successive windows is calculated; the index is found when the ratio is maximum [Eq. (1)]. This can be written as:

$$E(n) = \sum_{m=-N/2}^{N/2} w(m)x^2(n\,.\,h+m), \tag{1}$$

$$t_0 = h \times \arg\max_n(E(n+1)/E(n)), \tag{2}$$

where w is a smooth windowing function, and h the time step between two windows.

(4) Threshold (E): This method works on the energy of the RIR, which is windowed by rectangular windows $w(m)$; t_0 is here defined as the first time index where this local energy is K (typically $K=3$) times higher than its median running on past windows.

3.2 Spectrally based method

Since the spectra of the direct sound and the first reflections of an impulse response are very different from the background noise, a function based on spectra comparison is expected to give to good results, as seen in Ref. 8. One can expect an increase of low frequency components when looking at the direct sound.

(5) Mean over spectra (D_S): The idea is almost identical to D_E method, but ratios are calculated over spectra [Eq. (3)].

$$\tilde{E}(n) = \sum_k |X(n\,.\,h,k)|, \tag{3}$$

$$t_0 = h \times \arg\max_n(\tilde{E}(n+1)/\tilde{E}(n)), \tag{4}$$

where $X(n,k)$ is the short time Fourier transform of the signal $x(n)$, and h the time step between two windows.

3.3 Time frequency method

(6) Wavelet Transform (W): This method, often used for denoising,[9] is a natural tool for analyzing transient signals, since its time frequency resolution provides an increasingly finer time resolution at smaller scales. Let Ψ be a zero-mean real Gaussian wavelet. The wavelet transform of $x(t)$ is defined as

$$Wx(u,s) = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{s}}\Psi^*\left(\frac{t-u}{s}\right)dt, \tag{5}$$

where u is the translation parameter and s the scale factor.

In other words, $Wx(u,s)$ measures variations of $x(t)$ near u within an equivalent window of size s (Eq. (5)). When $s \to 0$, the decrease of the wavelet coefficients characterize the regularity of x around u. The onset of the transient signal is estimated by applying a threshold to the wavelet coefficients,[9] such as

$$Tm = \sigma_m\sqrt{2\log_e(N)}, \tag{6}$$

where σ_m is the standard deviation of the noise, and N the length of the signal. After inverse wavelet transform, the onset is estimated as the first nonzero sample.

3.4 Refinement step

The six methods presented above can be refined by a method based on linear prediction (L_P). Linear prediction analysis finds the coefficients of a finite impulse response linear filter that predicts the current value of the real-valued time series based on past samples, minimizing the prediction error in the least squares sense. L_P residual is computed on adjacent windows [t

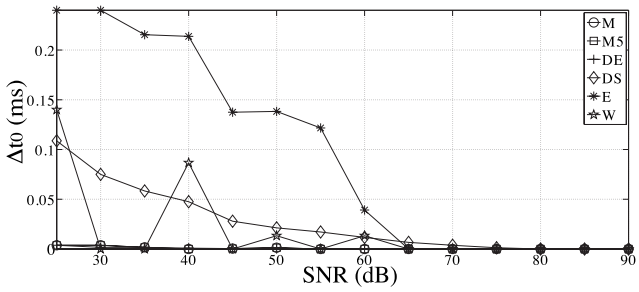


Fig. 1. Variations of onset times estimation as a function of *SNR*.

$-m:t]$ and $[t:t+m]$, with $m=0.8$ ms; t_0 is the time index that maximizes the likelihood of having a stationary Gaussian residual in both backward and forward windows, with a χ^2 goodness-of-fit test.[10] In other words, t_0 corresponds to a change point in the behavior of the system. This method is potentially very precise, but cannot be used as such, since it detects any spurious event regardless of its size. Instead, it is used only as a refinement step locally around the times provided by the previous methods. We call the six improved methods: ML_P, $M5L_P$, D_EL_P, D_SL_P, EL_P, and WL_P.

3.5 Reference method

An important issue regarding the evaluation of these methods is that the exact onset timing is not known (no absolute truth). These reference points could be determined manually by experts, at the cost of a tedious hand labeling and a potential lack of consistency between experts. Here, we decide to choose method D_SL_P as reference. The rationale for choosing this method is the following. First, on all the RIRs that we have at hand, D_E and D_S are the only methods which always return onset indices validated at hand by the authors (this is not always the case for M, M_5, E, W). As differences of onset estimation between methods D_E and D_S are inferior to 0.1 ms, we need to look at the enhanced methods. Indeed, method D_EL_P estimates the onset at the beginning of a small ripple that we call *precursor* (Figs. 3(c) and 3(d)), while method D_SL_P detects the onset at the end of the precursor. It is assumed that the precursor does not belong to the RIR itself, but is either an artifact of the pistol shot (maybe due to the cylinder rotation), or more probably an artifact of an anti-aliasing filter in the A/D conversion. Second, we assume that the most precise method should return the best onset index. L_P being a refinement step, this leaves D_SL_P as reference.

4. Comparison of the 12 automatic onset detection methods

4.1 Evaluation method

The 12 detection methods are tested over 100 audio wavfiles. These RIRs have been measured in Salle Pleyel in Paris,[3] according to the international standard,[1] with pistol shots as sound sources. For each RIR, the onset time is estimated by each of the 12 methods. From this, acoustical indices (C_{80}, T_C) are calculated and compared to the reference value, returned by method D_SL_P. Onset times that are not roughly consistent with the reference onset (i.e., not within 100 ms) are not taken into account for the statistics.

4.2 Robustness to noise

A way to assess the robustness of the presented estimators is to vary the background noise level of the RIRs. Figure 1 shows the mean variations of the onset time estimated by the first six methods on 100 experimental RIRs, with a signal to noise ratio (SNR) varying from 25 to 90 dB, by step of 5 dB. For each method, the reference onset time is the one obtained without adding noise to the signal. As expected from Sec. 3.3.3, the most robust method is W, followed by M and M_5; the worst method is E, followed by D_S.

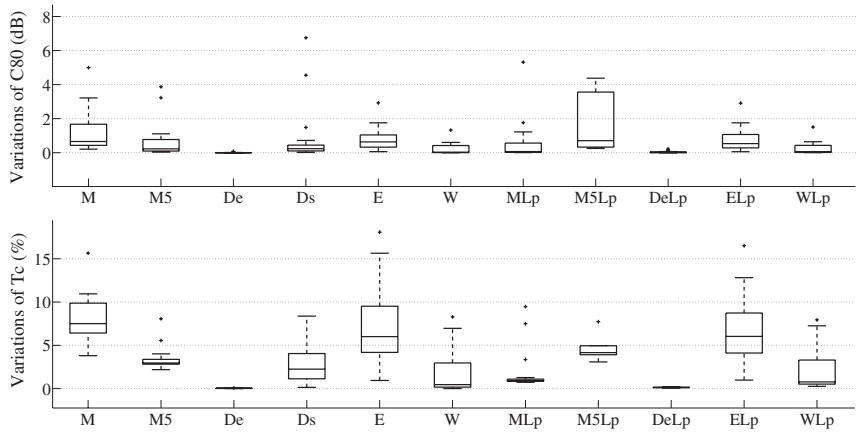


Fig. 2. Variations of C_{80} (top) and T_C (bottom) for the 11 methods, $D_S L_P$ being taken as reference.

4.3 Results and discussion

Except for large variations, which are discussed later, results (Fig. 2) show that adding the method L_P to any other method considerably improves the accuracy of onset detection, and hence, decreases errors, except for method M_5.

Methods M and M_5 are extremely robust to artifacts, but present important variations, since they do not account for a potential scattering effect. Figure 3 presents three different RIRs. The first one (a) starts very near its maximum. One can expect that detection results should not vary from a method to another. The second RIR (b) presents a long scattering effect ($\approx$30 ms), caused by a balcony. Methods M and M_5 provide bad estimations in that case. In cases (c) and (d), the onset is not the maximum, but a *precursor*, as introduced in Sec. 3.3.5. These small variations cannot be detected by M and M_5, but also by method E, since its threshold is not always adapted to the RIR's precursor amplitude. Moreover, because of its threshold, that the user has to set differently for each RIR, E is not suited to an automated analysis.

Methods D_E, D_S, and W seem to be particularly indicated for a quick and precise onset index determination, method W being more robust to noise than the other methods, as seen in Sec. 4.4.2. The resulting estimated C_{80} is always within ±0.01 dB (i.e., ±3%) of our reference method given by $D_S L_P$. Indeed, even if method L_P improves on other methods, the computation time is significantly increased for a negligible gain in precision (typically below 0.1%).

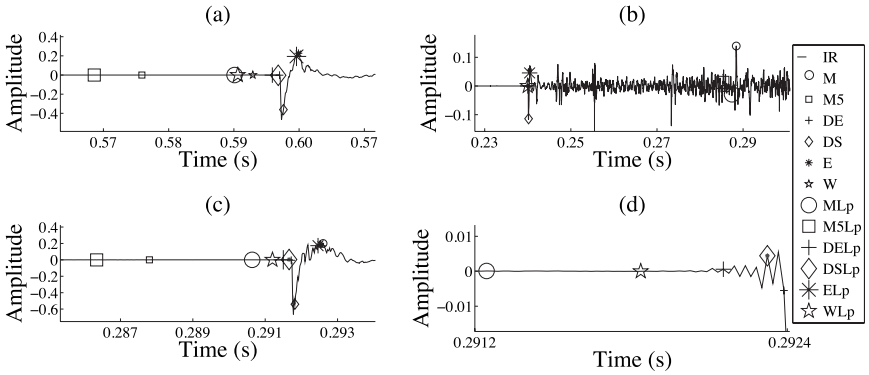


Fig. 3. Example of three different RIRs (note the different time scales). a) Simple case onset; b) Scattering effect; c) Precursor; d)Details of a precursor presented in (c).

4.4 Measurements with balloon bursts

The same experiment has been carried out on 100 RIRs measured with balloon bursts in Salle Pleyel (for the same source and receiver positions). Method $D_S L_P$ being also taken as reference, results slightly differ from those obtained with the pistol shots. Although method L_P improves all methods, except M_5, method D_S offers the best estimate ($\bar{\Delta}_{C_{80}}$=0.02 dB, $\sigma_{C_{80}}$=0.02 dB; $\bar{\Delta}_{T_C}$=0.3%, σ_{T_C}=0.01%), instead of D_E for pistol shots. This can be explained as follows. The pistol shot has a much sharper increase of energy than balloon burst,[11] typically 1.5 times faster. Thus, differences between pistol shots and balloon bursts are both spectral and temporal, and also related to their directivity, as explored in Refs. 11 and 12. Further studies are required to refine these claims. It should be noticed that nonlinear effects that may affect pistol shots are not detected by these methods since they intrinsically are parts of the response.

5. Conclusion

The main goal of this letter is to raise awareness on a loophole in the ISO 3382 standard for the computation of room acoustics indices. If one is interested in statistic and energetic indices such as T_C and C_{80}, a robust and precise method for determining the onset time is necessary. For instance, a variation of only 2 ms can generate high variations of clarity (around 1.5 dB, i.e., 55%). Furthermore, it is shown that for some of RIRs, there are large differences in the results given by obvious used onset detection methods, inducing significant differences in the acoustic indices.

Our experimental tests also show that methods based on energetical differences (method D_E), for pistols shots, on spectral differences (D_S) for balloon bursts, and on time-frequency analysis, such as wavelet transform (W), seem to provide reliable estimates, with a precision that is appropriate for most uses. Nevertheless, the computation time needed by method W, and the slight difference with results obtained with D_S, do not justify its use. These results highlight the inner spectral and temporal differences of frequently used sound sources. A statistically based refinement method is also a viable approach, but the slight gain in precision does not seem to justify the additional computational complexity. Further studies should test these methods on other sets of RIRs, including Ambisonics measurements of Salle Pleyel, and adjust the different important parameters such as window lengths, for an extended set of indices.

One may question the expected outcome of such detection algorithms. In the case of complex RIRs such as those with a precursor, what is the most relevant onset time from a perceptual point of view? Finding the beginning of the precursor may not be the best choice, since the precursor could be inaudible due to temporal masking effects. Such studies that are definitely beyond the scope of this letter, would require extensive listening tests. However, they remind us that these automatically generated signal processing indices are only meaningful if they provide information that has a *perceptual* relevance.

Acknowledgment

This work was partly supported by grants from Region Ile-de-France.

References and links

[1]ISO 3382, Acoustics-measurements of the reverberation time of rooms with reference to other acoustical parameters (1997).
[2]X. Pelorson, J.-P. Vian, and J.-D. Polack, "On the variability of room acoustical parameters: Reproducibility and statistical validity," Appl. Acoust. **37**, 175–198 (1992).
[3]G. Defrance, J.-D. Polack, and B.-F. Katz, "Measurements in the new Salle Pleyel," in *Proc. Int. Symp. Room Ac.* (Sevilla) (2007).
[4]V. L. Jordan, "A comprehensive musical criterion: The inversion index," JAES **23**, (2) 131–135 (1975).
[5]X. Meynial, J.-D. Polack, and G. Dodd, "Comparison between full-scale and 1:50 scale model measurements in theatre municipal, Le Mans," Acta Acust. **1**, 199–212 (1993).
[6]J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in musical signals," IEEE Trans. Speech Audio Process. **13**, 1035–1047 (2005).
[7]X. Meynial, G. Dodd, J.-D. Polack, and A.-H. Marshall, "All-scale model measurements: The MIDAS system,"

in *121st ASA Meeting, special session on auditorium measurements*, Baltimore (1991).
[8]D. J. Hermes, "Vowel-onset detection," J. Acoust. Soc. Am. **87**, 866–873 (1990).
[9]S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1999).
[10]M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes—Theory and Application* (Prentice–Hall, Englewood Cliffs. NJ, 1993).
[11]A. Nash, "On the acoustical characteristics of a balloon," in *Proc. Int. Symp. Room Acoustics (Sevilla)* (2007).
[12]D. Griesinger, "Beyond MLS-occupied hall measurement with fft techniques," 2004, URL http://world.std.com/~griengr/sweep.pdf.

Dolphin-inspired combined maneuvering and pinging for short-distance echolocation

Stephen E. Forsythe, Henry A. Leinhos, and Promode R. Bandyopadhyay
Naval Undersea Warfare Center, Newport, Rhode Island 02841
forsythese@npt.nuwc.navy.mil, leinhosha@npt.nuwc.navy.mil,
bandyopadhyaypr@npt.nuwc.navy.mil

Abstract: The biorobotic emulation of swimming and flying animals carrying out short-distance echolocation while maneuvering is considered. A simple and lightweight sonar for use on a small, maneuverable underwater vehicle for short-distance echolocation is explored. This sonar has four sensors and uses broadband, high-frequency signals to echolocate. The frequency-time characteristics of these signals are compared to those of bats and dolphins. The biosonar is paired with a biologically inspired, maneuverable, underwater vehicle, the combined use of sensors and maneuverability being analogous to animal behavior. Homing experiments have been carried out in an acoustic test facility where identification and localization of multiple targets is based on fusion of acoustic returns from multiple pings.

PACS numbers: 43.30.Vh, 43.30.Xm, 43.30.Yj, 43.80.Ka [JL]
Date Received: March 12, 2008 **Date Accepted:** June 14, 2008

1. Introduction

Dolphins swim incessantly with opportunity to ping repeatedly at a target from multiple directions and positions (Au and Simmons, 2007). Combined serial pinging and maneuvering offers an opportunity to quickly gain information with increased reliability about targets or surrounding. An animal's own motion and the motion in its surroundings are related to the processing of visual and auditory inputs received. The listening ability of some animals and their motion may be related due to functional reasons (Pollack and Martins, 2007). Between 1966 and 1998, a series of six conferences were held on bat and dolphin biosonar. Moore (2005) has reviewed the results. Müller and Kuc (2007) have reviewed bat-based engineering sonar systems. The present authors have developed a biorobotic vehicle that implements the high-lift hydrodynamic principle of swimming and flying animals (Ellington, 1984), which confers extreme maneuverability; they have also developed appropriate controllers to demonstrate these capabilities (Menozzi *et al.*, 2008). This vehicle has low self-noise, which allows experimentation in an acoustic test facility. A biology-inspired sonar paired with this biorobotic vehicle approximates the well-known echolocation capabilities of a dolphin (Bandyopadhyay, 2004 and 2005). This is the notional foundation of the present work. This approach of combining serial pinging with maneuverability receives some support from Yu *et al.* (2005), who recognized that a vehicle in motion can reduce noise and improve signal-to-noise ratio (SNR) by averaging over multiple pings. However, the effectiveness of averaging can be reduced by the relative motion between the sonar and the target. Yu *et al.* (2005) describe a method for estimating the sway and yaw motions that use multiple receive sensors. The motion estimates are used to synchronize the multiple pings. Their simulation suggests that the method is robust to noise and can improve the operating range of the sonar. Therefore, multiple pings from a vehicle maneuvering around a target, like an animal, can have high SNR if the vehicle motion is known and the pings are phase-referenced correctly. More specifically, the present authors develop the experimental technique of combining multiple pinging with animal-like vehicle maneuverability for target echolocation using only a handful of receivers. First, a biology-inspired, simple, lightweight sonar device is developed. This sonar is paired with the quiet biorobotic vehicle (named the

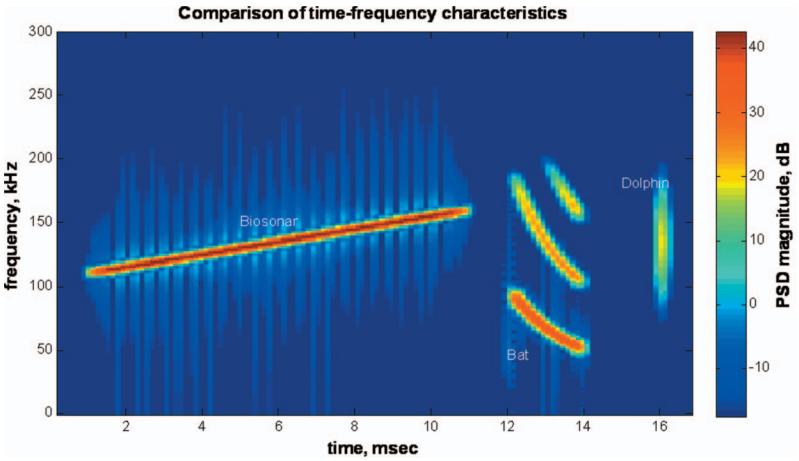


Fig. 1. Time-frequency analysis of three types of sonar.

BAUV—Biorobotic Autonomous Underwater Vehicle). The results on angular resolution and sensor-motion fusion of a homing experiment in an acoustic test facility using the sonar-equipped biorobotic vehicle are described.

The present work explores the design of simple, low-cost, lightweight, and low-power sonars for use in shallow waters. The proposed sonar is designed to approximate the capabilities of dolphin sonar in bandwidth, frequency range, and beamwidth due to their known discriminating abilities. However, we have used features of both bat and dolphin types of sonar. The bat-like frequency modulation (FM) sweep pulse affords a higher SNR than a dolphin-like click by increasing the total transmitted power per pulse. The processed output is presented as clicks with no regard for Doppler information per click (dolphin-like). The cross correlation of the return signal with the FM source waveform produces an impulse-like signal with a time width approximating that of the dolphin click; this is considered to be the raw sensory input of the system. Figure 1 compares the time-frequency content of a dolphin click, a bat chirp, and the biosonar signal. The signals were gathered from various sources and placed in series, and the short-time Fourier transform was used to clearly show the time-frequency behavior of all signals. The long pulse shown in Fig. 1 for the biosonar would be used for longer-range detection when SNR is a problem. As the range is shortened, shorter pulses would be used to avoid overlap of the return energy from multiple pings. Our biosonar uses the ITD (interaural time difference) to infer direction. Unlike the dolphin, our system uses four sensors to discriminate more readily in both the left-right and up-down directions. Biosonar research so far has not considered the role of the platform. It is not known how such platform motion would affect minimal-sensor broadband sonar performance.

2. Materials and methods

The following goals of the ITD sensor are defined: (1) <2-μs resolution with >20 dB SNR; (2) own speed is known to within 10%; (3) time delays of echoes are known to within <10 μs. The flat sonar head consists of one projector at the center and four receive sensors placed diametrically apart—two each in the horizontal and vertical planes (Fig. 3, Mm.1). The vehicle can move the sonar head forward, backward, or rotate it in any plane or axis at <5 kt. For our biosonar, both a horizontal and a vertical inter-sensor delay (ISD) analogous to the ITD can be calculated. Three-dimensional "perceptual space" based on the horizontal ISD, vertical ISD, and acoustic travel time maps into tentative images of objects in X, Y, Z in world coordinates. The percentage uncertainty in estimation of own speed implies that spatial discrimination is logarithmic in radius, yielding not a sphere, but an ellipsoid of estimated target location.

Mm. 1. Video recording of the motion of the present biosonar paired with the biorobotic BAUV vehicle freely swimming and maneuvering among the four targets in the NUWC Acoustic Test Facility. This is an.avi file of size 22 Mb.

The sonar was built in polyvinyl chloride (PVC) housing with circular pieces of 1–3 piezoceramic as the active elements. A single-board APEX power amplifier was used to provide the projector drive (using ±150 volts direct current (V DC) power supplies). The resulting ac voltage swing is about ±120 V. Kronhite preamplifier/filter modules were used for the four receive channel electronics. The unit was designed with mounting standoffs to attach it to the BAUV vehicle without impeding the unique motion of the biorobotic pectoral fin vehicle actuators (Mm. 1, Fig. 3). The transmitting voltage, a measure of sonar output versus voltage drive as a function of frequency, drops by 6 dB from 120 to 160 kHz. The receive voltage sensitivity (includes the preamplifiers), a measure of sonar response to input acoustic pressure, of the sonar as built remains mostly within 2 dB between sensors, flat between 135 and 160 kHz, and rises by generally 3 dB between 120 and 135 kHz.

Experimental data from an enclosed tank at the Naval Undersea Warfare Center (NUWC) in Newport, RI, shows the cross correlation (time series) resulting from a ping on a small target sphere placed in front of the sensor. The time series shows echoes from the target and from several other features nearby in the tank. Differences in arrival time between sensors are visible at this SNR (about 25 dB). Time differences on the order of less than 2 μs can be detected. Using the relation $\Delta t = \Delta x \sin(\theta)/c$, where Δt is the arrival time difference, Δx is the separation of the sensors, θ is the arrival angle, and c is the speed of sound in water, the corresponding angular separation can be calculated. For a minimum Δt of 2.5 μs, Δx of 15 cm, and c equal to 1500 m/s, the angular discrimination is 1° at the center of the sonar's field of view. The angular discrimination resolution was experimentally determined. As part of the calibration effort, FM sweeps were used to determine the position of a "bright" reflector—an air-filled, thin-walled plastic cylinder. The cylinder was set up <2 m away from the face of the sensor array. The overall dimensions of the calibration tank are 4.6 m deep $\times$ 4.6 m wide $\times$ 9.2 m long. A FM sweep from 100 to 140 kHz of 2 ms duration was used as the transmit waveform. The raw time series showed a strong, unambiguous return from the cylindrical target and the more diffuse returns from the walls of the tank and other structures within the tank. Early returns at $t=0$ due to the cross coupling between the co-located receivers and transmitter could be seen. The raw data are processed by cross correlation with the transmitted waveform. After cross correlation, a Hilbert transform is done to extract the envelope of the arriving energy. The absolute value of the resulting time series is then used to derive a smooth time envelope. Next, the time envelope is multiplied by a weighting function $w(t) \propto t^2$ to compensate for the distance to the target. This multiplication is used to normalize the target returns by removing the two-way travel distance from the target's return strength. The resulting time series more accurately shows the relative return strengths of multiple targets. To extract object information from the processed signals, local peaks (time and amplitude) are extracted from each of the four channels. Using the known dimensions of the array, a time window representing the maximum time delay between the two sensors of a horizontal or vertical pair is calculated (again using $\Delta t = \Delta x \sin(\theta)/c$). For an inter-sensor distance of 15.24 cm and signals arriving at less than 20° from the normal to the array face, this window is 35 μs wide. Peaks from a pair of sensors that are farther apart than this value cannot possibly be from a single object within the ±20° "cone of significance." By using a sliding window of this width to group candidate peaks, a potential object can be associated with peaks from all four channels that are within the sliding window. For each potential object, horizontal and vertical angles relative to the array face are computed by measuring the horizontal and vertical inter-peak arrival times. These arrival times are estimated to a high degree of accuracy (significantly less than the sampling interval) by using the five samples surrounding the peak as the basis for a least-squares fit to a parabola: $-y = at^2 + bt + c$. The peak time is given by $t = -b/2a$. Solving for θ in $\Delta t = \Delta x \sin(\theta)/c$ gives the horizontal and vertical angles to the object. The mean of the arrival times gives the object's distance using $d = t/2c$. After all potential object positions are calculated using the above method, the five

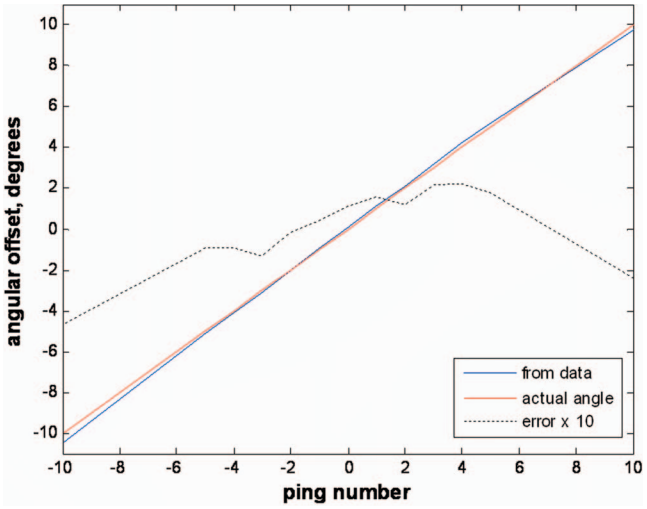


Fig. 2. Actual vs. estimated angle from array normal to target.

objects with the highest amplitude signals are retained for display and analysis. In the experiments, the target was selected by considering only those objects within 20 cm of the known target distance of 1.8 m. The array was rotated relative to the target to give effective horizontal angles to the target for successive pings of 10°, 5°, 4°, 3°, 2°, 1°, 0°, −1°, −2°, −3°, −4°, −5°, and −10° with respect to the face normal of the array. The angular estimates of horizontal angle using the peak data are shown in Fig. 2. The error (estimated angle minus actual angle) has been multiplied by 10 to make its range easier to read. Since the actual zero reference angle was not known more accurately than to within a few degrees, the mean of the estimated angles was subtracted from all estimated data. Looking at the errors after applying this correction, it can be said that the relative angular estimation performance of the array is excellent, rivaling the discrimination of 1.5° reported for some bats and 0.9° for some dolphins (Au, 1993, pp. 258 ff.).

Swimming and flying animals implement an unsteady high-lift principle of dynamic stall (Ellington, 1984). At the Naval Undersea Warfare Center, the authors have built a six-fin, rigid, cylindrical vehicle, called the BAUV (Fig. 3, Mm. 1). Each fin of the BAUV is independently imparted a rolling and pitching motion—the kinematic wing motions observed in flying insects. Thus, the fins implement the same high-lift principle as do swimming animals and flying insects. The result is that BAUV has animal-like maneuvering abilities (Menozzi *et al.*, 2008). For the present work, our biosonar system has been attached axially to the front of the BAUV (Fig. 3, Mm. 1). This biosonar-assembled BAUV vehicle was used in NUWC's Acoustic Test Facility (ATF) for target homing experiments. The facility contains 2.366 ml of fresh filtered water at a temperature of 19.7 °C. It is spanned by two movable bridges and overhead cranes. The test apparatus includes a three-dimensional positioner capable of movement over a 1.83 m cube to a 25.4 μm resolution; a coaxial rotator system that can position to 0.001° resolution; and an underwater camera/laser alignment system with VCR and/or computer recording capability of test events. The tank is partially lined with a thin wedge design anechoic lining that allows for measurements requiring long time durations at frequencies above 10 kHz. The full range of capability in frequency is 200 Hz to 5 MHz. The test system is VXI based with four channels capable of simultaneous data acquisition with independent delay capability and a sampling rate out to 20 MHz. The arbitrary waveform generation capability allows for the programming of any specialized waveforms.

3. Experimental fusion of multiple pinging and maneuvering

The fusion of the biosonar and the vehicle motion was carried out in the NUWC ATF. The vehicle was tethered. Multiple pings were used during the vehicle motion; target positions were

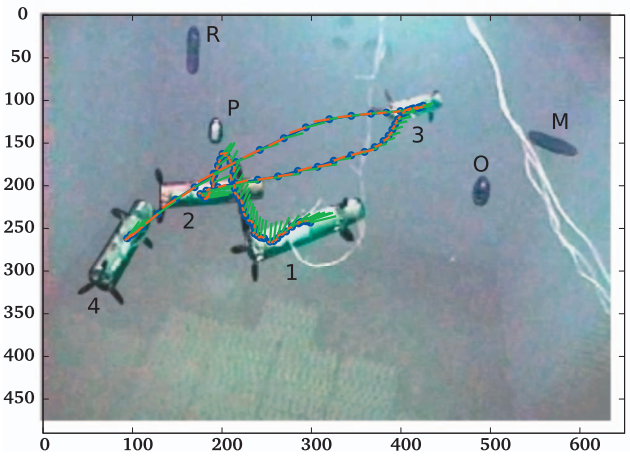


Fig. 3. Trajectory of the ITD-sonar installed BAUV vehicle (in the direction of 1–4) in the ATF at a depth of 4.6 m (Mm.1). The figure axes represent pixels. The blue circles represent the center of the vehicle, with the green lines indicating sonar-pointing vehicle (axis) direction and the red lines indicating the vehicle velocity vectors in relative scale. P (plastic cylinder), R (rock cylinder), O (oil-filled cylinder), and M (metal shell) are targets. The sonar head in the vehicle, visible beyond the fins, is pointing either toward the northeast corner or toward the north end of the tank. The distance between the two rear bottom corners of the tank is 12 m.

estimated by the fusion of data from multiple pings. In addition, the time-frequency signatures of a variety of targets were analyzed to discriminate targets by their composition. The experimental arrangement for tethered performance is shown in Fig. 3 and Mm. 1. The center of the world coordinates was taken to be the vehicle's center of rotation (0.53 m aft of the array face) at the vehicle's initial position. The vehicle was then moved as follows (one ping was done at each angle/location): rotate the vehicle from an initial angle of 35° to a final angle of −39° relative to the x axis in 2° increments; move the vehicle by 1 m in the x direction; rotate the vehicle from an initial angle of −39° to a final angle of 35° relative to the x axis in 2° increments; move the vehicle by 0.5 m in the x direction; rotate the vehicle from an initial angle of 35° to a final angle of −39° relative to the x axis in 2° increments. The result was a continuous path that can be described as rotate-move-rotate-move-rotate. Over the track, possible target estimates were formed using the original target estimation algorithm described earlier. The raw data were identified as possible objects by carrying out the following for each ping: Identify possible targets in local vehicle coordinates using the algorithm given above; translate possible target positions into world coordinates; assume that any possible target that is within a distance d (taken as 0.2 m for this experiment) of any other possible target from the previous ping is in fact a return from an actual object. Store the possible object position in world coordinates. The resulting possible object locations can be thought of as residing in a short-term memory. After possible objects were identified, the entire content of the short-term memory was analyzed together to further "solidify" the data into more-or-less permanent objects as follows: Group possible objects by sorting them into equivalence classes based on distance (i.e., two possible objects within a distance d of each other are actually views of the same object). Each equivalence class defines an object; develop statistics on the objects identified: Speed, spatial extent, orientation, etc.

No claims are made that actual biological processes are in any way analogous to the processing described above. It does seem reasonable, however, to make use of several layers of object representation, each building on a synthesis of data from the previous layers. Using data from the entire track described above, a picture of the significant objects in the environment was developed which showed both the raw target estimates taken from all 114 positions along the track and the statistics derived using the clustering algorithm. The algorithm mistakenly identifies a fifth target. This target results from spurious peaks in the original time-domain data that

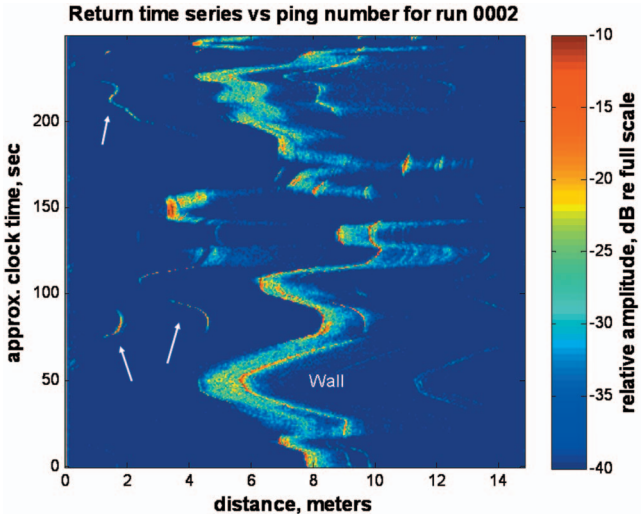


Fig. 4. Time series of processed echo returns as a function of calculated distance and clock time of the ping (ping rate is ≈2 pings/s). Arrows indicate targets.

are due to the resonant nature of the metal shell. The returns from the shell (second distribution from the right) show a spurious spatial dispersion due to the same effects. The other three targets are correctly shown as being compact cylinders that are positioned with their long axes vertically. The axes shown for each object are developed from a principal-axes analysis (using the Singular Value Decomposition) of the observations associated with each object; axes lengths are two standard deviation values for the spread of the observations.

A second homing experiment allowed the BAUV to swim freely while simultaneously pinging and acquiring acoustic data (Fig. 3). There was no independent sensor on board (e.g., for inertial guidance) to compare to the results of the pings, but an overall picture of the acoustic energy returned shows a target-rich environment, with target positions constantly changing as a result of the vehicle motion. The agility of the BAUV allows the acoustics to obtain many "views" of a target while the vehicle is in motion. In addition to swimming toward an object, the BAUV can circle the object and "crab" sideways to increase the effective array width, thus augmenting the binaural capability of the four receive sensors. Figure 4 shows an integrated picture formed by combining the processed pings over a period of 4 min while the BAUV swims through the tank. Figure 3 shows the position (blue) in 2 s intervals of the BAUV as it maneuvers between targets in the tank. The green line segments indicate vehicle axial direction and the red line segments indicate vehicle velocity. Between points 1 and 2, the BAUV descends to 4.6 m; between points 2 and 3, the BAUV approaches targets O and M; and from point 3 to point 4, the BAUV reverses to face targets P and R. In Fig. 4, there are multiple returns from the tank walls and the targets in the tank. Examples of the returns from the targets are the features at 1.8 m, 80 s and 2 m, 220 s. Wall echoes appear as diffuse ribbons at 6–8 m, with multiple reflections (e.g., from the corners and air-water boundary) such as the feature at 9–10 m, 130 s.

4. Conclusions

A biorobotic vehicle with an acoustic sensor has been shown to have the ability to form precise estimates of the relative positions of objects in the surrounding environment. This was accomplished with a simple acoustic array that has capabilities analogous to at least some of the known mechanisms for acoustic detection and location in dolphins. The absolute angular discrimination of $<1°$ seen in the present experiments is as good as the known capabilities of dolphins.

Acknowledgement

This work was sponsored by the Cognitive and Neuroscience Program of the Office of Naval Research.

References and links

Au, W. W. L. (**1993**). *Characteristics of Dolphin Sonar Signals. The Sonar of Dolphins* (Springer Verlag, New York).

Au, W. W. L., and Simmons, J. A. (**2007**). "Echolocation in dolphins and bats," Phys. Today **60**, 40–45.

Bandyopadhyay, P. R. (**2004**). "Guest Editorial: Biology-inspired science and technology for autonomous underwater vehicles," IEEE J. Ocean. Eng. **29**, 542–546.

Bandyopadhyay, P. R. (**2005**). "Trends in Biorobotic autonomous undersea vehicles," IEEE J. Ocean. Eng. **30**, 109–139.

Ellington, C. P. (**1984**). "The aerodynamics of hovering insect flight, IV. Aerodynamic mechanisms," Philos. Trans. R. Soc. London, Ser. B **305**, 79–113.

Menozzi, A., Leinhos, H., Beal, D. N., and Bandyopadhyay, P. R. (**2008**). "Open loop control of a multi-fin biorobotic underwater vehicle," IEEE J. Ocean. Eng. **33** (in press).

Moore, P. (**2005**). "Review of 'Echolocation in bat and dolphins'," J. Acoust. Soc. Am. **118**, 2755–2756.

Müller, R., and Kuc, R. (**2007**). "Biosonar-inspired technology: Goals, challenges and insights," Bioinsp. Biomim. **2**, S146–S161.

Pollack, G. S., and Martins, R. (**2007**). "Flight and hearing: Ultrasound sensitivity differs between flight-capable and flight-incapable morphs of a wing-dimorphic cricket species," J. Exp. Biol. **210**, 3160–3164.

Yu, L., Neretti, N., and Intrator, N. (**2005**). "Robust motion estimation improves underwater sonar accuracy," *CIHSPS 2005—IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, Orlando, FL, 31 March–1 April 2005.

Acoustic analysis of musical intervals in modern Byzantine Chant scales

Dimitrios Delviniotis, Georgios Kouroupetroglou, and Sergios Theodoridis
Department of Informatics and Telecommunications, University of Athens, Ilisia, GR-15784, Athens, Greece
ddelvis@di.uoa.gr, koupe@di.uoa.gr, stheodor@di.uoa.gr

Abstract: The goal of this work is to investigate experimentally the music intervals in modern Byzantine Chant performance and to compare the obtained results with the equal temperament scales introduced by the Patriarchal Music Committee (PMC). Current measurements resulted from pressure and electroglottographic recordings of 13 famous chanters singing scales of all the music genera. The scales' microintervals were derived after pitch detection based on autocorrelation, cepstrum, and harmonic product spectrum analysis. The microintervallic differences between the experimental values and the PMC's ones were statistically analyzed indicating large deviation of the mean values and the standard deviations. Significant interaction effects were identified among some genera and between ascending and descending scale directions.

PACS numbers: 43.75.Bc, 43.75.Rs, 43.75.St [TR]
Date Received: December 27, 2007 **Date Accepted:** July 17, 2008

1. Introduction

Byzantine Chant (BC) or Byzantine Music (BM) has been exclusively a monophonic vocal music (Wellesz, 1961) performed by chanters called psaltes. The term "echos" used in BM refers to the specialized meaning of the term "mode" (Arnold, 1983) denoting the musical scale being applied in a melody with a definite "tonic" or main note and specific music phrases, as well as its microintervallic structure or genus. There are three musical genera in BC: Diatonic (DI), Chromatic (CR)—subdivided in Malakon (CM) and Skliron (CS)—and Enharmonic (EN). The basic difference among these genera lies in their musical intervals. In modern BM, especially during the last 200 years, two main octave divisions have been proposed; first by Chrysanthos of Madytos (1832) and later by the Patriarchal Music Committee (1883). These approaches raised a number of discussions among those studying the BM. On the other hand, there are open issues related to the intervallic system on which a specific performer actually sings, along with the associated accuracy during his/her performance (Gabrielsson, 2003). Similar issues have been studied in other musical cultures (Sundberg, 1987) and types of music (Prame, 1997; Loosen, 1993; Ely, 1992; Fyk, 1995; Greer, 1970; Rakowski, 1990). Moreover, a corresponding system of music performance rules has been found to exist with regard to western music (Friberg *et al.*, 2006).

To the authors' knowledge there is no published systematic work to verify the results proposed by the PMC. The present study deals with the relative length of the musical intervals among notes in performed musical scales, considered as a whole. This approach was based on the finding that musical context may have a significant influence on intonation (Frances, 1958), and skilled musicians developed expertise for playing and perceiving typical musical sequences as a whole (Krumhansl, 1979). According to BM theory, a tetrachord is an interval of fourth (3:2) and there is a trend during the performance forcing the intervals inside the tetrachord to vary with respect to the direction of the note sequence performance (Chrysanthos from Madytos, 1832). The notes at the edges of the tetrachord remain almost unchanged, while the others inside it are moved upward or downward depending on the musical scale direction during the performance (PMC, 1883). This phenomenon was recently reported (Tsiappoutas *et al.*, 2004,

2006), but the conclusions were based on the performance of a two psaltes and two pieces of BM (not on musical scales). In this paper, the microintervallic structure of BM was studied as a function of two parameters: (a) musical genus and (b) ascending or descending musical scale direction.

2. The Patriarchal music committee's scales

The PMC (1883) proposed for the BC a 72-tone equal temperament system, with steps named "moria," each representing a frequency ratio of $2^{1/72}$, or 16.667 cents. PMC's values concern seven intervals: $(9/8)=12.23\text{m}/203.9\text{c}$[1] for all genera, $(9/8)\times(80/81)^2=9.65\text{m}/160.9\text{c}$ for DI, $(27/25)=7.99\text{m}/133.2\text{c}$ for DI and CM, $(9/8)\times(25/24)\times(80/81)^2=13.89\text{m}/231.5\text{c}$ for CM, $(9/8)\times(24/25)\times(80/81)^2=5.41\text{m}/90.2\text{c}$ for CS and EN, as well as $(9/8)\times(27/25)=20.23\text{m}/337.1\text{c}$ and $(25/24)=4.24\text{m}/70.6\text{c}$, for CS. Although the above-mentioned rather complex ratios, do not involve small enough numbers, in the sequel, they will be considered to be "justly" tuned intervals. All these values were approximated by PCM, according the 72-tone equal temperament system, into 12, 10, 8, 14, 6, 20, and 4 moria through truncating or rounding. Thus, the proposed PMC scales for the genera (along with their corresponding tetrachords) are DI: 12-10-8-12-12-10-8 (12-10-8), CM: 8-14-8-12-8-14-8 (8-14-8), CS: 6-20-4-12-6-20-4 (6-20-4), and EN: 12-12-6-12-12-12-6 (12-12-6), where between the two tetrachords of each scale a disjunctive interval of 12 moria exists.

3. Method

A total of 13 male subjects, average age 50.8 years, range 40–60, participated, following a systematic selection procedure. They were all famous psaltes, very experienced in BC performance, healthy, and without voice problems. They were asked to chant ascending and descending BC scales for the single Greek vowel /a/ in all genera, in a comfortable intensity, after a few warming-up exercises. The fundamental frequency range spanned from 132 up to 264 Hz. The pressure signal, captured in a studio by a condenser microphone (K2 RODE) at a constant distance of 30 cm, and the electroglottograph (EGG) signal were digitally recorded (console AWE16G) at a rate of 44.1 kHz, 16 bits. For each subject, a musical scale per genus was recorded five times, resulting in a corpus of 520 data scales. These recordings are part of the tagged DAMASKINOS prototype Acoustic Corpus of Byzantine Ecclesiastic Chant Voices (Chryssochoidis *et al.*, 2007). Four characteristic examples are presented in multimedia (Mm.) files.

Mm. 1. Diatonic scales sung by JA (1.494 kbytes).

Mm. 2. Enharmonicscales sung by JA (1.760 kbytes).

Mm. 3. Scliron Chromatic scales sung by AV (2.178 kbytes).

Mm. 4. Malakon Chromatic scales sung by AV (1.174 kbytes).

Three pitch detection algorithms (Rabiner and Schafer, 1978), i.e., the autocorrelation, the harmonic product spectrum (with eight replicas), and the cepstrum (with a lifter length corresponding to the range 50–500 Hz), implemented in MATLAB, were applied. A nonoverlapping 30 ms Hamming window was used to the pressure signal of each data scale. The autocorrelation algorithm was also applied to the EGG signal. The time domain autocorrelation method makes use of the filtering technique, leading to enhanced performance, while the other two frequency domain methods differ considerably in the way they exploit the spectral characteristics of a signal. Out of these four pitch methods, the one with the smooth pitch track and consisting of almost constant notes (which, however, include a vibrato), without octave's jumps and other errors, was selected for further analysis. Initially, the specification of the on- and offset of each note was based on the first derivative of a smoothed pitch track, in which the maxima and minima correspond to an increase and a decrease of the pitch, respectively. Only the extremes, over or under a threshold, were taken into further consideration. Then, the main "body" of a note

was specified to be at a certain pitch value if the rate and the extent of its vibrato were kept within certain limits (6.5 undulations per s and ±30 cents vibrato extent) (Sundberg, 1979). In order to reduce the influence of the transition parts at the boundaries of a note, a statistical analysis over all the frequencies inside each note was performed, so that the values closed to on- and offset of a note, which were larger than the median value of note's frequencies, plus or minus a standard deviation, were discarded. Since humans perceive the mean pitch in a logarithmic scale (Shonle and Horan, 1980), the use of moria and cents in pitch measurements is justified. Moria are used here for historical reasons. Notes with gliding tones (glissando) exceeding the threshold of 0.16 semitones (16c) per s, for 1 s tone duration were discarded, since their pitches would be somewhat problematic (d'Alessandro *et al.*, 1998).

The seven intervals of each eight-note scale were computed, relative to the starting note, for both ascending and descending scale directions. Thus, the total number of intervals for all the data scales was 3640. Whenever a pitch value was missed at some point of the scale, the next one was taken as the new starting point. The intervals of a tetrachord, along with the disjunctive one, were represented as, first (1st), second (2nd), third (3rd), and fourth (4th), respectively, for the genera DI, CM, CS, and EN. The following analysis was based on the assumption that the corresponding interval lengths of two tetrachords are identical, from a performance point of view. Each measured interval was compared with its counterpart of the PMC scale by calculating the difference $\Delta = |\text{measured}| - |\text{PMC}|$ with an accuracy of one decimal point, namely 0.1m/1.6c.

4. Results

The Δ values for all five trials of all psaltes were statistically analyzed. The intraperformers' variability was assumed to be comparable to the variability of each one of the data samples, since the mean value of all Δ for one performer did not differ from the values of the other performers by more than 3.3c. The average value of the standard deviation for all the performers was 31.6c for both ascending and descending scales. The one-way analysis of variance demonstrated closed relation among the performers for all intervals: $F=0.72$, $p=0.7053$, in descending scales and $F=0.54$, $p=0.8869$, in ascending scales.

All the mean Δ values do not exceed the range of −40c (min) to 33.3c (max). For both scale directions, the corresponding measures were −30c and 28.3c, respectively. The standard deviation varies between 21.6c and 50c. The average standard deviation for all the intervals was 30c, which was found to be comparable to the average value of the standard deviation for all the performers (31.6c).

Searching for possible trends in Δ values, with regard to the scale direction, the corresponding t-tests showed that for the intervals 2ndCM, 3rdCM, 3rdDI, 4thDI, 2ndEN, and 3rdEN (1st set) there is no statistically significant difference (95% confidence interval) between descending and ascending scales: $0.529 \leqslant p \leqslant 0.991$, while for the 3rdCS, 2ndDI, and 4thEN (2nd set), there is a small difference: $0.099 \leqslant p \leqslant 0.426$. For the rest of the intervals, namely 1stCM, 4thCM, 1stCS, 2ndCS, 4thCS, 1stDI, and 1stEN (3rd set), it appears that the Δ values differ between scale directions significantly ($p < 0.023$), except for the 1stEN ($p=0.087$). The corresponding differences between the means of Δ values for the sets of intervals (defined as ΔM), were measured to be smaller than 2.4c, for the first group of intervals, 4.4c, 4.9c, 4.2c, for the second one, and between 7.3c and 15.8c for the third one. The apparent trend in these findings is that the 1st interval of the tetrachord for all genera increases in the ascending with respect to the descending scale, while the 4th interval decreases for the two chromatic genera between the two scale directions (Fig. 1).

The t-test for the 1st, 2nd, and 3rd interval between the CM and CS genera showed that there are statistically significant (95%) differences in both scale directions. The p values were about zero ($p=0.000$) and the corresponding ΔM values were −13.5c, 57.2c, −45.1c for the ascending scale, and −14.5c, 64.5c, −49.8c for the descending one. The opposite signs in ΔM values between the adjacent intervals in the tetrachord indicate the opposite changes in length between them. In other words, when an interval that belongs to one chromatic genus increases,

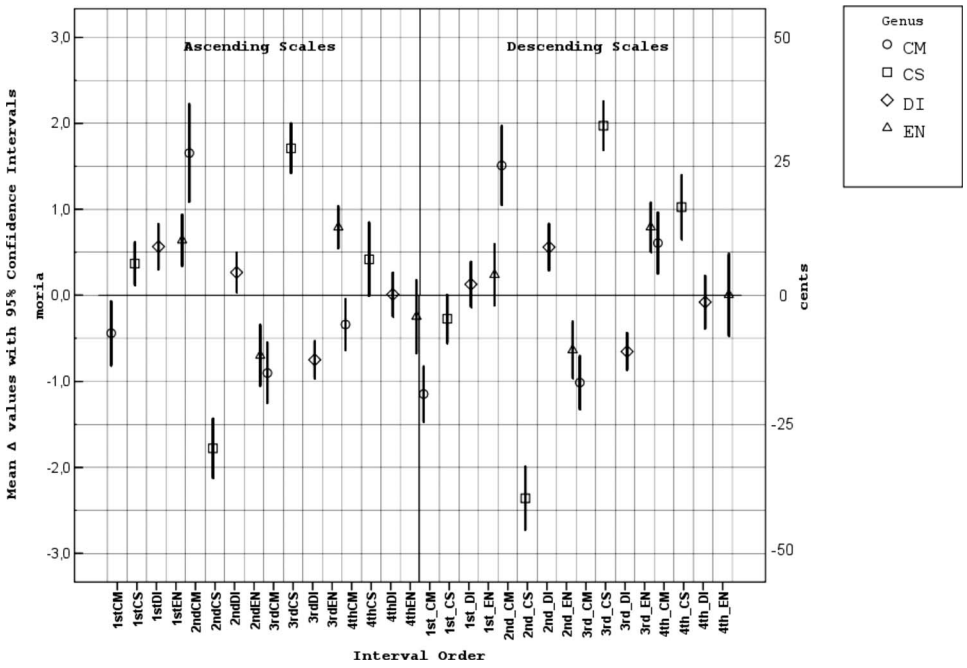


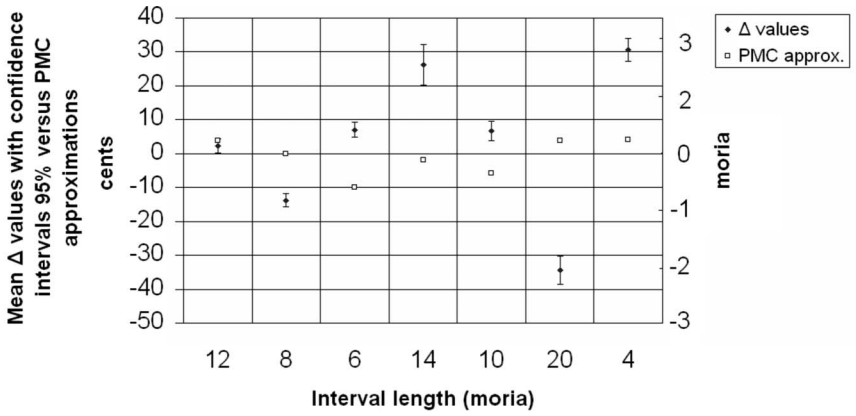
Fig. 1. Means of Δ values, over all performers, for each interval with their 95% confidence intervals, in ascending (left) and descending (right) scales. Means of Δ values for each genus are joined to each other by the same line as in the legend. Every vertical line corresponds to one interval located in axis according the genus order: CM-CR-DI-EN. The zero values of the ordinate refer to 72-tone temperament system.

the corresponding interval of the other genus decreases and visce versa (Fig. 1). The ΔM value for the 4th interval, between the two chromatic genera, was also found to be significant ($p=0.003$) in ascending direction of the scale (−13c), but not in descending ($p=0.117$) with value −7c. Between the DI and EN, for the 2nd and 3rd intervals, the ΔM values were computed to be 16.1c and −25.6c for the ascending scale, and 19.9c and −23.3c for the descending scale. Thus, they exhibit an explicit trend ($p=0.000$) to decrease for the 2ndEN interval relative to 2ndDI, and inverse, to increase for the 3rdEN relative to 3rdDI interval for both directions of the scale. The ΔM values, for the 1st and 4th intervals, between the diatonic and enharmonic genera, were found statistically insignificant ($0.269<p<0.754$) and were computed to be smaller than 4.3c.

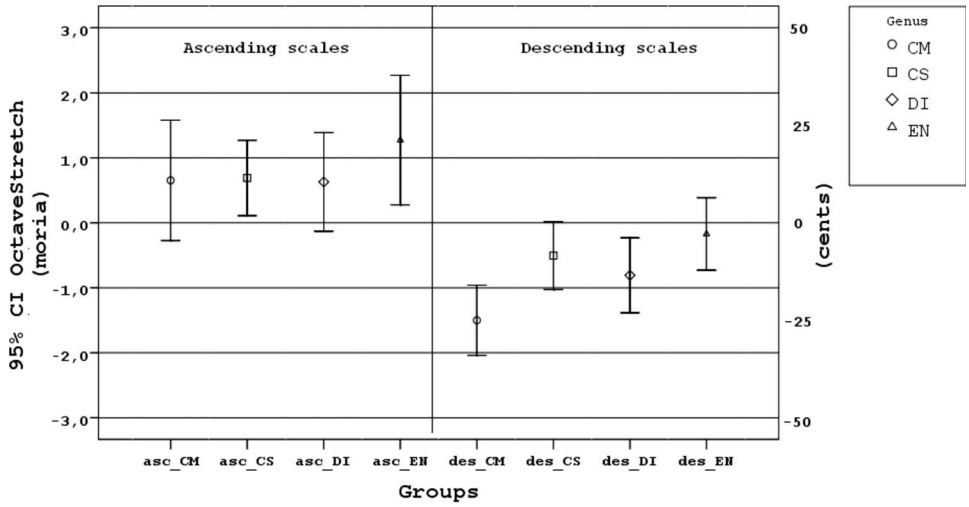
The less audible music interval, for PMC's musician experts, was considered to be 1m or 16.7c (PMC, 1883). This seems to differ from findings in other styles of singing for tones with vibrato, while it is in line for nonvibrato tones (Sundberg, 1987). Provided that two intervals in a scale differ at least 2m or 33.4c, any deviation more than 1m should be audible by musician experts (Burns, 1999).

5. Discussion

By considering that the interval values proposed by the PMC are approximations of their "just" counterparts and by examining their relations with the corresponding Δ values, the following remarks could be made: Although the interval of major second has been approximated by 12 moria for all the genera, this was found true only for the disjunctive interval and this can be also inferred from the proceedings of the PMC. Also, the two intervals of 8 moria (in DI and CM) and 6 moria (in EN and CS) have been considered to be identical. As it mentioned earller, the difference between the justly tuned intervals and the equal temperament ones proposed by the PMC are 0.23m/3.8c, −0.35m/−5.8c, −0.01m/0.2c, −0.11m/−1.8c, −0.59m/9.8c, 0.23m/3.8c, and 0.24m/4c for the intervals 12, 10, 8, 14, 6, 20, and 4 moria, respectively. The corresponding



(a)



(b)

Fig. 2. (a) Means of Δ values with confidence intervals (95%) and values of approximations of PMC's values, for each type of interval measured in moria. The zero values of the ordinate refer to "justly" tuned intervals of PMC. All the approximations of PMC's values do not fall into the confidence interval of a Δ mean except for the interval 12. (b) Mean octave stretch values for all genera and scale directions with confidence intervals (95%).

mean Δ values inside the confidence interval at the level of 95% were found 2.2±2.1c, 6.7±2.9c, −13.8±1.9c, 26.3±5.9c, 7.1±2.3c, −34.4±4.2c, and 30.6±3.3c [Fig. 2(a)]. Each one of the previous values does not belong in the confidence interval of the corresponding mean Δ value, showing an explicit deviation from that of the PMC's. The only exception is for the case of the disjunctive interval of 12 moria.

The statistically significant Δ*M* values between the CM and CS genera for the 1st, 2nd, and 3rd intervals and the ones between the DI and EN genera for 2nd and 3rd intervals indicate a strong interaction among them. The Δ values for adjacent intervals (without overlapping of their confidence intervals) have opposite signs between the two chromatic or diatonic genera (Fig. 1), except for the 1st CM-CS in the descending scale, and each absolute Δ*M* value between the two genera, for an interval, does not exceed the absolute difference of its PMC values, for

the two genera. For example, for the 1st interval between the two chromatic genera, the absolute ΔM value is 0.8m/13.3c in ascending scale, which does not exceed the absolute difference of its PMC values for the two genera: $|8-6|=2$m(33.3c). This finding also holds true for all the above-mentioned intervals: $(0<1\text{st}\ (13.3\text{c},15\text{c})<8-6(33.3\text{c}),\ 0<2\text{nd}\ (56.7\text{c},65\text{c})<20-14(100\text{c}),\ 0<3\text{rd}\ (45\text{c},50\text{c})<8-4(66.7\text{c}))$ (CM-CS), and $(0<2\text{nd}\ (16.7\text{c},25\text{c})<12-10(33.3\text{c}),\ 0<3\text{rd}\ (20\text{c},23.3\text{c})>8-6(33.3\text{c}))$ (DI-EN). An explanation for this could be that, while a chanter is trying to perform a very small or very large interval, as for example a chromatic one, he does not attain that (in the meaning of average values). The result is the production of an interval with intermediate length between its PMC values for the two genera. This phenomenon could be ascribed to an interaction between "relative" genera. Thus, we call this phenomenon of pitch alteration on the corresponding notes between two musical scales, *intra-genera interaction*.

The second factor that causes intervallic changes is the direction of the musical scale sequence. The influence of this factor can be apparent from ΔM values, for each interval, between the ascending and descending scale. The effect of intervallic change, due to the scale direction, has been mentioned in the literature on BC. However, what we know about it comes from the performers' experience. The trend for an interval to be modified, due to the scale direction change, may be based on the psychological effect of the attraction to the target note (Tsiappoutas *et al.*, 2004). As we mentioned before, in BM the scale is organized in the form of tetrachords or pentachords (a tetrachord including a disjunctive tone). That means as a performer chants the notes inside a tetrachord or pentachord starting from the tonal (starting note) and moving to the last note and then comes back he/she tends to drift them out of position. The most significant result of this effect is the alteration of the 1st interval, of which the main trend is to increase in length in ascending scale compared to one in descending scale, for all genera. When performing scales in the form of tetrachords, there is a tendency to modify the 3rd interval, as in the case of CS (small difference) or DI (very small difference), while in the case of EN there is a negligible modification (indicating a rigid interval). Also, the 3rd interval should appear an inverse change with respect to the 1st interval, (5c, −10c)(CS), (17c, −6.7c)(DI), (−6.7, 0.0c)EN, although there is a statistically insignificant modification. When performing in the form of pentachords, which is not rare in chanting a scale, the modification of the 4th interval is the main effect, as in the case of CM, CS, and EN genera. Such a phenomenon of a note's pitch change caused by a kind of attraction to the nearest note in scale has recently been reported as the attraction effect (Tsiappoutas *et al.*, 2006) but also it has been reported for several years in BM literature (Panagiotopoulos, 1981).

Provided that the Δ values were negative and positive ones, an expected mean of Δ values for all intervals inside an octave should be zero valued, however this is not the case. There exists a deviation from zero showing, of course, a trend existing in the performance and called octave stretch (Sundberg, 1972). Octave stretch concerns pure octave intervals, as well as sums of successive intervals of seconds inside an octave scale. Average values of all subjects' octave stretches with their corresponding confidence intervals were computed for each musical genus and scale direction [Fig. 2(b)]. This computation revealed that averages of octave stretches among different genera of same scale direction did not differ more than 10c in ascending and 21.7c in descending scale, but between the directions of scale there was an important difference [Fig. 2(b)] 46.7c as the maximum value. The average octave stretch for all genera and scale directions was computed to be 0.7c.

The former estimated octave stretch could explain the small alteration of the 3rd interval due to the scale direction for all genera. It can be suggested that by scaling down the music scale, especially in descending order, any change in the interval length could be masked by that stretch effect.

In order to have an estimation of the contribution for each of the above-mentioned factors of alteration of intervals, a score was fabricated from the sum of the absolute Δ values for each contributor divided by the number of values. Thus, the score value was measured to be 20.8c per interval, concerning the intra-genera interaction, and 7.4c per interval, due to the scale

direction. Under the assumption that the two contributors are independent of each other, the smallest contribution should be due to the direction of the musical sequence.

All the above-mentioned measurements concern sung notes extracted from a scale, namely each tone was sung in the context of a preceding or subsequent tone in that scale. It is known that a sung tone in such a context changes its pitch tending to come close to the next tone toward the direction of scale (Fyk, 1997). This results in an alteration of the music interval between the two successive tones and this is verified from our data.

6. Conclusions

A comparative analysis was conducted between experimental values of BC intervals during performance of musical scales and the ones of PMC's scales. Differences in interval lengths were measured from 0.1c up to 39.3c, while the biggest differences were found for the chromatic genera. All interval deviations are either positive or negative depending on the intra-genera interaction, the scale direction, and octave stretch effects. The corresponding intervals of the two chromatic genera deviate from PCS values in reverse way between them, which can be ascribed to the effect of the intra-genera interaction. Thus, DI and CM interact with EN and CS, respectively. Differences in interval sizes between the ascending and descending musical scales, during the performance, were measured. Thus, inside a pentachord or tetrachord, the first interval length increase and the fourth or third interval length decrease, in ascending scales, and, inversely, in the descending ones. Between the two effects of the intervals alteration, the one of the intra-genera interaction was observed to have the biggest influence in intervals deviation.

Acknowledgments

This work was cofunded by European Community Funds and Greek National resources under the project "AOIDOS" of the EPEAEK program.

References and links

[1]Where in this paper the value of musical intervals are given in both moria (m) and in cents (c), the symbol/is used between them

Arnold, D. (**1983**). *The New Oxford Companion to Music* (Oxford University Press, Oxford).
Burns, E. (**1999**). "Intervals, scales, and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic, San Diego).
Chrysanthos from Madytos. (**1832**). (Michele Weis Press, Trieste) (available in digital form from *ANEMI: The Digital Library of Modern Greek Studies* at http://anemi.lib.uoc.gr/metadata/4/4/3/metadata-01-0000443.tkl).
Chryssochoidis, G., Delviniotis, D., and Kouroupetroglou, G. (**2007**). "A semi-automated tagging methodology for Orthodox Ecclesiastic Chant Acoustic corpora," in *Proceedings of the Fourth Sound and Music Computing (SMC07) Conference*, Lefkada, Greece, 11–13 July 2007, pp. 126–133.
d'Alessandro, C., Rosset, S., and Rossi, J. (**1998**). "The pitch of short-duration fundamental frequency glissandos," J. Acoust. Soc. Am. **104**, 2339–2348.
Ely, M. C. (**1992**). "Effect of timbre on college woodwind players' intonational performance and perception," J. Res. Music Edu. **40**, 158–167.
Frances, R. (**1958**). *La Perception de la Musique*, (Vrin, Paris). Translated by W. J. Dowling. (**1988**). as *The Perception of Music* (Erlbaum, Hillsdale, N.J.).
Friberg, A., Bresin, R., and Sundberg, J. (**2006**). "Overview of the KTH rule system for music performance," Adv. Cog. Psychol. **2**, 145–161.
Fyk, J. (**1995**). *Melodic intonation, psychoacoustics, and the violin* (Organon, Zielona Góra, Poland).
Fyk, J. (**1997**). "Intonational protention in the performance of melodic octaves on the Violin," in *Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology* (Springer, Berlin).
Gabrielsson, A. (**2003**). "Music performance research at the millennium," Psycho. Music **31**, 221–272.
Greer, R. D. (**1970**). "The effect of timbre on brass-wind intonation," in *Studies in the Psychology of Music*, edited by E. Gordon (University of Iowa Press, Iowa City), Vol. **6**, pp. 65–94.
Krumhansl, C. L. (**1979**). "The psychological representation of musical pitch in a tonal context," Cogn. Psychol. **11**, 346–374.
Loosen, F. (**1993**). "Intonation of solo violin performance with reference to equally tempered, Pythagorean, and just intonations," J. Acoust. Soc. Am. **93**, 525–539.
Panagiotopoulos, D. (**1981**). *Theory and Praxis of Ecclesiastical Byzantine Music*, (Soter, Athens).
Patriarchal Music Committee (PMC). (**1883**). *Elementary Teaching of Ecclesiastical Music*, (Constantinople)

(reprint, Koultoura, Athens, **1978**).
Prame, E. (**1997**). "Vibrato extent and intonation in professional Western lyric singers," J. Acoust. Soc. Am. **102**, 616–21.
Rabiner, L., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).
Rakowski, A. (**1990**). "Intonation variants of musical intervals in isolation and in musical contexts," Psycholo. Music **18**, 60–72.
Shonle, J. I., and Horan, K. E. (**1980**). "The pitch of vibrato tones," J. Acoust. Soc. Am. **67**, 246–252.
Sundberg, J. (**1979**). "Perception of singing," STL-QPSR **20**, 1–48.
Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb).
Sundberg, J., and Lindqvist, J. (**1972**). "Musical octaves and pitch," J. Acoust. Soc. Am. **54**, 922–929.
Tsiappoutas, K., Ioup, G., and Ioup, J. (**2004**). "Measurement and analysis of Byzantine chant frequencies and frequency intervals," J. Acoust. Soc. Am. **116**, 2581.
Tsiappoutas, K., Ioup, G., and Ioup, J. (**2006**). "Frequency tracking of ecclesiastical Byzantine music frequency intervals," J. Acoust. Soc. Am. **119**, 3440.
Wellesz, E. (**1961**). *A History of Byzantine Music and Hymnography* (Clarendon, Oxford).

ACOUSTICAL NEWS—USA

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

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

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2007.

Independent Auditors's Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America (the "Society") as of December 31, 2007 and 2006 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

We conducted our audits in accordance with the auditing standards generally accepted in the United States of America. Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes consideration of internal control over financial reporting as a basis for designing audit procedures that are appropriate in the circumstances, but not for the purpose of expressing an opinion on the effectiveness of the Society's internal control over financial reporting. Accordingly, we express no such opinion. An audit includes examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements, assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation. We believe that our audits provide a reasonable basis for our opinion.

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of December 31, 2007 and 2006. and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

O'Connor Davies Munns & Dobbins, LLP
June 9, 2008
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
31 DECEMBER 2007 and 2006

	2007	2006
Assets		
Cash and cash equivalents	$ 2,796,600	$ 2,078,642
Accounts receivable, net	300,893	314,044
Investments	9,714,086	8,098,936
Property and equipment, net	34,579	50,880
Other assets	400,688	361,294
	$13,246,846	$10,903,796
Liabilities and Net Assets		
Liabilities		
Accounts payable and accrued expenses	$ 462,430	$ 429,259
Deferred revenue	1,539,161	1,442,037
Total liabilities	$ 2,001,591	$ 1,871,296
Net assets		
Unrestricted	$10,035,476	$ 7,994,438
Temporarily restricted	909,779	738,062
Permanently restricted	300,000	300,000
Total net assets	$11,245,255	$ 9,032,500
	$13,246,846	$10,903,796

 0001-4966/2008/124(4)/1877/18/$23.00

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2007 AND 2006

	2007	2006
Changes in Unrestricted Net Assets		
Revenues:		
Dues	$ 710,461	$ 688,698
Publishing—JASA	2,446,387	2,342,613
Standards	352,458	362,555
Spring Meeting	240,594	307,337
Fall Meeting	214,452	451,537
Other member services revenue	21,319	22,579
Other	156,006	180,069
Net assets released from restrictions	88,269	81,226
	$ 4,229,946	$4,436,614
Expenses:		
Publishing	$ 1,517,417	$1,491,176
Standards	396,502	386,729
Spring Meeting	237,305	359,404
Fall Meeting	287,469	297,535
Member Services	225,336	234,768
Other	517,701	440,196
Administration	575,819	531,828
	$ 3,757,549	$3,741,636
Net Income from Operations	$ 472,397	$ 694,978
Non-operating activities:		
Interest and dividends	$ 359,489	$ 256,643
Realized gain on investments	380,352	259,648
Unrealized gain on investments	777,681	204,266
	$ 1,517,522	$ 720,557
Change in Unrestricted Net Assets Before Adoption of Recognition Provisions of FASB Statement No. 158	$ 1,989,919	$1,415,535
Effect of adoption of FASB Statement No. 158	51,119	—
	$ 2,041,038	$1,415,535
Changes in Temporarily Restricted Net Assets		
Contributions	$ 22,637	$ 15,426
Interest and dividends	40,799	31,288
Realized gain	64,556	44,069
Unrealized gain	131,994	34,670
Net assets released from restrictions	(88,269)	(81,226)
Increase in temporarily restricted net assets	$ 171,717	$ 44,227
Change in net assets	$ 2,212,755	$1,459,762
Net assets at beginning of year	9,032,500	7,572,738
Net assets at end of year	$11,245,255	$9,032,500

Annual Reports of Technical Committees (See November issue for additional reports)

Acoustical Oceanography

I first want to thank my predecessor, Ross Chapman. Ross served as Chair of the Acoustical Oceanography Technical Committee for three years and AO is healthy and vibrant due in no small part to Ross's efforts. *Fall 2007 Meeting (New Orleans, Louisiana)*. The Technical Committee on Acoustical Oceanography (AO) sponsored two special sessions: "Storms and Intense Air-Sea Interactions" organized by Jeffrey Nystuen and "Deep and Shallow Seismic Sensing of Geological Structure organized by Altan Turgut which was also sponsored by Underwater Acoustics. AO also had two regular sessions entitled "Marine Sediment Properties and Inversions" and "Acoustic Effects of Internal Waves and Other Finescale Oceanography." AO also was a co-sponsor of three special sessions: "Acoustic Applications for Hurricane and Storm Preparedness in the Ocean Bottom" with Physical Acoustics, "Design of Distributed Surveillance and Oceanographic Monitoring Systems" with Underwater Acoustics and Signal Processing, and a named session for Leon Sibul with Signal Processing in Acoustics and Underwater Acoustics. Special thanks go to Jim Lynch who served as the AO representative to the Technical Program Organizing Meeting for New Orleans. The Best Student Paper Awards in Acoustical Oceanography went to David Barclay (first prize) of the University of California, San Diego for his paper "The effect of grain shape on the porosity of marine sediments" and to Megan Ballard (second prize) of Pennsylvania State University for her paper "Improved perturbative inversion schemes for obtaining bottom geoacoustic properties in shallow water." *Spring 2008 Meeting (Paris, France)*. The 2007 Medwin Prize in Acoustical Oceanography was awarded to Mark Trevorrow of Defence R&D Canada Atlantic for his contributions to acoustic measurement of bubbles, plankton and fish. Dr. Trevorrow presented the AO Prize Lecture entitled "The use of moored inverted echosounders for monitoring near-surface processes." In conjunction with the European Conference on Underwater Acoustics (ECUA), AO sponsored six special sessions in Paris: "Acoustical Oceanography of Polar Environments" organized by Peter Mikhalevsky and Hanne Sagen, "Marine Ecosystem Acoustics" organized by Kenneth Foote and Olav Godoe, "Geoacoustic Characterization of the Ocean Bottom and Geoacoustic Inversion" organized by Peter Gerstoft and Dag Tollefsen, "Passive Acoustic Tomography" organized by Karim Sabra and Sergio Jesus, "Acoustic Characterization of Sea Floor Habitats" organized by Christian de Moustier, Dick Simons and Xavier Lurton, and "Rapid Environmental Assessment" organized by John Osler and Yann Stephan. In addition, two regular sessions were organized: "General Topics in Acoustical Oceanography" and "Acoustical Tomography and Long Range Propagation." AO also was a co-sponsor of eight special sessions: "Animal Sonar Systems" with Animal Bioacoustics, Underwater Acoustics, and Signal Processing in Acoustics, "Seabed and Sea Surface Interaction Measurements and Modeling" with Underwater Acoustics and ECUA, "Impact of Environmental Variability on Mid-Frequency Sonar Performance" with Underwater Acoustics and ECUA, "Bayesian Signal Processing" with Signal Processing in Acoustics and ECUA, "Model-Based Signal Processing" with Signal Processing in Acoustics and ECUA, "Acoustics in Marine Archeology," with Engineering Acoustics, Underwater Acoustics, and ECUA, "Ultrasonic Acoustic MEMS" with Engineering Acoustics, and "Acoustically Activated Bubble Dynamics and Applications" with Physical Acoustics, and Biomedical Ultrasound/Bioresponse to Vibration. Special thanks go to Andone Lavery who served as the AO representative to the Technical Program Organizing Meeting for Paris. Congratulations to the following students for Best Paper awards in AO: First Prize (Poster) Paul Roberts, UCSD, "Application of multiple angle acoustic scatter to remote fish classification," First Prize (Presentation) Meghan Ballard, PSU, "Three dimensional geoacoustic inversion on the New Jersey Shelf," Second Prize (Poster) Aleksandra Kruss, IOPAN, Poland, "Estimation of macrophytes using single-beam and multibeam echosounding for environmental monitoring of arctic fjords (Kongsfjord, West Svalbard Island), and Second Prize (Presentation) David Barclay, UCSD, "Doppler Geo-Spectroscopy in the Makai Experiment." Information on these and related matters is available on the new Acoustical Oceanography Technical Committee website. It can be reached through the ASA web page by clicking on "Committees."

JAMES H. MILLER
Chair

Architectural Acoustics

TCAA had a very successful year in 2007-08. At the Fall 2007 meeting in New Orleans, TCAA special sessions and their organizers included "Architectural Acoustics: Sound Systems in Large Rooms and Stadia" David S. Woolworth and Timothy W. Leishman, "Better Than the Real Thing—Rock, Pop, and All That Jazz" Alex Case and Tony Hoover, "Impact and Footfall Noise I" and "Impact and Footfall Noise II" David Braslau, "Acoustics of Rehearsal Facilities" Damian Doria, 'Computational Room Acoustics and Room Acoustics Measurements" David S. Woolworth, and "Acoustics of Modular Construction" Edward Duncan. David S. Woolworth and Angelo Campanella served as the TCAA representatives to the Technical Program Organizing Meeting for this meeting.

The Spring 2008 meeting was actually a global conference on acoustics entitled Acoustics 08, held in Paris, France. TCAA chaired a record 18 special sessions with chairs from our technical committee joined by their European colleagues as co-chairs. Special session titles and chairs included: "Acoustics of Open-Plan Spaces I" Kenneth Roy, "Acoustics of Concert Halls I" Takayuki Hidaka, "Speech Segregation in Rooms I" Barbara Shinn-Cunningham, "Surround Sound Acoustics I" Alex Case and Tony Hoover, "Low Frequency Absorption: Mechanisms, Measurement Methods, and Application I" Peter D'Antonio, "Acoustics of Concert Halls II" Takayuki Hidaka, "Low Frequency Absorption: Mechanisms, Measurement Methods and Application II" Peter D'Antonio, "Acoustics and Electroacoustics of Small Rooms" Jiri Tichy and Mendel Kleiner, "Acoustics and Privacy in Healthcare Facilities I: Emerging Policy Around the World" David Sykes, "Acoustics of Opera Houses I" Robin Glosemeyer Petrone, "Comparison of US and European Standards in Building/Room Acoustics I" Kenneth Roy, "Acoustics of Opera Houses II" Robin Glosemeyer Petrone, "Acoustics and Privacy in Healthcare Facilities II: Emerging Research Around the World I" Bill Cavanaugh, "Acoustics of Open-Plan Spaces II" Kenneth Roy, "Acoustics of Opera Houses III" Robin Glosemeyer Petrone, "Acoustics of Concert Halls II" Takayuki Hidaka, "Low Frequency Absorption: Mechanisms, Measurement Methods and Application III" Peter D'Antonio, "Comparison of US and European Standards in Buildings" Room Acoustics II" Kenneth Roy, "Surround Sound Acoustics II" Alex Case and Tony Hoover, "Case Studies and Design Approaches" Byron Harrison, "Prediction Methods in Building Acoustics I" Berndt Zeitler, "Architectural Acoustics Potpourri I" David Bradley, "Prediction Methods in Building Acoustics II" Berndt Zeitler, "Measuring Methods and Uncertainty in Building Acoustics I" Brandon Tinianov, "Archeological Acoustics I" David Lubman, "Archeological Acoustics II" David Lubman, "Speech Segregation in Rooms II" Barbara Shinn-Cunningham, "Architectural Acoustics Potpourri III" Byron Harrison, "Acoustics and Privacy in Healthcare Facilities II: Emerging Research Around the World II" Bill Cavanaugh, "Case Studies and Design Approaches II" Byron Harrison, "Prediction Methods in Building Acoustics III" Berndt Zeitler, "Measuring Methods and Uncertainty in Building Acoustics II" Brandon Tinianov, "New Frontiers in Room Acoustical Modeling I" Murray Hodgson, "Airborne and Impact Sound Insulation I" Berndt Zeitler, "Airborne and Impact Sound Insulation II" Berndt Zeitler, "Coupled Volume Acoustics I" Jason Summers, "New Frontiers in Room Acoustical Modeling II" Murray Hodgson, "New Measurement Parameters in Performing Arts Spaces I" Lily Wang, "New Measurement Parameters in Performing Arts Spaces II" Lily Wang, "Coupled Volume Acoustics II" Jason Summers. David Adams crossed the ocean in February to serve as the TCAA representative to the Technical Program Organizing Meeting for this conference. There was an enormous effort on the part of these chairs and the high quality papers presented within the sessions. Thanks to each of them for the dedication in support of the TC.

Achieving good classroom acoustics continues to be an important topic that TCAA members are strongly advocating. In addition to permanent construction, the standard is being expanded to address the large number of modular or 'portable' classrooms in use today. Resolution of the topic is expected next year. Also the US Green Building Council has finalized the standard for LEED (Leadership in Energy and Environmental Design) encompassing school projects. Many of our TCAA members have provided input on the original document and the followup public comment period. Minimum acoustical performance is a required prerequisite plus two additional credits may be awarded for exemplary acoustical performance.

For almost three years, the joint subcommittee with the Technical Committee on Noise concerning *Speech Privacy*, chaired by Greg Tocci and

David Sykes, has been very active. The document "Draft Interim Sound and Vibration Design Guidelines for Hospitals and Healthcare Facilities" for the American Institute of Architects (AIA) and the American Hospital Association (AHA), is complete and in distribution. Likewise with the "Green Guide for Health Care," a joint project of the Center for Maximum Potential Building Systems and Health Care Without Harm. The subcommittee has convened well attended special sessions at every ASA meeting since their formation, including two in Paris. More information about this subcommittee's activities may be found on the website: www.speechprivacy.org.

Our newest subcommittee on *Green Building Acoustics* continues to influence green building standards and the acoustical issues they sometimes create. Special sessions are planned for both the Miami and Portland meetings. This group can be found at the webpage: http://groups.google.com/group/asa-gba.

This year the TCAA again offered Best Student Paper Awards. Two awards were presented for papers given at the New Orleans meeting; First prize went to John Magee from the University of Hartford, while a commendation award was given to Gordon Rubin of RPI. Two winners were selected for papers given at the Paris meeting: Anne Guthrie of Rensselaer Polytechnic Institute for first prize, and Yun Jing from Rensselaer Polytechnic Institute second prize.

Congratulations to TCAA members who became ASA Fellows this year. New Fellows include Elizabeth Cohen, Masayuki Morimoto, Tony Nash and Lily Wang.

Many thanks to the following individuals for their hard work this past year in ASA on behalf of TCAA. Red Wetherill is our representative on the Medals and Awards Committee. Ron Freiheit serves on the Membership Committee, and George Winzer is TCAA representative to the ASA Committee on Standards. Lauren Ronsse from the University of Nebraska will serve as our newly elected Student Council representative. The Associate Editors in Architectural Acoustics are Lily Wang and Ning Xiang for *JASA*, Ning Xiang for *JASA Express Letters*. Tony Hoover serves on the Editorial Board for *Acoustics Today*. Finally, thanks to Alex Case for his service as TCAA Secretary.

BRANDON TINIANOV
Chair

Engineering Acoustics

The Fall 2007 meeting of the ASA was a small but tight and well-structured meeting. Engineering Acoustics (EA) sponsored one full-day special session entitled "Infrasonic Instrumentation," organized and chaired by Allan Zuckerwar and Qamar Shams. This session was exceptional in its breadth of interdisciplinary interest and participation. This special session can serve as a model for future special sessions in Engineering Acoustics in that it successfully leveraged the role of EA as a common thread across all disciplines within the society.

It is fitting, then, that Allan Zuckerwar was honored with the Silver Medal in Engineering Acoustics for "contributions to the theory and practical development of transducers and their use in fundamental measurements." Congratulations to Allan for recognition well deserved.

The winner of the student paper competition was Tim Marston for "Comparison of measurement and theory for the acoustic pressure field in an infrasonic calibrator through the isentropic to isothermal transition."

For the Spring 2008 ASA joint meeting (Acoustics '08) members of the EA committee organized eight special sessions with the help of colleagues from our sister committees in the SFA and the EAA. The chair sends his personal gratitude to all the members who volunteered their time and effort to organize such an unusually large number of special sessions. Steve Thompson substituted for Kim Benjamin on the Medals and Awards Committee, and Ilene Busch-Vishniac for Tom Howardh on the Membership Committee.

Gunner Rasumussen received a Lifetime Achievement award from the EAA for "his outstanding contribution in the field of acoustic measurements and instrumentation" in the development and commercialization of measurement microphones at B&K and G.R.A.S.

In the spirit of the joint meeting, the student paper competition was opened up to members of all attending societies. The winners were Thierry Le Van Suu for "Thin film thermo-viscous damping in miniature condenser microphones," and Cyril Meynier for "Reproducibility and life duration assessment of cMUT transducers."

As was previously reported in the *ECHOES* newsletter, Gerhard Sessler received the Technology Award of the Eduard Rhein Foundation for his work on electret microphones with Jim West, who was awarded a US National Medal of Technology in 2006 for the collaboration.

Daniel Warren
Chair

Musical Acoustics

During 2007-2008 the Technical Committee on Musical Acoustics (TCMU) was chaired by James Beauchamp. Representatives to ASA committees were: James P. Cottingham, Membership; Uwe J. Hansen, Medals and Awards; Ian Lindevald, ASACOS; Eric A. Dieckman, Student Council, and James W. Beauchamp, Technical Council. Associate Editors were Diana Deutsch (JASA, Express Letters), Neville H. Fletcher (JASA), Thomas D. Rossing (JASA), and James W. Beauchamp (POMA). Technical Program Organization Committee (TPOM) representatives were Scott McDermott (New Orleans) and Antoine Chaigne (Paris). Those appointed or reappointed as TCMU members for 2008-2011 were James W. Beauchamp, George A. Bissinger, Annabel J. Cohen, James P. Cottingham, Evan B. Davis, Diana Deutsch, Uwe J. Hansen, Peter L. Hoekje, James H. Irwin, Ian M. Lindevald, Robert Pyle, Gary P. Scavone, Brad Story, George Tzanetakis, and Chris E. Waltham. Recently Paul Wheeler, a professor of electrical and computer engineering at Utah State University, was elected the new Chair of the TCMU for 2008-2011.

At the 154th Meeting in New Orleans in November-December, 2007, TCMU presented three special sessions: 2aMU/2pMU, "Musical Pitch Tracking and Sound Source Separation Leading to Automatic Music Transcription I & II," organized and chaired by James W. Beauchamp; 3eMU, "Science and Performance," organized and cochaired by Uwe J. Hansen and Tyrone M. Porter; and 4pMU, "Session in Honor of Max Mathews with Electronic and Computer Music Concert," organized and cochaired by Thomas D. Rossing and Julius O. Smith III. In addition, TCMU cosponsored the Architectural Acoustics sponsored special session "Even Better Than the Real Thing—Rock, Pop, and All That Jazz." Moreover, there was a session of contributed papers at this meeting: "Musical Acoustics and Instrumental Acoustics" chaired by Scott McDermott. Special musical performances in conjunction with the sessions were given by the Marlon Jordan Jazz Quartet on 3eMU; Rohan Krishnamurthy (South Indian drum) on 4aMU; and Maureen Chowning (soprano), Jon Appleton (radio baton), and Max V. Mathews (electronic violin) on 4pMU.

Like the other groups, TCMU was extremely well represented at the 155th meeting at Paris in terms of special sessions. There were so many sessions that this may be have been the first time that Music sessions had to be scheduled in parallel. (Fortunately, most Music sessions were held in the same conveniently located room just opposite the escalators on Level 2 of the Palais des Congrès.) The following special sessions were presented: 1pMUa/2pMUc, "Acoustic Measurements on Wind Instruments I & II," organized and cochaired by Noam Amir and David B. Sharp; 1pMUa/2pMUc, "Acoustic measurements on wind instruments," organized and cochaired by Shigeru Yoshikawa and Judit Angster; 1pMUc, "Acoustics and Psychoacoustics of Pipe Organs," organized and cochaired by Jonas Braasch and David M. Howard; 2aMUa/2pMUa, "Brass Instrument Acoustics I & II," organized and cochaired by Thomas Moore and Jöel Gilbert; 2aMUb/2pMUd, "Interaction Between Instrument and Instrumentalist I & II," organized and cochaired by Gary Scavone and Xavier Boutillon; 2pMUe, Singing Voice and Source-Filter Interaction," organized and cochaired by Brad Story and Nathalie Henrich; 3aMUa, "Interdisciplinary Research on the Science of Singing: A Tribute to Johan Sundberg," organized and cochaired by Annabel Cohen and Sten Ternstrom; 4aMU/4pMUf, "Virtual Musical Instruments," organized and cochaired by Julius O. Smith and Antoine Chaigne; 4pMUa/4pMUe/5aMUa, "Control of natural and synthetic musical sounds," organized and cochaired by Marcelo Wanderley and Jean Kergomard; 5aMUd/5aMUe, "Signal representations and models of musical sounds," organized and cochaired by Bertrand David and James W. Beauchamp. In addition, TCMU cosponsored the Architectural Acoustics sponsored special sessions 1pAAd/2pAAj, "Surround Sound Acoustics I & II" and 2aAAe/2pAAb/2pAAf," "Acoustics of Opera Houses I, II & III." Moreover, there were several sessions of contributed papers at this meeting: 3aMUa/4pMUc, "Plucked Stringed Instruments I & II," cochaired by François Gautier and Chris Waltham; 3pMUa/4pMUb, "Bowed and Keyboard Stringed Instru-

ments I & II," cochaired by Knut Guettler and Claudia Fritz; 3pMUb/4pMUd, "Wind Instruments I & II," chaired by Seiji Adachi & Simon Félix; 5aMUc/5aMUf, "Plucked and Struck Idiophones I & II," chaired by Tom Rossing and Charles Besainou. Music was also prevalent as an extracurricular activity at the Paris meeting. The first night an organ concert was presented at the Saint Sulpice Cathedral in Paris. On Tuesday and Wednesday there were tours of the Lutheries Acoustique Musique (LAM) musical acoustics lab in Paris's 15th arrondissment. On Thursday and Friday there were tours of the Institut de Recherche et Coordination Acoustique/Musique (IRCAM). Also, on Thursday there was a guided tour of "Pipe Organs of Paris" with organs ranging in age from 400 years to the present.

TCMU continues to promote student involvement in musical acoustics by continuing to sponsor the ASA Best Student Paper Awards in Musical Acoustics. Intense interest in this competition is demonstrated by an increasing number of entries: There were 10 entries for the New Orleans meeting and 32 for the Paris meeting. The winners for New Orleans were Rohan Krishnamurthy (Kalamazoo College), who received First Place for "Acoustics of the mridangam: Study of a new design of a South Indian drum," and Edward L. Toussaint (Lawrence Univ.), who won Second Place for "Reed vibration, pressure, and airflow in Western free-reed instruments." At Paris the winners were Jer-Ming Chen (Univ. New South Wales, Sydney), who received First Place for "How to play the first bar of *Rhapsody in Blue*," and Edgar Berdahl (Stanford Univ.), who won Second Place for "Applications to passivity theory to the active control of acoustic musical instruments."

TCMU continues to maintain a very useful website (http://www.public.coe.edu/~jcotting/tcmu/). It currently contains links to 12 musical acoustics laboratories and associations and numerous websites on acoustics education, as well as close to 200 links to musical acousticians, giving their names and specific interest topics. A new feature added in 2007-2008 is a list of musical acoustics papers published in JASA with links to the abstracts for these papers. A number of papers on musical acoustics topics were published in JASA during 2007-2008. There were papers on such topics as the acoustics of winds, strings, organ, piano, African mbira, and musical pitch perception.

JAMES BEAUCHAMP
Chair 2005—2008

Noise

TC Noise was well represented at the 154th meeting of ASA in Spring 2007 in New Orleans with 7 Special Sessions again showing the widespread aspects of noise: Architectural Acoustics, Musical Acoustics, Signal Processing in Acoustics, and Noise: Even Better than the Real Thing-Rock, Pop, and all that Jazz—Alexander U. Case and K. Anthony Hoover; Lawn, Yard, and Portable Noise and Noise Generation, Modeling, and Measurement-Richard D. Godfrey and Karl B. Washburn; Rain Noise-Robert C. Coffein. Soundscape Developments: Case Studies and Best Practices-Brigitte Schulte-Fortkamp and Bennett Brooks; Noise and Animal Bioacoustics: Advances in Measurement of Noise and Noise Effects on Animals and Humans in the Environment I+II–Ann E. Bowles and Brigitte Schulte-Fortkamp; Architectural Acoustics, Noise, and ASA Committee on Standards; Acoustics of Modular Construction-Edward C. Duncan.

TC Noise was proud to announce that James E. West, Research Professor, Department of Electrical and Computer Engineering, The Johns Hopkins University Whiting School of Engineering, received the 2006 National Medal of Technology "for co-inventing the electret microphone while working with Gerhard Sessler at Bell Labs in 1962." Jan Achenbach, Walter P. Murphy and Distinguished McCormick School Professor in the Department of Civil and Environmental Engineering at Northwestern University, received the 2005 National Medal of Science for seminal contributions to engineering research and education in the area of wave propagation in solids and for pioneering the field of quantitative non-destructive evaluation. U. S. President George W. Bush presented the 2005 and 2006 National Medals of Science and Technology at a White House ceremony on Friday, July 27, at 1:50 p.m. The ceremony was carried live by satellite feed and Web cast on the White House Web site.

Rich Peppin was the Noise representative at the Technical Program Organizing Meeting for New Orleans. The Noise Young Presenter Award went to Yun Jing for his paper "Sound field modeling in urban squares using a diffusion model" and to Cole V. Duke for his paper "Measurement and evaluation of blade passage frequency fluctuations." Brandon Tinianov and Vladimir Ostashev were announced as new Fellows.

For final conclusions on this meeting it should be highlighted that the joint session "Noise and Noise Effects on Animals and Humans in the Environment I+II" was planned and understood as a first step in collaboration with respect to measures and methodologies in research in both areas of Animal Bioacoustics and Noise. The Soundscape approach got much attention and the session will be continued at the ASA meeting in Miami November 2008. There is also a report on this session in *Acoustics Today* Volume 4, Issue 2, April 2008.

The efforts of several volunteers should be recognized: Nancy Timmerman is the Noise representative on the Medals and Awards Committee, John Erdreich was the representative on the Membership Committee until this meeting and Richard Peppin is the representative on the ASA Committee on Standards. Cole Duke is the Student Council representative and, following the practice initiated a few meetings ago, the secretary for the TC-Noise meetings.

The Noise web page [http://www.nonoise.org/quietnet/tcn/] is maintained by Les Blomberg. Joe Posey is the new coordinator for the Noise Young Presenter Awards. For the meeting in New Orleans the TC Noise Chair took over this responsibility. Serving as Associate Editor for *JASA Express Letters* is Mike Stinson and *JASA* Associate Editors are: Keith Attenborough, Kenneth Cunefare, Vladimir Ostashev, and Brigitte Schulte-Fortkamp.

The 155th ASA meeting in Paris joint with EAA, SFA, ECUA, was an overwhelming meeting. TC Noise collaborated with EuroNoise and was responsible for 30 sessions covering Classroom acoustics-Louis Sutherland, Luigi Maffei and David Lubman; Cultural variations in sound/noise assessment-Catherine Guastavino and Danièle Dubois; Measurement of occupational noise exposure-William Murphy and Beat Hohmann; Noise from wind power projects-Eddi Duncan and Kerstin Persson Waye; Soundscape and community noise-Brigitte Schulte-Fortkamp and Bennett Brooks; Aeroacoustics-Philip Morris and Christophe Bailly; Tire-road noise from the road perspective-Paul Donavan and Fabienne Anfosso-Lédée; Prominent discrete tones-Robert Hellweg and Lothar Schmidt; Sound quality tools and applications-Wade Bray, Klaus Genuit and Etienne Parizet; Airframe noise measurement, prediction, and control-Joe Posey and Denis Gely; Sound and vibration from explosions in air-Roger Waxler and Keith Attenborough; Environmental noise mapping-Kenneth Cunefare and Gaetano Licitra; Session in honor of Henning von Gierke-Paul Schomer and Brigitte Schulte-Fortkamp; Noise, vibration and acoustics for medical and research facilities and their occupants-James West, Jack Evans and Marc Asselineau; Sleep disturbances and other health effects-Lily Wang, Barbara Griefahn and Jacques Lambert; Vibration perception-Patricia Davies and Michael Griffin; Physical and psychophysical evaluation of vehicle exterior noise-Paul Donavan and Hugo Fastl; Fan noise and low-Mach number rotating blade noise-Scott Morris and Michel Roger; Car acoustics-Luc Mongeau and Virginie Maillard; Action planning and global solutions for urban noise-Klaus Genuit and Itziar Aspuru; Acoustic performance of energy efficient building products-Anthony Hoover and Jacques Roland; Soundscape in the heritage of urban and natural areas-Giovanni Brambilla and Bennett Brooks; Potential to reduce tire/road noise-Stuart Bolton and Ernst-Ulrich Saemann; Noise mapping techniques and uncertainties-Richard Peppin and Wolfgang Probst; Noise, structure borne noise from building technical equipment, and ground borne noise from railways-James Nelson and Michel Villot; EU projects for aircraft noise reduction-Dominique Collin; Source identification in radiation and scattering-Stuart Bolton and Peter Møller Juhl; Time-domain modeling methods in acoustics-Paul Calamia and Maarten Hornikx.

Keith Attenborough and Brigitte Schulte-Fortkamp were our representatives at the Technical Program Organizing Meeting for Paris. Noise had 53 applications for the Young Presenter Award! Our new fellow is Dick Botteldooren. TC Noise collaborated also successfully with TC Noise EAA; for the Technical committee meeting 51 participants were counted, 37 from ASA and 14 from EAA.

Let me close my report citing Elaine Moran: "Acoustics'08 Paris will be talked about for decades to come. The contacts made at that meeting will undoubtedly result in collaborations across oceans and countries which can only be good for acoustics."

BRIGITTE SCHULTE-FORTKAMP
Chair

Psychological and Physiological Acoustics

Reflecting P&P's policy of emphasizing spring meetings, the fall meeting in New Orleans was a small one. P&P sponsored two sessions. We thank Bob Turner for stepping in on short notice to be the representative to the Technical Program Organizing Meting.

The Paris meeting was quite an experience. Although impossibly busy, the science was at a high level. There were 16 sessions sponsored by P&P, 13 of them special sessions. We thank the organizers of the special sessions: Christian Lorenzi and Brian Moore (Role of Temporal Fine Structure in Speech and Non-Speech Perception for Normal and Hearing-Impaired People), Blas Espinoza-Varas and Alain de Cheveigne (Cross-Spectral Auditory Integration), Huanping Dai and Birger Kollmeier (Auditory Perception and Signal Processing by Prostheses), Steve Colburn, Bernhard Seeber, and Bernhard Laback (Cochlear Implants: Going Beyond the Envelope), Patricia Davies and Hugo Fastl (Applications of Psychoacoustics), Sridhar Kalluri and Michael Akeroyd (Binaural Perception by Hearing Aid Wearers), Jont Allen and Sarah Watkins (Acoustic Features and Speech Perception), Mounya Elhilali and Daniel Pressnitzer (Integrated Approaches to Auditory Scene Analysis), Erv Hafter and Armin Kohlrausch (Jens Blauert and His Contributions), John Neuhoff and Anna Preiss (Auditory Perception of Sound Source Properties), Jeremy Marozeau and Sabine Meunier (Loudness, From Controlled Stimuli to Environmental Sounds), Chris Shera and Arturo Moleti (Otoacoustic Emissions, From Cochlear Modeling to Experimental Techniques, and Back), and Durand Begault and Lauri Savioja (Computational Auralization). Co-sponsors of P&P sessions included ASACOS, Architectural Acoustics Noise, Signal Processing in Acoustics, Computational Acoustics, and Speech Communication. P&P was a co-sponsor of 4 other sessions. We thank Armin Kohlrausch and Alain de Cheveigne and the TC Chair for their work in organizing the P&P sessions, Bill Yost as meeting organizing committee co-chair, and Bill Hartmann as technical organizing committee co-chair.

The P&P open meeting was well-attended, and was jointly run by the ASA TC chair and Armin Kohlrausch, representing EAA. There was discussion of the need for nomination of more fellows to the ASA. Incoming chair Andrew Oxenham will be organizing a committee to determine likely candidates for fellowship. Peggy Nelson announced that the executive council was very likely to pass a proposal to have JASA papers go directly to PubMed, in fulfillment of requirements from many granting agencies that publications from funded research be available to the general public. Rich Freyman informed the group that when submitting papers to JASA, they should send a cover letter requesting the associate editor, as the AEs do not see the computer requests submitted by authors. The P&P Associate Editors provided an overview of their efforts. We thank the continuing P&P Associate Editors, Rich Freyman, Ruth Litovsky, John Middlebrooks, Brenda Lonsbury-Martin, Brian Moore, Bill Shofner, and Magda Wojtczak for their efforts on our behalf. We also thank Erick Gallun, our AE for POMA.

We express our thanks to Mike Heinz for helping to organize the TC ballot and for counting the ballots, and, thanks to members of the P&P community who so willingly added their names to the ballot. The newly elected members of the Technical Committee are Jose Alcantara, Emily Buss, Erick Gallun, Sridhar Kalluri, Gerald Kidd, Bernhard Seeber, and Dan Tollin. We thank our outgoing members, Michael Akeroyd, Alain de Cheveigne, Brent Edwards, Mike Heinz, and Lynne Werner.

We thank Bill Hartmann, our representative to the Medals and Awards Committee, Lynne Werner, our representative to the Membership Committee, and Brent Edwards, our representative to ASACOS. We also appreciate the efforts of our member of the executive council, Peggy Nelson. Also congratulations to the new fellows, Michael Akeroyd, Brian Glasberg, and Christophe Micheyl.

As I pass the torch to the incoming chair, Andrew Oxenham, the P&P Technical Initiatives continue unchanged. The initiatives include travel support for invited speakers, student receptions, and homepage maintenance. Suggestions for uses of funds, including innovations such as workshops, satellite meetings, etc., are welcome (oxenham@umn.edu).

ELIZABETH A. STRICKLAND
Chair, 2005-2008

Signal Processing in Acoustics

The Signal Processing Technical Committee (SPTC) has been very active during the period of this report. We organized 2 sessions for the 154th ASA meeting in New Orleans, November 27 to December 1, 2007. We also organized 5 sessions and cosponsored another 5 for the 155th ASA meeting in Paris, June 29 to July 4, 2008. The Paris meeting was held jointly with the European Acoustics Association and the Société Française d'Acoustique. We sponsored a competition for best paper by a young presenter for both meetings, the Gallery of Acoustics competition in Paris, and a student challenge problem.

At the New Orleans meeting, a special session in honor of Leon Sibul was organized by Lee Culver. Leon was one of the original organizers of the SPTC and a mentor to many signal processors in ASA. The Best Paper by a Young Presenter Award was given to Joris Vanherzeele from the Vrije Universiteit Brussel in Brussels, Belgium, for his presentation "Acoustic source identification using a generalized regressive discrete Fourier series for tomographic reconstruction." We wish to thank Joe Posey for representing SPTC during the Technical Program Organizing Meeting (TPOM).

The meeting in Paris was the largest recorded gathering of acousticians in history. The SPTC organized 5 special sessions and co-sponsored 11 sessions. Three special sessions were organized on time-reversal methods in acoustics: "Overview of time reversal in acoustics," "Time reversal methods for array imaging and signal processing," both organized by David Chambers and Claire Prada, and "Biomedical applications of time-reversal" organized by Oleg Sapozhnikov and Mickael Tanter. Eliza Michalopoulou and Simon Godsill organized a special session entitled "Bayesian signal processing," and Sean Lehman and Christian Pichot organized a session entitled "Model-based signal processing." In addition there were five regular sessions, including a session on "Advances in acoustic sensors and networks for defense applications" organized and chaired by Brian Ferguson and Pierre Naz. A competition for best student paper was held, with the entries currently being scored. We wish to thank Jean-Pierre Hermand for representing SPTC at the TPOM.

This year, the Gallery of Acoustics, organized by Sean Lehman, was held at the Paris meeting. This is a competition for the best audiovisual presentation of acoustical phenomena, with entries judged by both technical and aesthetic criteria. We had five entries: "Sound generated by impact on thin ice" by Mikhail Mironov, A. P. Pyatakov, and P. A. Pyatokov, "Outstanding sound pressure distribution on the outlet of spiral ducts" by Wojciech Lapka and Czeslaw Cempel, "Time reversal in solids (or how to focus wave energy)" by Pierre-Yves Le Bas, and Brian E. Anderson, "Acoustic imaging to increase noise awareness" by Stephan Paul and William D. Fonseca, and "Speech…Rhythm" by Volker Dellwo. These entries showed a great variety of creativity, artistic flair, and ingenuity in presenting a broad range of acoustic phenomena. Attendees of the meeting were asked to vote on each entry. The winners of first prize were Wojciech Lapka and Czeslaw Cempel for their video presentation on pressure distributions in spiral ducts.

Each year Signal Processing sponsors a Student Challenge problem that can be found in the SP section of the ASA web site. This year's problem asked the student to extract a sinusoidal communication signal from a moving vehicle out of noise in the presence of a jamming signal. Paul Gendron organized the contest and judging. The First prize of $1000 was awarded to Berke Gur from the University of Massachusetts Lowell.

We wish to thank Edmund J. Sullivan and William Carey, our signal processing Associate Editors of the *Journal of the Acoustical Society of America*, Jim Candy, our signal processing Associate Editor for *JASA Express Letters*, and Sean Lehman, our signal processing Associate Editor of *Proceedings of Meetings on Acoustics*, for the selfless service to the Society and SPTC.

DAVID H. CHAMBERS
Chair

Speech Communication

The Speech Technical Committee (TC) supports the activities, meetings, publications, etc., for the largest technical area in the Society. This report covers the ASA meeting in New Orleans, Lousiana and the Acoustics'08 meeting in Paris, France. The current members of the TC are Augustine Agwuele, Jean Andruski, Norma Barroso, Fredericka Bell-Berti, Patrice Beddor, Tessa Bent, Lynne Bernstein, Ocke-Schwen Bohn, Norma Antonanzas Barroso, Suzanne E. Boyce, Ann Bradlow, Dani Byrd, Kate Bunton, Roger Chan, Rebeka Campos-Astorkiza, Alexander Francis, Bruce Gerratt, Kenneth Grant, Susan Guion, Helen Hanson, Racheael Holt, Markus Iseli,

Paul Iverson, Diane Kewley-Port, Jody E. Kreiman, Jelena Krivokapic, Anders Lofqvist, Andrew Lotto, Benjamin Munson, Shrikanth Narayanan, Terrence Nearey, Peggy Nelson, Dwayne Paschall, Douglas, O'Shaughnessy, Joseph Perkell, Astrid Schmidt-Nielsen, Christine, Shadle, Rahul Shrivastav, Mitchell Sommers, Rajka Smiljanic, Brad Story, Joan Sussman, Scott Thomson, Gary G. Weismer and Zhaoyan Zhang. Many of these members are new to the TC. I want to welcome them and thank them for their support.

Other STC members who assisted us by serving on committees were Freddie Bell-Berti and Diane Kewley-Port—Executive Council, Jody Kreiman—Membership Committee, Shrikanth Narayanan—ASACOS, Anders Loqvist—Medals and Awards Committee, Corine Bickley and Ann Syrdal—Standards Committee on TTS Systems, Carol Espy-Wilson—Rossing Prize Committee, Jennell Vick and Chritian Stilp—Student Representative, and Brad Story who maintains our web page. The continuing Associate Editors are: Speech Production—David Berry, Anders Lofqvist and Christine Shadle, Speech Perception—Kenneth Grant, Paul Iverson, Mitchell Sommers, Joan Sussman, Rochelle Newman and Allard Jongman; and Speech Processing—Douglas O'Shaughnessy.

The two ASA meetings went very smoothly this year because of the particular efforts of several members: First, we are grateful to our Technical Program Organizing Meeting (TPOM) representatives, who sorted papers, arranged the technical programs, and determined presentation rooms. They were Emily Tobey for the New Orleans meeting and Paul Iverson for the Paris meeting. Second, we greatly appreciate the effort of the coordinators for student judging at the meetings: Peter Assmann in New Orleans and Suzanne Boyce, Ocke-Schwen Bohn and Helen Hanson in Paris (over 70 papers were judged at the Paris meeting). We applaud Ingo Titze for receiving the Silver Medal in Speech Communication at the New Orleans meeting, and Pat Kuhl for receiving the Gold Medal Award of the ASA at the Paris meeting. We also are very happy to welcome Sarah Hawkins, Brad Story and Dani Byrd as new fellows of the ASA in the Speech Communication TC.

Student Activities

In our continuing effort to promote student participation in ASA meetings, the Speech Technical Committee sponsored two student activities at each meeting, a competition with a cash award for best student presentation and an evening reception in New Orleans and a student luncheon in Paris. The student reception, which is joint with other technical committees, is intended to allow students to meet more senior ASA members informally. The receptions were well attended. The student papers were judged by STC members and the winners were awarded $300 for first prize and $200 for second prize. At the Salt Lake City meeting, the first place winner was Yolanda Holt, Ohio State University for "Are vowels in African American English affected by Southern vowel shift?" The second place winner was Wendy Herd, University of Kansas for "Incomplete neutralization of /t/ and /d/ in flapping environments: An analysis of the North Midland dialect of American English."

At the New Orleans meeting, the first place winner was Youngsok Jung, Havard-MIT division of Health Science and Technology, for "Acoustic articulatory evidence for quantal vowel categories: The feature [low]." The second place winner was Melissa Baese, Northeastern University, for "Enhancing learning on foreign-accent adaptation with a combination of active training and passive stimulus exposure." A special thanks to all of the reviewers of the student papers.

Special Session and Special Talks

To create stimulating and focused sessions we sponsor special sessions every year, which focus on themes of interest to the speech community. In New Orleans, special sessions addressed "Role of Attention in Speech Perception" chaired by Alexander L. Francis, "Auditory and Somatosensory Feedback in Speech Production I" and "Auditory and Somatosensory Feedback in Speech Production II," chaired by Anders Lofqvist. There was a Hot Topics presentation given by Ann Bradlow and Carol Espy-Wilson on "Speech-in-Noise Perception and Recognition: What humans *do* and what computers *should do* to handle speech in noise." In Paris, there were twelve special sessions: "Articulatory modeling and control of speech and singing organs" cochaired by Maureen Stone and Phil Hoole, "Multi-modal speech technology" cochaired by Gerasimos Potamianos and Gerard Bailly, "How do physical and motor knowledge matter to speech?" cochaired by Carol Fowler and Mikko Sams, "Acoustics of speech production: aeroacoustics and phonation" cochaired by David Berry and Xavier Pelorson, "Speech recognition in noisy environments" cochaired by Oliveir Siohan and Georges Linares, "Measurement of sociophonetic variation in speech" cochaired by Jonathan Harrington and Ewa Jacewicz, "Neurobiology of speech perception" cochaired by Paul Iverson and Christope Pallier, "Speaker identification by machine" chaired by Louis-Jean Boe, "Cross-language speech perception and production" cochaired by Yue Wang and Dawn Behne, "Speech prosody and how it relates to segmental aspects of speech," cochaired by Hansjorg and Mixdorff.

I would like to thank the Technical Council for all of their support and for being such an agreeable group with which to work.

CAROL ESPY-WILSON
Chair

USA Meetings Calendar

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

	2008
10-14 Nov	156th Meeting of the Acoustical Society of America, Miami, FL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
	2009
18-22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
26-30 October	158th Meeting of the Acoustical Society of America, San Antonio, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
	2010
19-23 May	158th Meeting of the Acoustical Society of America, Baltimore, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
15-19 November	2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org]

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

Volumes 1-10, 1929-1938: JASA, and Contemporary Literature, 1937-1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10

Volumes 11-20, 1939-1948: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of Print

Volumes 21-30, 1949-1958: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75

Volumes 31-35, 1959-1963: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90

Volumes 36-44, 1964-1968: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.

Volumes 36-44, 1964-1968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print

Volumes 45-54, 1969-1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound)

Volumes 55-64, 1974-1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound)

Volumes 65-74, 1979-1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound)

Volumes 75-84, 1984-1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

Volumes 85-94, 1989-1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

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

Volumes 105-114, 1999-2003: JASA and Patents. Classified by subject and indexed by author and inventor. Pp.616, Price: ASA members $50; Nonmembers $90 (paperbound)

Members of Technical and Administrative Committees of the Acoustical Society of America

The Technical and Administrative Committees listed below have been appointed by the President with the approval of the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2009.

Technical Committees 2008-2009

Acoustical Oceanography

James H. Miller, Chair to 2010

Term to 2011
Kelly J. Benoit-Bird
Gerald L. D'Spain
Timothy F. Duda
David M. Farmer
Gary J. Heald
Jean-Pierre Hermand
Paul C. Hines
Daniela Di Iorio
David P. Knobles
Timothy G. Leighton
Ying-Tsong Lin
Jennifer Miksis-Olds
Daniel Rouseff
Emmanuel K. Skarsoulis
Dajun Tang
Altan Turgut
Joseph D. Warren
Thomas C. Weber

Term to 2010
Kyle M. Becker
N. Ross Chapman
Grant B. Deane
Brian D. Dushaw
Matt A. Dzieciuch
Peter Gerstoft
Oleg A. Godin
John K. Horne
Bruce M. Howe
James F. Lynch
Anthony P. Lyons
Gopu Potty
Ralph A. Stephen
Alexandra I. Tolstoy
Kathleen E. Wage
Peter F. Worcester

Term to 2009
Mohsen Badiey
Michael J. Buckingham
Dezhang Chu
John A. Colosi
Christian de Moustier
Stan E. Dosso
Kenneth G. Foote
D. Vance Holliday
Andone C. Lavery
Zoi-Heleni Michalopoulou
Jeffrey A. Nystuen
David R. Palmer
Simon D. Richards
Martin Siderius
Aaron M. Thode

Ex officio:
Peter F. Worcester, member of Medals and Awards Committee
Kenneth G. Foote, member of Membership Committee
Anthony P. Lyons, member of ASACOS
Lora J. Van Uffelen, member of Student Council

Animal Bioacoustics

Richard R. Fay, Chair to 2009

Term to 2011
Judith C. Brown
Sheryl L. Coombs
Edmund R. Gerstein
Charles R. Greene
Darlene R. Ketten
Bertel Mohl
Cynthia F. Moss
Paul E. Nachtigall
Peter L. Tyack
David G. Zeddies

Term to 2010
Michael L. Dent
James J. Finneran
Mardi C. Hastings
Charlotte W. Kotas
Marc O. Lammers
David A. Mann

Marie A. Roch
Gary J. Rose
Joseph A. Sisneros
Term to 2009
John R. Buck
Christopher W. Clark
William C. Cummings
Robert J. Dooling
David K. Mellinger
Arthur N. Popper
Annemarie Surlykke
Ex officio:
James M. Simmons, member of Medals and Awards Committee
Richard R. Fay, member of Membership Committee
Ann E. Bowles, member of ASACOS
Alison Stimpert, member of Student Council

Architectural Acoustics

Brandon D. Tinianov, Chair to 2010

Term to 2011
Wolfgang Ahnert
Robert B. Astrom
Seth E. Bard
Christopher N. Blair
John S. Bradley
Christopher N. Brooks
Angelo J. Campanella
Zhixin Chen
Quinsan Ciao
Robert C. Coffeen
Russell A. Cooper
David B. Copeland
Peter D'Antonio
Felicia M. Doggett
Bill Dohn
Timothy J. Foulkes
Ronald R. Freiheit
Richard D. Godfrey
Matthew V. Golden
Tyrone Hunter
Clare M. Hurtgen
J. Christopher Jaffe
Yun Jing
Thomas E. Kaytt
Jeff P. Kwolkoski
Stephen J. Lind
David Lubman
Ralph T. Muehleisen
Matthew L. Nickerson
Boaz Rafaely
Jack E. Randorff
Jonathan Rathsam
H. Stanley Roller
Steven R. Ryherd
B. Schulte-Fortkamp
Kerrie G. Standlee
Noral D. Stewart
Michelle C. Vigeant
Michael Vorländer
Term to 2010
Russell L. Altermatt
Russ Berger
Warren E. Blazier
Joseph F. Bridger
Norm Broner
Bennett M. Brooks
Steven M. Brown
Richard H. Campbell
Andrew C. Carballeira
Emily L. Cross
F. M. del Solar Dorrego
Erin L. Dugan
Edward C. Duncan
Jesse J. Ehnert
Ronald T. Eligator
Michael Ermann
Adam R. Foxwell
Ronald R. Freiheit
Klaus Genuit
Kenneth W. Good, Jr.
Lewis S. Goodfriend
Bradford N. Gover
Pamela J. Harght
Mark A. Holden
K. Anthony Hoover
Jerald R. Hyde
Jodi Jacobs
Basel H. Jurdy
David W. Kahn
Brian F.G. Katz
Michael P. Kerr
Jonathan E. Kirchner
Brian J. Landsberger
Martha M. Larson
Jonathan S. Leonard
Gary S. Madaras
Benjamin E. Markham
David L. Moyer
Paul B. Ostergaard
Dennis A. Paoletti
Stephen W. Payne
Richard F. Riedel
Lauren Ronsse
Hari V Savitala
Benjamin C. Seep
Neil A. Shaw
J. Michael Spencer
Rose Mary Su
Jeff D. Szymanski
Richard H. Talaske
Michelle Vigeant
Alicia J. Wagner
Ewart A. Wetherill
George E. Winzer
Term to 2009
Nils-Ake Andersson
C. Walter Beamer, IV
Leo L. Beranek
Sergio Beristain
Jim X. Borzym
Erica Bowden
David T. Bradley
David Braslau
Todd L. Brooks
Courtney B. Burroughs
Paul T. Calamia

Alexander U. Case
William J. Cavanaugh
Dan Clayton
Jessica S. Clements
Elizabeth A. Cohen
David A. Conant
Damian Doria
John Erdreich
Robin S. Glosemeyer
Timothy E. Gulsrud
Byron W. Harrison
Robert D. Hellweg
Murray R. Hodgson
Ian B. Hoffman
Jin Yong Jeon
Jian Kang
Bertram Y. Kinzey, Jr.
Mendel Kleiner
Alexis D. Kurtz
Timothy W. Leishman
Jerry G. Lilly
Edward L. Logsdon
Peter A. Mapp
David E. Marsh
Gregory A. Miller
Hideo Miyazaki
Matthew A. Nobile
Christian Nocke
Bruce C. Olson
Cornelius H. Overweg
Richard J. Peppin
Stephen D. Pettyjohn
Scott D. Pfeiffer
Norman H. Philipp
James E. Phillips
Jens Holger Rindel
Carl J. Rosenberg
Kenneth P. Roy
Hiroshi Sato
Melvin L. Saunders
Ron Sauro
Paul D. Schomer
Kevin P. Shepherd
Yasushi Shimizu
Gary W. Siebein
Abigail E. Stefaniw
Christopher A. Storch
Jason E. Summers
Louis C. Sutherland
Jiri Tichy
Nancy S. Timmerman
Gregory C. Tocci
Alfred C.C. Warnock
George P. Wilson
Ning Xiang
Ex officio:
Dana S. Hougland, member of Medals & Awards
Ronald R. Freiheit, member of Membership Com.
George E. Winzer, member of ASACOS
Lauren Ronsse, member of Student Council

Biomedical Ultrasound/Bioresponse to Vibration

Jeffrey A. Ketterling, Chair to 2011

Term to 2011
Michalakis A. Averkiou
Timothy A. Bigelow
Parag V. Chitnis
Saurabh Datta
Sara Davis
Caleb H. Farny
Jonathan Mamou
Stuart B. Mitchell
Neil R. Owen
Oleg A. Sapozhnikov
Armen Sarvazyan
Mark E. Schafer
Eleanor P. Stride
Jahan Tavakkoli
Gail R. ter Haar
Matthew W. Urban
Keith A. Wear
Mark S. Wochner
Term to 2010
Constantin-C. Coussios
Mathias Fink
Sheryl M. Gracewski
Seyed H.R. Hosseini
Ronald E. Kumon
Pascal P. Laugier
T. Douglas Mast
Thomas J. Matula
Robert J. McGough
James A. McAteer
James G. Miller
Todd W. Murray
Michael L. Oelze
Neil R. Owen
Ronald A. Roy
Preston S. Wilson
Suk Wang Yoon
Evgenia A. Zabolotskaya
Term to 2009
Emmanuel Bossy
Anthony J. Brammer
Robin O. Cleveland
Mostafa Fatemi
Yuri A. Ilinskii
Peter J. Kaczkowski
Vera A. Khokhlova
Oliver D. Kripfgans
James C. Lacefield
Yuri A. Pishchalnikov
Tyrone M. Porter
Ex officio:
Lawrence A. Crum, member of the Medals and Awards Committee
Christy K. Holland, member of Membership Committee
Peter J. Kaczkowski, member of ASACOS
Lucie Somaglino, member of Student Council

Engineering Acoustics

Daniel M. Warren, Chair to 2009

Term to 2011
Steven R. Baker
David A. Brown

Stephen C. Butler
Robert D. Corsaro
Robert M. Drake
Stephen E. Forsythe
Brian H. Houston
Robert M. Koch
Christopher C. Lawrenson
L. Dwight Luker
Arnie L. Van Buren
Kenneth M. Walsh
Joseph F. Zalesak

Term to 2010

Mahlon D. Burkhard
Fernando Garcia-Osuna
Charles S. Hayden
Jan F. Lindberg
Elizabeth A. McLaughlin
Alan Powell
Roger T. Richards
Kenneth D. Rolt
Neil A. Shaw
James F. Tressler

Term to 2009

Stanley L. Ehrlich
Gary W. Elko
Robert D. Finch
Guillermo C. Gaunaurd
Thomas R. Howarth
Dehua Huang
Sung Hwan Ko
Victor Nedzelnitsky
James M. Powers
P.K. Raju
Stephen C. Thompson
James E. West
George S. K. Wong

Ex officio:

Kim C. Benjamin, member of Medals and Awards Committee
Thomas R. Howarth, member of Membership Committee
Mahlon D. Burkhard, member of ASACOS
Douglas Wilcox, member of Student Council

Musical Acoustics

Paul A. Wheeler, Chair to 2011

Term to 2011

James W. Beauchamp
George A. Bissinger
Annabel J. Cohen
James P. Cottingham
Evan B. Davis
Diana Deutsch
Uwe J. Hansen
Peter L. Hoekje
James H. Irwin
Ian M. Lindevald
Robert W. Pyle
Gary P. Scavone
Brad H. Story
George Tzanetakis
Christopher E. Waltham

Term to 2010

R. Dean Ayers
Judith C. Brown
Courtney B. Burroughs
John R. Buschert
Joel Gilbert
Thomas M. Huber
Bozena Kostek
Barry Larkin
Daniel O. Ludwigsen
Thomas R. Moore
Thomas D. Rossing
David B. Sharp
Julius O. Smith
William J. Strong
Joe Wolfe

Term to 2009

Rolf Bader
Xavier Boutillon
Jonas Braasch
Murray D. Campbell
Rene E. Causse
Antoine J. Chaigne
Neville H. Fletcher
Nicholas J. Giordano
J. M. Harrison
William M. Hartmann
William L. Martens
James M. Pyne
Daniel A. Russell
Punita G. Singh
Sten O. Ternstrom
Shigeru Yoshikawa

Ex officio:

Uwe J. Hansen, member of Medals and Awards Committee
James P. Cottingham, member of Membership Committee
Ian M. Lindevald, member of ASACOS
Eric Dieckeman, member of Student Council

Noise

Brigitte Schulte-Fortkamp, Chair to 2009

Term to 2011

Seth E. Bard
Elliott H. Berger
Ann E. Bowles
Frank H. Brittain
Steven M. Brown
Mahlon D. Burkhard
Robert D. Collier
Lawrence S. Finegold
Samir N. Y. Gerges
Richard D. Godfrey
Matthew V. Golden
Murray R. Hodgson
Jerry G. Lilly
Stephen J. Lind
David Lubman
George A. Luz
Matthew L. Nickerson
Matthew A. Nobile
Richard J. Peppin
Robert A. Putnam
Jack E. Randorff

Stephen I. Roth
Paul D. Schomer
Michelle E. Swearingen
Nancy S. Timmerman
Brandon D. Tinianov
Gregory C. Tocci
Lily M. Wang

Term to 2010

Martin Alexander
Brian E. Anderson
Keith Attenborough
John P. Barry
Leo L. Beranek
Arno S. Bommer
Dick B. Botteldooren
Giovanni Brambilla
James O. Buntin
John C. Burgess
Jim R. Cummins
Kenneth A. Cunefare
Paul R. Donavan
Andre Fiebig
Ronald R. Freiheit
Klaus Genuit
David C. Haser
Gerald C. Lauchle
George C. Maling
Thomas R. Norris
John P. Seiler
Noral D. Stewart
Louis C. Sutherland
Jiri Tichy
D. Keith Wilson
Ning Xiang
Yuzhen Yang

Term to 2009

Sergio Beristain
Susan B. Blaeser
Erica E. Bowden
Bennett M. Brooks
Ilene J. Busch-Vishniac
Angelo J. Campanella
William J. Cavanaugh
Gilles A. Daigle
Patricia Davies
Damian J. Doria
Connor R. Duke
Jesse J. Ehnert
Tony F. W. Embleton
John Erdreich
David J. Evans
Bradford N. Gover
Robert D. Hellweg
Tyrone Hunter
William W. Lang
Richard H. Lyon
Alan H. Marsh
Ralph T. Muehleisen
William J. Murphy
Kenneth P. Roy
Kevin P. Shepherd
Scott D. Sommerfeldt
Kerrie G. Standlee
George S.K. Wong

Ex officio:

Nancy S. Timmerman, member of Medals and Awards Committee
Bennett M. Brooks, member of Membership Committee
Richard J. Peppin, member of ASACOS
Cole R. Duke, member of Student Council

Physical Acoustics

Ronald A. Roy, Chair to 2010

Term to 2011

Robert T. Beyer
Robin O. Cleveland
Lawrence A. Crum
E. Carr Everbach
Kenneth E. Gilbert
Robert A. Hiller
R. Glynn Holt
Bart Lipkens
Thomas J. Matula
Ralph T. Muehleisen
John S. Stroud
Richard L. Weaver
Preston S. Wilson

Term to 2010

David T. Blackstock
David A. Brown
John A. Burkhardt
Kerry W. Commander
Bruce C. Denardo
Kent L. Gee
Logan E. Hargrove
D. Kent Lewis
Julian D. Maynard
Albert Migliori
James G. Miller
George Mozurkewich
Lev A. Ostrovsky
Andrea Prosperetti
Neil A. Shaw
Victor W. Sparrow
Richard Stern
Michelle E. Swearingen
Roger M. Waxler

Term to 2009

Anthony A. Atchley
Yves H. Berthelot
James P. Chambers
Charles C. Church
Nico Declercq
Kenneth G. Foote
Mark F. Hamilton
David I. Havelock
Philip L. Marston
Peter H. Rogers
James M. Sabatier
Philip S. Spoor
Larry A. Wilen
D. Keith Wilson
Evgenia A. Zabolotskaya

Ex officio:

Thomas J. Matula, member of Medals and Awards Committee
Steven L. Garrett, member of Membership Committee
Richard Raspet, member of ASACOS

Todd A. Hay, member of Student Council

Psychological and Physiological Acoustics

Andrew J. Oxenham, Chair to 2011

Term to 2011
Huanping Dai
Christian Lorenzi
Christophe D. Micheyl
Roy D. Patterson
Daniel Pressnitzer
Brian Roberts
Term to 2010
Sid P. Bacon
Qian-Jie Fu
Ruth Y. Litovsky
Robert A. Lutfi
Kim S. Schairer
Christopher Shera
Edward J. Walsh
Beverly A. Wright
Term to 2009
Katherine H. Aerhart
Richard Freyman
Amy R. Horwitz
Jody E. Kreiman
Glenis R. Long
Enrique A. Lopez-Poveda
Robert S. Schlauch
Stanley E. Sheft
Ex officio:
William M. Hartmann, member of the Medals and Awards Committee
Lynne A. Werner, member of Membership Committee
Brent W. Edwards, member of ASACOS
Suzanne P. Carr, member of Student Council

Signal Processing in Acoustics

David H. Chambers, Chair to 2009

Term to 2011
Joseph A. Clark
David I. Havelock
Jean-Pierre Hermand
George E. Ioup
Juliette W. Ioup
Sean K. Lehman
William L. Martens
Zoi-Heleni Michalopoulou
Natalia A. Sidorovskaya
Kevin B. Smith
David C. Waddington
Ning Xiang
Term to 2010
Max Deffenbaugh
Alireza A. Dibazar
Gary W. Elko
Kassiani Kotsidou
Patrick J. Loughlin
Alan W. Meyer
Daniel J. Sinder
David C. Swanson
Robert C. Waag
Preston S. Wilson
Lixue Wu
Term to 2009
James V. Candy
William M. Carey
Leon Cohen
Geoffrey S. Edelson
Stanley L. Ehrlich
Brian Ferguson
Paul J. Gendron
Peter Gerstoft
William M. Hartmann
Kevin D. Heaney
William S. Hodgkiss
Paul D. Hursky
John M. Impagliazzo
Jens M. Meyer
Hassan Namarvar
Joe W. Posey
James C. Preisig
Brian Rapids
Edmund J. Sullivan
Ex officio:
William J. Carey, member of Medals and Awards Committee
Ning Xiang, member of Membership Committee
Charles F. Gaumond, member of ASACOS
John H. Camin, member of Student Council

Speech Communication

Carol Espy-Wilson, Chair to 2010

Term to 2011
Fredericka Bell-Berti
Ann R. Bradlow
Kate E. Bunton
Dani M. Byrd
Rebeka Campos-Astorkiza
Rachel Frush Holt
Zhaoyan Zhang
Term to 2010
Augustine Agwuele
Patrice S. Beddor
Tessa C. Bent
Helen M. Hanson
Diane Kewley-Port
Jody E. Kreiman
Andrew J. Lotto
Shrikanth S. Narayanan
Dwayne Paschall
Christine H. Shadle
Rahul Shrivastav
Term to 2009
Jean E. Andruski
Lynne E. Bernstein
Ocke-Schwen Bohn
Suzanne E. Boyce
Ann R. Bradlow
Bruce R. Gerratt
Kenneth W. Grant
Paul E. Iverson
Benjamin R. Munson
Peggy B. Nelson
Mitchell S. Sommers
Brad H. Story
Joan E. Sussman

Ex officio:
Anders Lofqvist, member of Medals and Awards Committee
Jody E. Kreiman, member of Membership Committee
Shrikanth S. Narayanan, member of ASACOS
Christian Stilp, member of Student Council

Structural Acoustics and Vibration

Sean F. Wu, Chair to 2009

Term to 2011
Dean E. Capone
Joel M. Garrelick
Peter C. Herdic
Teik C. Lim
Thomas J. Royston
Angie Sarkissian
Jeffrey S. Vipperman
Richard L. Weaver
Term to 2010
Jeffrey E. Boisvert
Stephen C. Conlon
Linda P. Franzoni
Robert C. Haberman
Rudolph Martinez
Koorosh Naghshineh
Donald G. Pray
Michael F. Shaw
Term to 2009
Joseph M. Cuschieri
David Feit
Sabih I. Hayek
Philip L. Marston
James E. Phillips
Earl G. Williams
Ex officio:
Courtney B. Burroughs, member of Membership Committee
Mauro Pierucci, member of Medals and Awards Committee
Sabih I. Hayek, member of ASACOS
Brian T. Thornock, member of Student Council

Underwater Acoustics

Kevin LePage, Chair to 2009

Term to 2011
Micheal A. Ainslie
Juan I. Arvelo, Jr.
Pierre-Philippe Beaujean
Shira L. Broschat
Nicholas P. Chotiros
Richard L. Culver
Claire Debever
Geoffrey F. Edelmann
Peter Gerstoft
John H. Glattetre
Brian T. Hefner
Jean-Pierre Hermand
Charles W. Holland
Shawn F. Johnson
David P. Knobles
Bruce K. Newhall
John C. Osler
Kevin B. Smith
Alexandra I. Tolstoy
Joshua D. Wilson
Term to 2010
Kyle M. Becker
David C. Calvo
Jee Woong Choi
Christian P. de Moustier
Stan E. Dosso
Steven I. Finette
Kenneth G. Foote
David Fromm
Christohper H. Harrison
Zoi-Heleni Michalopoulou
Tracianne B. Neilsen
Robert I. Odom
Marshall H. Orr
Gregory J. Orris
James C. Preisig
Martin Siderius
William L. Siegmann
Jixun Zhou
Term to 2009
Ralph N. Baer
John R. Buck
Chi-Fang Chen
David R. Dowling
Roger C. Gauss
Frank S. Henyey
Paul C. Hines
Chen-Fen Huang
Marcia J. Isakson
Finn B. Jensen
Sunwoong Lee
James H. Miller
John R. Preston
Purnima Ratilal
Karim G. Sabra
John B. Schneider
Ralph A. Stephen
Alexander G. Voronovich
Kevin L. Williams
Lisa M. Zurk
Ex officio:
Henrik Schmidt, member of Membership Committee
Eric I. Thorsos, member of Medals and Awards Committee
Robert M. Drake, member of ASACOS
Megan Ballard, member of Student Council

Administrative Committees 2008—2009

Archives and History

Julian D. Maynard, Chair to 2010

Term to 2011
Anthony A. Atchley
Leo L. Beranek
William J. Cavanaugh
Steven L. Garrett
Logan E. Hargrove
Allan D. Pierce
Victor W. Sparrow
Term to 2010
Carol Y. Espy-Wilson
E. Carr Everbach

William W. Lang
Richard Stern
David C. Swanson
Term to 2009
Jont B. Allen
David I. Havelock
Wesley L. Nyborg
Richard J. Peppin
William J. Strong

Audit

Wayne M. Wright, Chair to 2009

Term to 2011
Peter H. Dahl
Term to 2010
Michael R. Stinson

Books[+]

David L. Bradley, Chair to 2011

Term to 2011
James P. Cottingham
Jeffrey A. Nystuen
Neil A. Shaw
Term to 2010
Richard Stern
Brandon D. Tinianov
Term to 2009
Juan I. Arvelo
Jerry H. Ginsberg
Philip L. Marston
Ex officio:
Allan D. Pierce, Editor-in-Chief

College of Fellows Steering

Diane Kewley-Port and Stephen C. Thompson,
Acting Cochairs to 2010

Term to 2010
Peter G. Cable
M. David Egan
Uwe J. Hansen
Diane Kewley-Port
Thomas D. Rossing
Term to 2009
Thomas J. Matula
Scott D. Sommerfeldt
Stephen C. Thompson
Beverly A. Wright
William J. Cavanaugh, ex officio as past Chair
Richard H. Lyon, ex officio as past Chair

Education in Acoustics

James M. Sabatier, Chair to 2009

Term to 2011
David T. Blackstock
Courtney B. Burroughs
James P. Chambers
Robin O. Cleveland
Kenneth A. Cunefare
D. Michael Daly
Mary Florentine
R. Glynn Holt
Murray S. Korman
Luc Mongeau
Peggy B. Nelson
Tyrone M. Porter
Neil A. Shaw
Roger Waxler
James E. West
Douglas Wilcox
Term to 2010
George A. Bissinger
David A. Brown
Stanley A. Cheyne
Robert D. Collier
Lawrence A. Crum
Corinne M. Darvennes
Bruce C. Denardo
Margaritis S. Fourakis
Carl K. Frederickson
Carole E. Gelfer
Daniel O. Ludwigsen
Sharon Y. Manuel
Philip L. Marston
Ralph T. Muehleisen
Andrew A. Piacsek
Thomas D. Rossing
Ronald A. Roy
Dawn R. Schuette
Scott D. Sommerfeldt
Robert A. Walkling
George S.K. Wong
Term to 2009
William A. Ahroon
Takayuki Arai
Anthony A. Atchley
Fredericka Bell-Berti
Suzanne E. Boyce
Robert D. Celmer
Annabel J. Cohen
E. Carr Everbach
Thomas B. Gabrielson
Steven L. Garrett
Kent L. Gee
Uwe J. Hansen
Katherine S. Harris
Elizabeth S. Ivey
Joie P. Jones
Amy T. Neel
P. K. Raju
Deborah M. Rekart
Daniel A. Russell
M. Roman Serbyn
Victor W. Sparrow
Emily A. Tobey

Ethics and Grievances

Janet M. Weisenberger, Chair to 2011

Term to 2009
William J. Cavanaugh
Gerald L. D'Spain
Barbara G. Shinn-Cunningham

International Research and Education

David T. Blackstock, Chair to 2010

Term to 2011
Keith Attenborough
Sergio Beristain
Vera A. Khokhlova
Brigitte Schulte-Fortkamp
Term to 2010
Malcolm J. Crocker
Samir N.Y. Gerges
Konstantin A. Naugolnykh
Oleg Sapozhnikov
Michael Vorländer
Term to 2009
Gilles A. Daigle
Lily M. Wang
Suk Wang Yoon

Investments

William A. Yost, Chair to 2010

Term to 2011
Anthony A. Atchley
William M. Harmtmann
Term to 2009
Ilene J. Busch-Vishniac
David Feit, Treasurer, ex officio

Medals and Awards

David L. Bradley, Chair to 2009

Term to 2011	
Dana S. Hougland	Architectural Acoustics
Mauro Pierucci	Structural Acoustics and Vibration
James A. Simmons	Animal Bioacoustics
Eric I. Thorsos	Underwater Acoustics
Term to 2010	
Kim C. Benjamin	Engineering Acoustics
William M. Carey	Signal Processing in Acoustics
Uwe J. Hansen	Musical Acoustics
Anders Lofqvist	Speech Communication
Peter F. Worcester	Acoustical Oceanography
Term to 2009	
Lawrence A. Crum	Biomedical Ultrasound/Bioresponse to Vibration
William M. Hartmann	Psychological and Physiological Acoustics
Thomas J. Matula	Physical Acoustics
Nancy S. Timmerman	Noise

Meetings

Clark S. Penrod, Chair to 2011 and Fall 2009
Harry A. DeFerrari, Fall 2008, Miami
Fred C. DeMetz, Fall 2007, New Orleans
Diane Kewley-Port, Vice President-Elect
Elaine Moran, ASA Office Manager, ex officio
Charles E. Schmid, Executive Director, ex officio
Victor W. Sparrow, Vice President
William A. Yost, Spring 2008, Paris
Lisa Zurk, Spring 2009, Portland
to be appointed, Spring 2010, Baltimore

Membership

Peter H. Rogers, Chair to 2009

Term to 2011	
Bennett M. Brooks	Noise
Courtney B. Burroughs	Structural Acoustics and Vibration
Christy K. Holland	Biomedical Ultrasound/Bioresponse to Vibration
Term to 2010	
Kenneth G. Foote	Acoustical Oceanography
Ronald R. Freiheit	Architectural Acoustics
Steven L. Garrett	Physical Acoustics
Ning Xiang	Signal Processing in Acoustics
Henrik Schmidt	Underwater Acoustics
Term to 2009	
James P. Cottingham	Musical Acoustics
Thomas R. Howarth	Engineering Acoustics
Jody E. Kreiman	Speech Communication
Richard R. Fay	Animal Bioacoustics
Lynne A. Werner	Psychological and Physiological Acoustics

Panel on Public Policy

Edward J. Walsh, Chair to 2010

Term to 2011
William M. Hartmann
Sabih I. Hayek
Peggy B. Nelson
Victor W. Sparrow
Term to 2010
Angelo J. Campanella
Lawrence A. Crum
Ellen S. Livingston
James H. Miller
Term to 2009
Mardi C. Hastings
David Lubman
William A. Yost
Gilles A. Daigle, ex officio as immediate Past President
George V. Frisk, ex officio as Vice President
Charles E. Schmid, ex officio as Executive Director
Paul D. Schomer, ex officio as Standards Director

Public Relations

Geoffrey F. Edelmann, Chair to 2009

Term to 2011
William M. Carey
Katherine H. Kim
Bart Lipkens
Ellen S. Livingston
Andrew A. Piascek
Brigitte Schulte-Fortkamp
Term to 2010
Kelly J. Benoit-Bird
Diana Deutsch
E. Carr Everbach
Christy K. Holland
Brenda L. Lonsbury-Martin
James H. Miller
Joe W. Posey
Barbara Shinn-Cunningham
Stephen C. Thompson

Term to 2009
Paul A. Baxley
Ann E. Bowles
Paul D. Hursky
Jack E. Randorff
Barbara J. Sotirin
Lora J. Van Uffelen
Kathleen E. Wage
Allan D. Pierce, Editor-in-Chief, ex officio
Elaine Moran, ASA Office Manager, ex officio
Charles E. Schmid, Executive Director, ex officio
Thomas D. Rossing, Echoes Editor, ex officio

Publication Policy

Brenda L. Lonsbury-Martin, Chair to 2009

Term to 2011
James F. Lynch
Philip L. Marston
Term to 2010
Charles C. Church
Mardi C. Hastings
Term to 2009
Jont B. Allen
David I. Havelock
Whitlow W. L. Au, President-Elect, ex officio
Allan D. Pierce, Editor-in-Chief, ex officio

Regional Chapters

Juan I. Arvelo, Cochair to 2011
Elizabeth A. McLaughlin, Cochair to 2011

Kent L. Gee	Brigham Young Univ. Student Chapter
Angelo J. Campanella	Central Ohio
Robert M. Keolian	Central Pennsylvania
Ernest M. Weiler	Cincinnati
Rebecca Mercuri	Delaware Valley
Gary W. Siebein	Florida
Shaun D. Anderson	Georgia Institute of Tech. Student Chapter
Timothy J. Foulkes	Greater Boston
Robert D. Celmer	Univ. of Hartford Student Chapter
Michael J. Anderson	Inland Northwest
Robert C. Coffeen	Univ. of Kansas Student Chapter
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Richard F. Riedel	Metropolitan New York
Sergio Beristain	Mexico City
William V. Slaton	Mid-South
Roger T. Richards	Narragansett
Michelle Vigeant	Univ. of Nebraska Student Chapter
Richard F. Riedel	New York
George A. Bissinger	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
David Lubman	Orange County
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
Juan I. Arvelo	Washington, D.C.
Thomas M. Disch	Wisconsin

James M. Sabatier, Chair, Education in Acoustics, ex officio
David Feit, Treasurer, ex officio
Cole Duke, Student Council representative, ex officio

Rules and Governance

William M. Hartmann, Chair to 2011

Term to 2011
Anthony A. Atchley
Donna L. Neff
Term to 2010
William J. Cavanaugh
Floyd Dunn
Term to 2009
Elaine Moran
Charles E. Schmid

Prizes and Special Fellowships

Wayne M. Wright, Chair to 2009

Term to 2011
Uwe J. Hansen
Term to 2010
Anthony A. Atchley
Constantine Trahiotis
Term to 2009
Fredericka Bell-Berti
James E. West

Standards

Executive Committee
Paul D. Schomer, Chair (Standards Director)
Robert D. Hellweg, Vice Chair
Susan B. Blaeser, Standards Manager, ex officio

S1 Representation
Philip J. Battenberg, Chair S1 and ASA rep. on S1
Richard J. Peppin, Vice Chair S1
Alan H. Marsh, ASA rep. on S1
Paul D. Schomer, ASA alternate rep. on S1
S2 Representation
Ronald L. Eshleman, Chair S2
Ali T. Herfat, Vice Chair S2
Sabih I. Hayek, ASA rep. on S2
Bruce E. Douglas, ASA alternate rep. on S2

S3 Representation
Craig A. Champlin, Chair S3 and ASA rep. on S3
David A. Preves, Vice Chair S3
Mahlon D. Burkhard, ASA alternate rep. on S3

S12 Representation
Robert D. Hellweg, Chair S12
William J. Murphy, Vice Chair S12
Bennett M. Brooks, ASA rep. on S12
David Lubman, ASA alternate rep. on S12

International TAGs (ex officio)
Paul D. Schomer, Chair, U.S. TAG for ISO/TC 43 and ISO/TC 43/SC1
David J. Evans, Chair, U.S. TAG for ISO/TC 108
Victor Nedzelnitsky, U.S. Technical Advisor for IEC/TC 29

ASA Technical Committee Representatives
Victor W. Sparrow, Chair of ASA Technical Council, ex officio
Anthony P. Lyons, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
Angelo J. Campanella, Architectural Acoustics
Peter J. Kaczkowski, Biomedical Ultrasound/
Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Diana Deutsch, Musical Acoustics
Richard J. Peppin, Noise
Richard Raspet, Physical Acoustics
Brent W. Edwards, Psychological and Physiological Acoustics
Charles F. Gaumond, Signal Processing in Acoustics
Shrikanth S. Narayanan, Speech Communication
Sabih I. Hayek, Structural Acoustics and Vibration Robert M. Drake,
Underwater Acoustics

ASA Officers
David Feit, Treasurer, ex officio
Charles E. Schmid, Executive Director, ex officio

Past Chair of ASACOS (ex officio)
Tony F. W. Embleton

Associate Editors for Standards News–JASA (ex officio)
Susan B. Blaeser
Paul D. Schomer

Student Council

Michael Canney	Chair
Megan Ballard	Underwater Acoustics
John Camin	Signal Processing in Acoustics
Suzanne P. Carr	Psychological and Physiological Acoustics
Eric Dieckeman	Musical Acoustics
Cole R. Duke	Noise and Regional Chapters Committee Liaison
Todd A. Hay	Physical Acoustics
Lauren Ronsse	Architectural Acoustics
Lucie Somaglino	Biomedical/Bioresponse
Christian Stilp	Speech Communication
Alison Stimpert	Animal Bioacoustics
Brian Thornock	Structural Acoustics and Vibration
Lora J. van Uffelen	Acoustical Oceanography
Douglas Wilcox	Engineering Acoustics

Tutorials

Lily Wang, Chair to 2009

Term to 2011
Kenneth A. Cunefare
David R. Dowling
Barbara G. Shinn-Cunningham
Term to 2010
Paul E. Barbone
Micheal L. Dent
Michelle E. Swearingen
Term to 2009
Ann R. Bradlow
James V. Candy
James P. Chambers
Charles E. Schmid, Executive Director, ex officio

Women in Acoustics

Lisa M. Zurk, Chair to 2009

Term to 2011
Judy R. Dubno
Mardi C. Hastings
Andone C. Lavery
Tracianne Nelson
Lily M. Wang
Term to 2010
Helen M. Hanson
Marcia J. Isakson
Jennifer L. Miksis-Olds
Peggy B. Nelson
Term to 2009
Kathryn W. Hatlestad
Carolyn J. Richie
Diane Kewley-Port, ex officio as Vice President-Elect

Acoustics Today Editorial Board

Richard Stern, Chair
Elliott H. Berger
Carol Espy-Wilson
K. Anthony Hoover
Allan D. Pierce
Thomas D. Rossing
Brigitte Schulte-Fortkamp

External Affairs Council

Peggy B. Nelson, Chair
Michael R. Stinson, Vice Chair
Peter H. Dahl, Member
Whitlow W. L. Au, President-Elect
Allan D. Pierce, Editor-in-Chief
Charles E. Schmid, Executive Director
David T. Blackstock
David L. Bradley
Geoffrey F. Edelmann
Clark S. Penrod
Edward J. Walsh
Lily M. Wang

Internal Affairs Council

Wayne M. Wright, Chair
Fredericka Bell-Berti, Vice Chair
Andrea M. Simmons, Member
Gilles A. Daigle, Past President
Charles E. Schmid, Executive Director
Paul D. Schomer, Standards Director
Juan I. Arvelo
Michael S. Canney
William M. Hartmann
Julian D. Maynard
Elizabeth A. McLaughlin
Janet M. Weisenberger

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer
Physics Department, Georgetown University, Washington, DC 20057

International Meetings Calendar

Below are announcements of meetings and conferences to be held abroad. Entries preceded by an * are new or updated listings.

October 2008

3–5 **7th International Conference on Auditorium Acoustics**, Oslo, Norway (ioa.org.uk).

6–8 **Acoustics Week in Canada**, Vancouver, B.C., Canada (www.caa-aca.ca).

14–15 **Underwater Noise Measurement, Impact and Mitigation**, Southampton, UK (www.ioa.org.uk/viewupcoming.asp).

21–22 **Institute of Acoustics (UK) Autumn Conference 2008**, Oxford, UK (ioa.org.uk).

21–23 **International Conference on Low Frequency Noise and Vibration**, Tokyo, Japan (www.lowfrequency2008.org).

21–24 **acústica 2008**, Coimbra, Portugal (www.spacustica.pt).

26–29 **inter-noise 2008**, Shanghai, China (www.internoise2008.org).

27–31 **XX Session of the Russian Acoustical Society**, Moscow, Russia (rao.akin.ru).

30–31 **Autumn Conference of the Swiss Acoustical Society**, Lausanne, Switzerland (www.sga-ssa.ch).

November 2008

2–5 **IEEE International Ultrasonics Symposium**, Beijing, China (www.ieee.org/conf/ius_2008).

5–7 **Iberoamerican Acoustics Congress (FIA 2008)**, Buenos Aires, Argentina (www.adaa.org.ar).

14–18 **20th Session of the Russian Acoustical Society**, Moscow, Russia (www.akin.ru).

20–21 **Reproduced Sound 24**, Brighton, UK (ioa.org.uk).

24–26 **Australian Acoustical Society National Conference**, Geelong, Vic., Australia (www.acoustics.asn.au).

30–5 **18th Biennial Congress of the Australian Institute of Physics**, Adelaide, Australia (www.aipc2008.com).

December 2008

9–12 **19th International Acoustic Emission Symposium**, Kyoto, Japan (www.19iaes-kyoto.com/index.html).

17–19 **Symposium on the Acoustics of Poro-Elastic Materials (sapem 2008)**, Bradford, UK (sapem2008.matelys.com).

January 2009

11–17 **International Congress on Ultrasonics**, Santiago, Chile (www.icu2009.usach.cl).

March 2009

23–26 **International Conference on Acoustics (NAG/DAGA 2009)**, Rotterdam, The Netherlands (www.nag-daga.nl).

April 2009

5–9 **Noise and Vibration: Emerging Methods (NOVEM 2009)**, Oxford, UK (www.isvr.soton.ac.uk/NOVEM2009).

13–17 **2nd International Conference on Shallow Water Acoustics**, Shanghai, China ([soon]:www.apl.washington.edu).

19–24 **International Conference on Acoustics, Speech, and Signal Processing**, Taipei. R.O.C. (icassp09.com).

May 2009

19–22 *__XXI Session of the Russian Acoustical Society__, Moscow, Russia (www.akin.ru/main.htm).

June 2009

17–19 *__3rd International Conference on Wind Turbine Noise__, Aalborg, Denmark (www.windturbinenoise2009.org)

July 2009

5–9 **16th International Congress on Sound and Vibration**, Krakow, Poland (www.icsv16.org).

August 2009

23–28 **Inter-noise 2009**, Ottawa, Ont., Canada (www.internoise2009.com).

September 2009

6–10 **InterSpeech 2009**, Brighton, UK (www.interspeech2009.org).

14–18 *__5th Animal Sonar Symposium__, Kyoto, Japan (cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html).

15–17 *__Autumn Meeting of the Acoustical Society of Japan__, Koriyama, Japan (www.asj.go.jp/index-en.html).

19–23 **IEEE 2009 Ultrasonics Symposium**, Rome, Italy (e-mail: pappalar@uniroma3.it).

23–25 **TECNIACUSTICA2009**, Cádiz, Spain (www.sea-acustica.es).

October 2009

26–28 **Euronoise 2009**, Edinburgh, UK (www.euronoise2009.org.uk).

June 2010

13–16 **INTERNOISE2010**, Lisbon, Portugal, (www.internoise2010.org).

August 2010

23–27 **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (www.ica2010sydney.org).

September 2010

26–30 **Interspeech 2010**, Makuhari, Japan (www.interspeech2010.org).

August 2011

27–31 *__Interspeech 2011__, Florence, Italy (www.interspeech2011.org).

September 2011

4–7 *__International Congress on Ultrasonics__, Gdansk, Poland.

 0001-4966/2008/124(4)/1897/1/$23.00

BOOK REVIEWS

Malcolm J. Crocker
John Wiley & Sons, New Jersey, 2007

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

Handbook of Noise and Vibration Control

Malcolm J. Crocker

John Wiley & Sons, New Jersey, 2007 1584 pp. Price: $195.00 (hardcover) ISBN: 0471395994

The intent of this book appears to be to serve as a complete reference for those working in the field of noise and vibration control. In terms of the number of topics covered and the sheer amount of information provided, it probably largely succeeds in accomplishing that objective. The editor has put together an impressive list of authors (148 contributing authors) to write the chapters in the handbook. These individuals are well known in their fields and the result is a level of expertise represented in the chapters that cannot be faulted.

The book consists of 130 chapters divided into 11 major sections: Fundamentals of acoustics and noise; Fundamentals of vibration; Human hearing and speech; Effects of noise, blast, vibration, and shock on people; Noise and vibration transducers, analysis equipment, signal processing, and measuring techniques; Principles of noise and vibration control and quiet machinery design; Industrial and machine element noise and vibration sources: Prediction and control; Transportation noise and vibration: Sources, prediction, and control; Interior transportation noise and vibration sources: Prediction and control; Noise and vibration control in buildings; and Community and environmental noise and vibration prediction and control.

The first four sections of the handbook are largely background material, with the first two sections covering the fundamentals of acoustics and vibration, and the next two sections essentially covering the effects of noise and vibration on people. Together, these sections provide almost as many pages (about 400) as many textbooks on the fundamentals of acoustics and vibration. However, in the opinion of this reviewer, the presentation of this material is not as effective as it could be, given the number of pages devoted to the subject. The subject matter is tutorial in nature and would be of greatest use to those who have had little exposure to acoustics and vibration. This type of material is often most effectively presented in a developmental style of textbooks which builds on previous material, provides sufficient details to understand the results presented, and provides appropriate examples of the principles being discussed as part of that development. However, this book is a handbook and is written in that style, with topics being covered in a very concise manner (with details often omitted) and presenting results that often are only loosely tied together. As a result, the material in these first four sections is largely unnecessary for those already familiar with the subject matter and is likely rather difficult to follow for those who have not been exposed to the material previously. In addition, there is a discrepancy in the clarity and usefulness of the various chapters in these sections, at least partly due to multiple authors contributing to the work. Some chapters are very clear in their writing style and serve their purpose well. Other chapters are not nearly as well written and, in some cases, will be of little help to the reader unless he/she is already familiar with the subject matter or supplements the material in this handbook with other sources. (This discrepancy in clarity and usefulness appears in multiple places throughout the handbook and is probably largely unavoidable when a large number of authors are utilized.)

The handbook is arranged in a sort of hierarchical fashion. Each of the 11 sections begins with an introductory chapter on the subject covered in that section, generally written by the editor of the handbook. This introductory chapter overviews the subject matter for that section of the handbook, identifies some of the basic issues to be covered in the chapters that follow in that section, and generally provides a brief synopsis of some of the more important results that are presented in greater detail in each of the succeeding chapters. This would possibly seem a little redundant if the reader is reading the handbook from "cover to cover," but actually provides a potentially nice resource for someone looking for a more high-level review of the subject matter without getting into many of the details. For such people, these introduction chapters would be very useful. If additional information and details are needed, then that can be found in the additional chapters that follow the introduction for each section.

As one example of how this structure is organized, Part VIII covers the subject area of "Transportation Noise and Vibration: Sources, Prediction, and Control" and focuses on external noise radiated from various transportation vehicle sources. The introductory chapter very briefly introduces such topics as noise emission, internal combustion engines, muffler design, tire/road noise, aerodynamic sources, gear noise, jet engine noise, and aircraft propeller noise, including helicopter rotor noise. After the introductory chapter, there are ten additional chapters covering these topics: Internal combustion engine noise prediction and control: Diesel and gasoline engines; Exhaust and intake noise and acoustical design of mufflers and silencers; Tire/road noise: Generation, measurement, and abatement; Aerodynamic sound sources in vehicles: Prediction and control; Transmission and gearbox noise and vibration prediction and control; Jet engine noise generation, prediction, and control; Aircraft propeller noise: Sources, prediction, and control; Helicopter rotor noise: Generation, prediction, and control; Brake noise prediction and control; and Wheel-rail interaction noise prediction and its control. It can be seen from these topics that a broad range of transportation vehicles are covered (ships are covered in another section). This type of structure is typical of all the sections in the handbook.

In many of the chapters (throughout the entire handbook), there is a fairly heavy focus on describing the noise radiation and vibration mechanisms associated with various noise and vibration sources. Some, but not all, of the chapters then go beyond that to outline potential noise control solutions or to identify how to make use of the material presented in the chapter. One example of this is Chap. 50, which covers the topic "Use of Near-Field Holography in Noise and Vibration." After developing the basic theory of near-field acoustic holography (NAH), some of the difficulties that arise, and various developments to try and address those difficulties, the author includes a section that provides a "cookbook recipe" for implementing NAH, which would be very helpful to assist someone new to the field in implementing this tool in a noise control application. There are a number of other chapters that are also very good at outlining the basic theory and results associated with their respective topics, as well as clear and concise guidance on how to implement the material in practical noise control applications.

Unfortunately, there are also some chapters that are not successful in providing a clear overview of the subject matter along with guidance on how it can be used in practice. In some cases, those chapters present material that is fairly well written and provide some useful information to help the reader understand the noise source addressed in the chapter, but which will be of little help to a noise control engineer in determining how to apply the material to a practical noise control application. In a few cases, the chapters are written at a level and/or in a style that may be of little help to the reader unless he/she is already knowledgeable in the field. Fortunately, these chapters represent a minority of the chapters, so I do not want to dwell on these few weaknesses at the expense of overlooking all that is good about the handbook.

In conclusion, there is a wealth of information included in this handbook, covering a very wide range of topics. Much of it is well written, and the book would serve as a good resource to those in the field of noise and vibration control. While there are a few weaknesses in certain parts of the handbook, it would certainly serve as a very useful reference in helping to solve problems encountered or to at least develop some of the understanding needed to make effective use of other resources available.

SCOTT D. SOMMERFELDT
Brigham Young University,
Provo, UT 84602

 0001-4966/2008/124(4)/1898/1/$23.00

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:

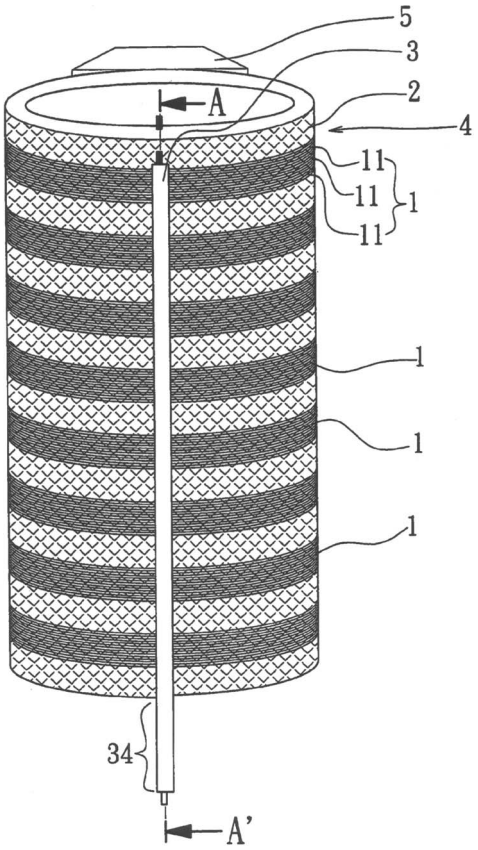
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
DAVID PREVES, *Starkey Laboratories, 6600 Washington, Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,339,447

43.20.Ye HIGH-VOLTAGE TRANSFORMER COIL WITH ACOUSTIC WAVE GUIDING FUNCTION

Jiann-Fuh Chen *et al.*, assignors to Unelectra International Corporation
4 March 2008 (Class 333/242); filed 1 December 2005

A waveguide **3** made of a nonconducting material is attached to cast resin transformer coil **4**. An acoustic transducer is attached at the end of extension **34**, which is a suitable distance from the high voltages that may be present in coils **1**. Acoustic signals produced by insulation breakdown between wires **11** are transmitted via the waveguide. The acoustic signal can be analyzed to determine such partial insulation breakdowns before catastrophic failure.—NAS



7,313,247

43.38.Ja AUTOMOTIVE INTERIOR TRIM PANEL WITH INTEGRAL ACOUSTIC CHAMBER

Robert Tilli *et al.*, assignors to Ford Global Technologies, LLC
25 December 2007 (Class 381/389); filed 25 April 2005

A scheme to utilize all available volume within an automobile door or other panels enables better sound reproduction at the lower frequencies. The arrangement that is described is compatible with side impact protection and can eliminate the need for loudspeakers in the rear package tray, thus saving weight and cost.—KPS

7,324,013

43.38.Ja SAFETY ALARM SYSTEM

Earl Edward Esson, assignor to Preco Electronics, Incorporated
29 January 2008 (Class 340/903); filed 2 November 2004

A safety system for use on moving vehicles provides an added level of safety by not only indicating motion, but changing the sound pattern when danger is imminent. An object detection sensor and a measure of ambient noise are both used to modify the warning signals' character and amplitude.—KPS

7,327,235

43.38.Ja ALARM SOUND OUTPUTTING DEVICE FOR VEHICLE AND PROGRAM THEREOF

Takashi Nagata, assignor to Denso Corporation
5 February 2008 (Class 340/435); filed in Japan 21 June 2004

Devices that detect the presence of other vehicles could improve situational awareness for an automobile driver. This patent concerns the presentation of such information to the driver. It is proposed that virtual sound sources be used, thus conveying the relative location of other vehicles.

 0001-4966/2008/124(4)/1899/6/$23.00

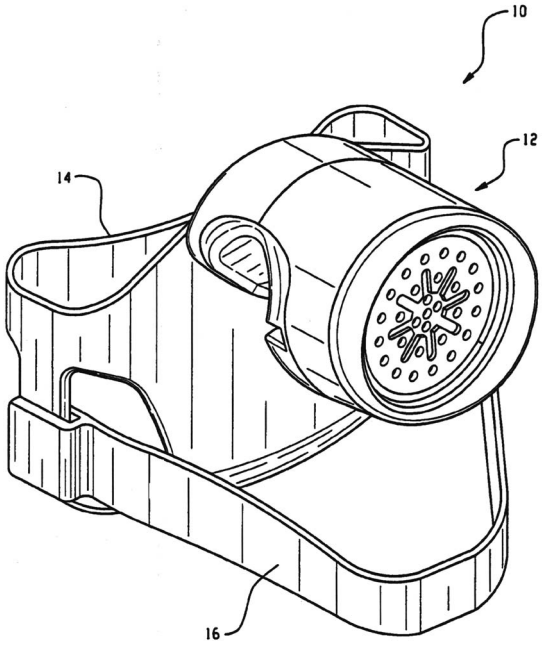
Head-related transfer functions are used to synthesize the source location using the car's sound system.—KPS

7,333,627

43.38.Ja AUXILIARY PLAYPEN SPEAKER

Todd Ventrola and Thomas J. Witman, assignors to Evenflo Company, Incorporated
19 February 2008 (Class 381/388); filed 18 June 2004

Speaker housing **12** with integral resilient clamp can be placed on the railing of a playpen or crib (but not necessarily limited to same) and secured by repositionable strap **16** while satchel **14** can hold a radio, CD player, or



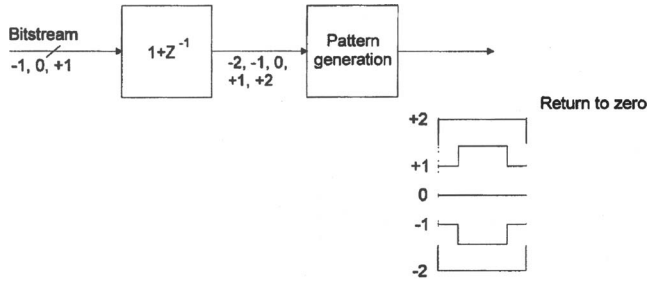
similar device to power the speaker "to provide musical enjoyment to children within a room or a playpen or crib."—NAS

7,336,794

43.38.Lc HIGH EFFICIENCY DRIVER FOR MINIATURE LOUDSPEAKERS

Claus Erdmann Fürst *et al.*, assignors to Sonion A/S
26 February 2008 (Class 381/117); filed 2 December 2002

A scheme using return to zero (RTZ) encoding and pattern generation blocks prior to the H-bridge in a miniature transducer mountable digital electronics power package is described. The clock frequency, RTZ encoding, pattern generation, and H-blocks can be varied as needed to meet



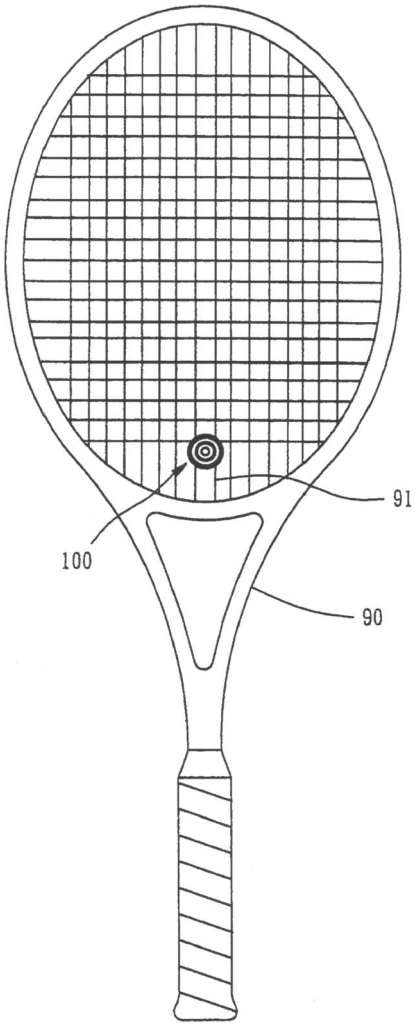
sound quality, emission of and susceptibility to EMI, and pricing constraints.—NAS

7,335,118

43.40.Kd SHOCK ABSORBING AND SOUND PRODUCING DEVICE FOR TENNIS RACKET

Hung-Fu Sung, Taichung Hsien, Taiwan
26 February 2008 (Class 473/461); filed 6 December 2006

Shock absorbing and sound producing assembly **100** can be mounted proximate to the handle, as shown in the figure, or distal to the handle of a tennis racket **90** in the network of strings **91**. The device thus reduces the amount of shock transmitted to the player while also producing a sound



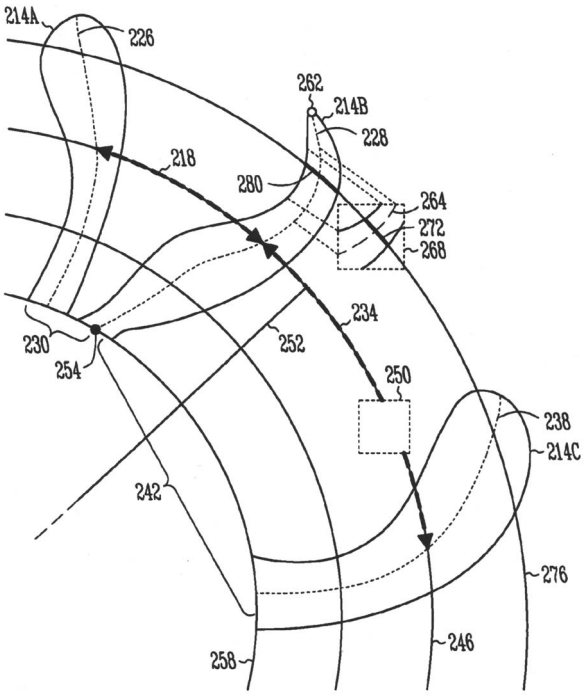
when the air flowing through the device is normal to the device, which is said to indicate whether the player has hit the ball correctly or not.—NAS

7,300,244

43.50.Gf COOLING FAN NOISE REDUCTION APPARATUS, SYSTEMS, AND METHODS

Eric Baugh *et al.*, assignors to Intel Corporation
27 November 2007 (Class 415/119); filed 30 September 2005

A computer processor chip cooling fan that should not emit a blade tone sound is claimed where fan blades are mounted with unequal spacings **218-234**, are bent **226-228-238**, are out of plane with one another, and are of



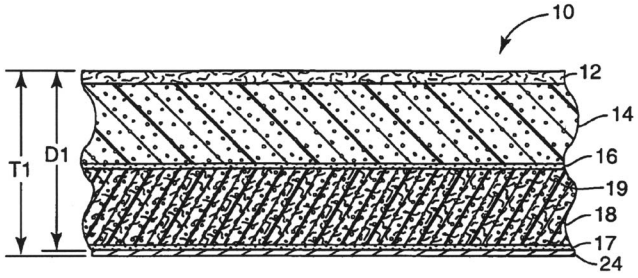
different shapes **214A-214B-214C**, as are also the struts that support the fan shaft and motor.—AJC

7,320,739

43.50.Gf SOUND ABSORPTIVE MULTILAYER COMPOSITE

Delton R. Thompson, Jr. and Xiaohe Liu, assignors to 3M Innovative Properties Company
22 January 2008 (Class 156/308.2); filed 1 July 2004

A multilayer composite is designed to absorb sound in applications such as automobile headliners. The key features are a surface layer, **12**, with a resistance of 500–4000 mks Rayls, an open cell foam or fibrous pad, **14** (<2000 Rayls), an air permeable core, **18** (<100 Rayls), and a barrier, **24**. The importance of each of these features is demonstrated. Some interesting examples are given in which commercial headliners are slightly modified,



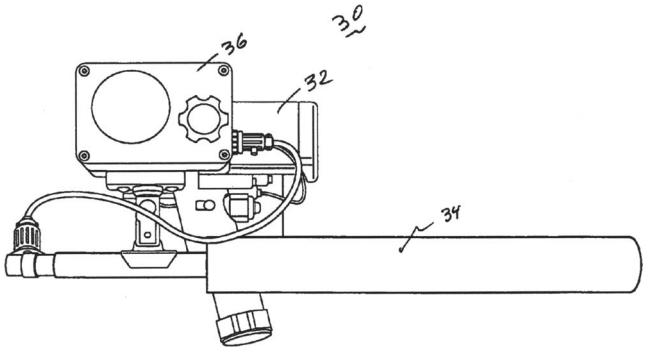
resulting in significant increases in acoustic performance. Manufacturing details are given along with specific material selections.—KPS

7,305,886

43.50.Lj NOISE DETECTING APPARATUS

Clark J. Radcliffe, assignor to Board of Trustees of Michigan State University
11 December 2007 (Class 73/647); filed 7 June 2005

A handheld device to measure noise from passing boats is intended for use by law enforcement authorities. It consists of a distance range finder, a directional microphone, a controller, and associated electronic components. Thus, a noise measurement can be associated with a known source distance. The controller is designed to subtract out the previously measured ambient noise and adjust the measured level to a reference distance of 50 m. The adjustment is based on a selectable x dB per doubling of distance, where x



is empirically determined. Data acquired using a prototype device seem to indicate that source directivity and propagation phenomena preclude the immediate adoption of the device by law enforcement authorities.—KPS

7,364,014

43.55.Ev SOUND ABSORBING BODY, SOUND ABSORBING STRUCTURAL BODY AND METHOD OF MANUFACTURING THESE BODIES

Hirofumi Goda *et al.*, assignors to Prime Polymer Company, Limited
29 April 2008 (Class 181/293); filed in Japan 26 April 2002

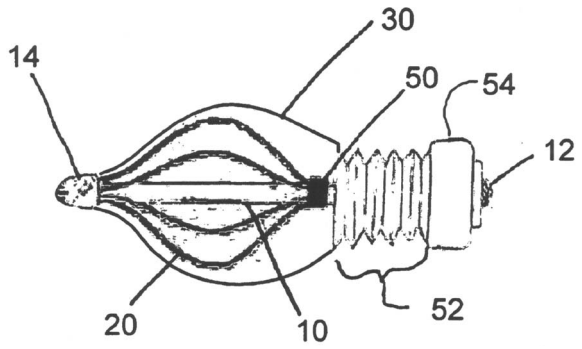
This product is comprised of a single three-layered molded body that both contains and absorbs sound. The two outer layers are a dense unexpanded fiber-containing thermo-plastic resin and the inner layer is an expanded version of the same resin. One of the outer layers is perforated so that the inner surface can absorb sound. The body is all of one piece from the same injection molding, rather than being made of separate multiple laminations. The patent also covers the manufacturing process, whereby one side of the mold is retracted or pulled away just enough for the inner part of the resin injection to expand—hence the foamed sound absorbing interior layer.—CJR

7,362,875

43.66.Ts BALLOON-EXPANDABLE HEARING DEVICE FITTING SYSTEM AND SELF-EXPANDING HEARING DEVICE

Gary M. Saxton *et al.*, assignors to Sonic Innovations, Incorporated
22 April 2008 (Class 381/322); filed 2 April 2004

Goals are to reduce custom hearing aid manufacturing time and to improve wearer comfort when the devices are used for long periods of time. A self-expanding hearing device is comprised of a malleable, tube-shaped plastic or metal body, a flexible membrane for contacting the ear canal wall, and a flexible frame having several rods for supports. The wearer



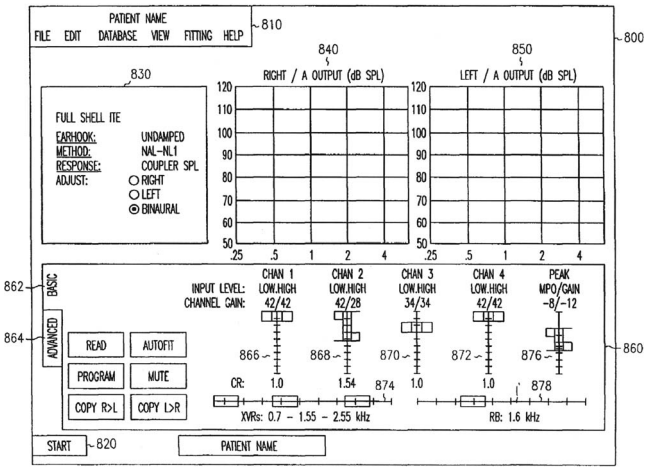
compresses the frame for insertion into the ear canal. When the frame is released, it expands via a springlike effect until the device conforms to and is retained in the ear canal.—DAP

7,366,307

43.66.Ts PROGRAMMABLE INTERFACE FOR FITTING HEARING DEVICES

Jerry L. Yanz *et al.*, assignors to Micro Ear Technology, Incorporated
29 April 2008 (Class 381/60); filed 11 October 2002

A streamlined graphical user interface for hearing aid programming provides, on a single common reference axis, a visual representation and a means for controlling multiple parameters that are related by a programmable constraint. An example is the amount of gain provided at a particular frequency for low and high level input signals may be controlled with two sliders operating on a single reference axis. The amount of movement of



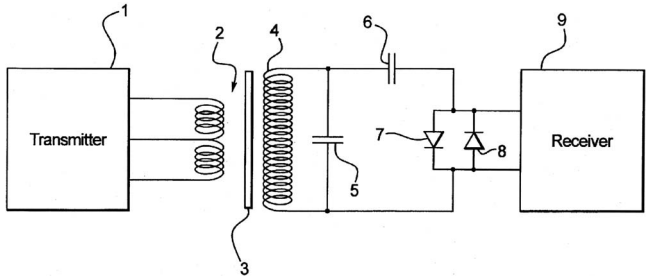
such sliders may be limited to bounds defined by the constraint. These interactions are conveyed visually to the hearing aid dispensing professional.—DAP

7,366,316

43.66.Ts DEVICE TO TRANSMIT AND RECEIVE DATA FOR REMOTE CONTROL OF HEARING DEVICES

Juergen Reithinger, assignor to Siemens Audiologische Technik GmbH
29 April 2008 (Class 381/331); filed 4 February 2004

The size and cost of a remote control is reduced by winding the transmitter and receiver coils around the same ferromagnetic core. A capacitor



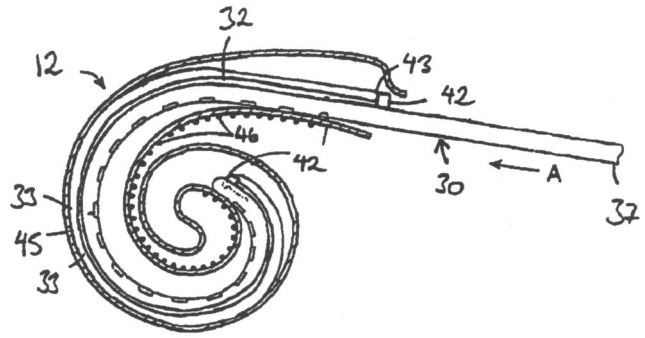
protects the receiver from overvoltage and is used also to tune the receiver oscillator frequency.—DAP

7,367,992

43.66.Ts CURVED COCHLEAR IMPLANT ELECTRODE ARRAY

Fysh Dadd, assignor to Cochlear Limited
6 May 2008 (Class 623/10); filed in Australia 12 March 2001

An implantable electrode array adopts a straight configuration for insertion deep into the cochlea and, once in place, a spirally curved shape so



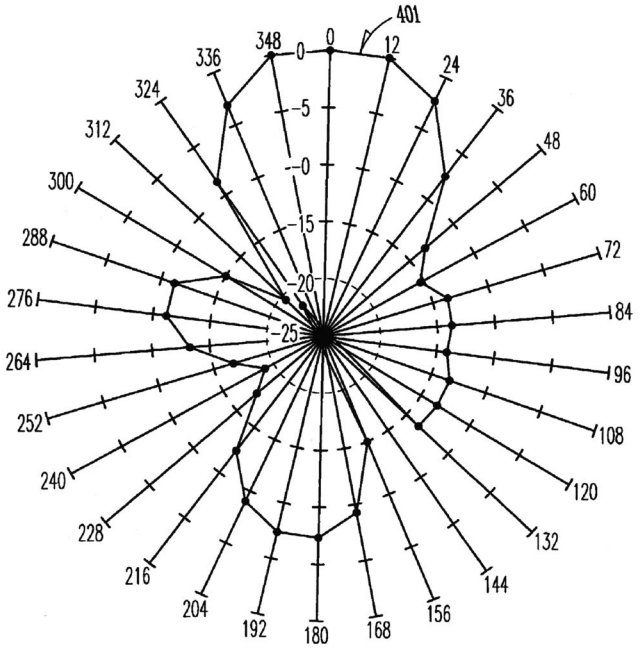
that adequate intensity stimulation at specific frequencies is applied to appropriate auditory nerves with minimal cross-electrode interference.—DAP

7,369,669

43.66.Ts DIOTIC PRESENTATION OF SECOND-ORDER GRADIENT DIRECTIONAL HEARING AID SIGNALS

Lawrence Hagen *et al.*, assignors to Micro Ear Technology, Incorporated
6 May 2008 (Class 381/313); filed 15 May 2002

Hearing devices that are worn on each ear of a binaural fitting, having either a switch-selectable omnidirectional or a first-order directional microphone system, are combined via a wireless or wired communication link across the head to produce either a first-order or second-order directional



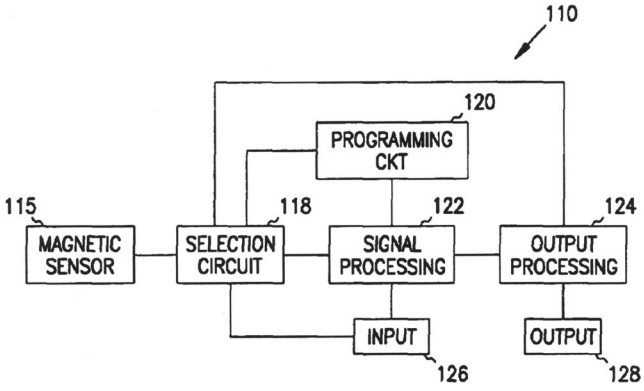
system, respectively. Gain and phase of the signals from the two ears are adjusted when they are mixed to obtain the desired polar directivity pattern.—DAP

7,369,671

43.66.Ts SWITCHING STRUCTURES FOR HEARING AID

Mike K. Sacha *et al.*, assignors to Starkey Laboratories, Incorporated
6 May 2008 (Class 381/331); filed 16 September 2002

A magnetic field sensor in a bridge circuit determines which of three different magnetic field strengths is present. A magnetically activated switch then switches inputs, outputs, or signal processing parameters accordingly. Inputs that can be selected include omnidirectional and directional microphones and an induction coil. Switching the signal processing parameters may vary the amount and type of filtering performed on the input signal. The



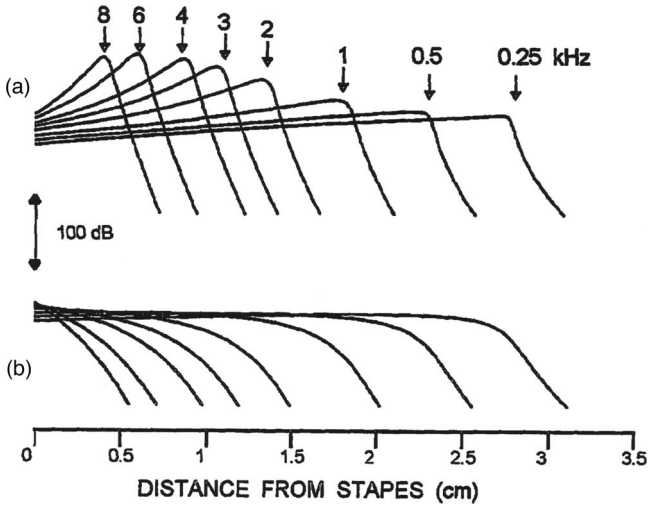
magnetic signal may contain a signal to inductively recharge the hearing aid battery and programming control information as well as programming data.—DAP

7,366,656

43.72.Ar METHOD APPARATUS AND SYSTEM FOR PROCESSING ACOUSTIC SIGNALS

Miriam Furst-Yust *et al.*, assignors to Ramot At Tel Aviv University Limited
29 April 2008 (Class 704/200.1); filed 19 February 2004

Following an extensive description of the mammalian hearing system, a speech analyzer in the form of a mechanical analog of the human cochlea is described. The approach taken seems to be that if frequency-dependent delays and gains can be properly implemented, the resulting acoustic transducer will provide superior performance as a hearing aid or as a speech recognition input device. A major aspect of the discussion is the contribution of the outer hair cells to the vibratory characteristics of the membrane.



Presumably, however, this contribution would be included in the aforementioned frequency-dependent delays and gains. The patent does include a sentence to the effect that the system could be implemented by other, nonmechanical techniques.—DLR

7,366,984

43.72.Ar PHONETIC SEARCHING USING MULTIPLE READINGS

Daryn E. Robbins *et al.*, assignors to Microsoft Corporation
29 April 2008 (Class 715/535); filed 29 July 2005

A system is described for entering Chinese characters into a computer using a system of phonetic codes. In order to handle cases in which the desired input character could be represented by any one of multiple phonetic codes, a target string of likely search candidates is stored in the user's database. When a phonetic code is entered, all possible matching Chinese characters are checked against the target string. Any hits are presented to the user as the most likely input characters.—DLR

7,363,230

43.72.Gy AUDIO DATA PROCESSING APPARATUS AND AUDIO DATA DISTRIBUTING APPARATUS

Yasuhiro Matsunuma, assignor to Yamaha Corporation
22 April 2008 (Class 704/500); filed in Japan 1 August 2002

A distributed audio network streams MP-3 encoded audio data from a server to client devices via a wireless network. To maintain good sound quality when network conditions fluctuate, processing resources are freed up to allow changes to be made in the bit rate of the encoded audio data. A detector monitors for a free processor to encode the divided PCM data that has overlapping sections resulting from several processors acting in parallel.

After encoding, the divided data are recombined, ensuring that the overlapping sections do not create any artifacts caused by dividing the data.—DAP

7,366,979

43.72.Kb METHOD AND APPARATUS FOR ANNOTATING A DOCUMENT

Steven Spielberg and Samuel Gustman, assignors to Copernicus Investments, LLC
29 April 2008 (Class 715/512); filed 9 March 2001

The patent describes an audio/video storage and playback device intended to aid a writer or editor of audio/video program materials by allowing the user to attach editing notes or annotations to the audio/video stream in such a way that the notes may be easily reviewed or modified during the editing process. One aspect of the invention would be that editing notes could be attached to program materials and reviewed using a variety of physical devices. The patent notes that prior systems of annotation typically required a specific hardware implementation. Editors need to be able to move the material easily across a variety of devices, such as an automobile sound system, a cell phone, a regular phone, a personal computer, or a thin client attached to a network.—DLR

7,366,766

43.72.Ne WEB-BASED SPEECH RECOGNITION WITH SCRIPTING AND SEMANTIC OBJECTS

John Kroeker and Oleg Boulanov, assignors to Eliza Corporation
29 April 2008 (Class 709/217); filed 23 March 2001

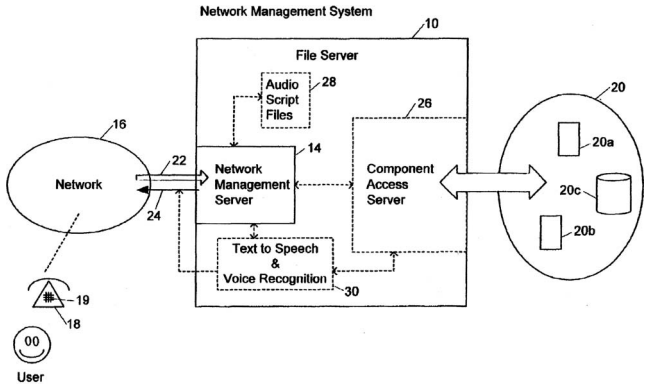
Transactional speech applications, including conducting interviews for surveys over a telephone automatically without human or server-side intervention, are implemented using Web-based technologies, such as voice over IP, together with speech recognition. A speech application script defines at least one application context, represented as categories of interpretation, to control the speech recognition system so a context-independent result is generated. A result, corresponding to all possible context-dependent interpretations of the audio input, is returned to the script.—DAP

7,369,996

43.72.Ne SYSTEM AND METHOD FOR VOICE BASED NETWORK MANAGEMENT

Darryl Sladden *et al.*, assignor to Cisco Technology Incorporated
6 May 2008 (Class 704/270.1); filed 15 August 2003

Management of network devices including routers, switches, and gateways is conducted using speech commands or dual tone multifrequency (DTMF) signaling via telephone. Users on the telephone respond to audio voice scripts generated by a server, using Voice Extensible Markup Language (VxML), or other forms of communication. User authorization is performed with automatic number identification of voice interface devices.



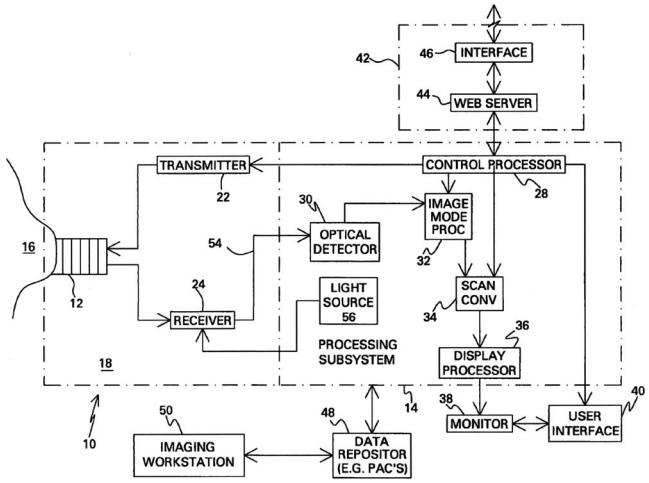
Management requests from users include controlling the number of network connections and verifying router status.—DAP

7,367,945

43.80.Vj ULTRASOUND SYSTEM

Samhita Dasgupta *et al.*, assignors to General Electric Company
6 May 2008 (Class 600/443); filed 29 March 2004

The ultrasound probe that transmits and receives ultrasound signals in this system includes an optical connection with electro-optic modulators that encode optical signals with electrical signals. Optical detectors that produce



corresponding demodulated signals are in the system console. The demodulated signals are processed to form an image.—RCW

LETTERS TO THE EDITOR

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

Factors contributing to comodulation masking release with dichotic maskers (L)

Emily Buss[a)] 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 2007; revised 10 July 2008; accepted 11 July 2008)

Detection threshold for a pure tone signal centered in a narrow band of noise may be reduced by inclusion of additional flanking masker bands, provided that they share coherent amplitude modulation (AM) across frequency. This comodulation masking release (CMR) associated with coherent AM across frequency is often much smaller if the signal and on-signal masker are presented to one ear and the flanking masker band(s) are presented contralaterally. An experiment was carried out to explore the role of peripheral effects (e.g., suppression) and central effects (e.g., grouping) in this finding. As frequently reported, CMR was smaller when two or more flanking maskers were presented contralaterally to the signal than when presented ipsilaterally. An intermediate condition, where a subset of flanking maskers was presented to each ear, provided comparable benefit to presenting all flankers ipsilateral to the signal. This result suggests that central effects may play a significant role in the reduced dichotic CMR under some conditions.
 [DOI: 10.1121/1.2968685]

PACS number(s): 43.66.Dc, 43.66.Rq [BCM] Pages: 1905–1908

I. INTRODUCTION

The detection threshold for a pure tone in a narrow band of noise is often lowered by the introduction of maskers at remote frequencies that have the same pattern of amplitude modulation as the on-signal masker, a finding described as comodulation masking release (CMR). This CMR effect can be demonstrated with a single wide band of noise or with multiple narrow band noise maskers distributed in frequency (Hall *et al.*, 1984). While CMR is often described in terms of the across-channel cues associated with coherent amplitude modulation (AM) across frequency, coherent AM is also associated with cues at the signal frequency that could improve performance, such as the envelope modulation associated with beating between neighboring bands (Schooneveldt and Moore, 1987) or suppression (Oxenham and Plack, 1998; Ernst and Verhey, 2006). Identifying the contributions of within- and across-channel cues has been the subject of several studies (e.g., Schooneveldt and Moore, 1987; Carlyon *et al.*, 1989) and a secondary goal of many more. It is widely believed that "true CMR," based on across-channel effects, can be small relative to the total masking release observed for closely spaced masker bands.

Previous studies have attempted to discriminate within- from across-channel effects in CMR using modeling (Verhey *et al.*, 1999), manipulation of stimulus features thought to disrupt across-channel processes (e.g., asynchronous gating; Dau *et al.*, 2004), adjustment in the spectral proximity of masker bands (Schooneveldt and Moore, 1987), and dichotic stimulus presentation, wherein flanking masker bands are presented contralateral to the signal and on-signal band (Cohen and Schubert, 1987; Schooneveldt and Moore, 1987). Each of these approaches has both strengths and weaknesses. For example, recent work in our laboratory suggests that asynchronous gating can disrupt both within- and across-channel processes. Dichotic CMR results are often interpreted cautiously; some reports suggest that the cues underlying dichotic CMR may differ from those responsible for monaural CMR (e.g., Ernst and Verhey, 2008), though study of the combination of cues across ears has failed to uncover any substantive differences between monaural and dichotic CMR (Schooneveldt and Moore, 1989). This concern aside, the dichotic CMR paradigm is arguably the most straightforward means of eliminating within-channel cues while holding other stimulus features constant. For this reason, dichotic CMR was chosen for further study in the present investigation.

Dichotic CMR findings have been quite variable, ranging from no dichotic CMR to values comparable to those

[a)]Electronic mail: ebuss@med.unc.edu

reported for monaural CMR. The wide variety of stimulus parameters used across studies could play a role in the range of results. Two studies that reported a dichotic CMR comparable to that found with monaural stimuli used two maskers which played continuously throughout a run (Cohen and Schubert, 1987; Schooneveldt and Moore, 1987). Using gated maskers, Hicks and Bacon (1995) reported no evidence of dichotic CMR, and Ernst and Verhey (2006) found no dichotic CMR for gated stimuli unless the flanking band level greatly exceeded the on-signal band level. Studies using monaural stimuli often report greater CMR for continuous than gated maskers (Fantini *et al.*, 1993), an effect that may be inversely related to the number of flanking bands (Hatch *et al.*, 1995). One goal of the present study was to test the hypothesis that dichotic CMR is similarly affected by masker gating, with a smaller dichotic CMR for gated as opposed to continuous stimuli.

Several reports of monaural CMR suggest that the outputs of auditory channels carrying information about the masker complex must be perceptually grouped in order to support the beneficial effects of coherent masker modulation. This conclusion rests primarily on the reduction or elimination of CMR under conditions of asynchronous onset of the on-signal and flanking maskers (Grose and Hall, 1993; Dau *et al.*, 2004). While all of the above-cited gated dichotic CMR studies used synchronous onset across bands, contralateral masker presentation could itself introduce a strong segregation cue in that the signal and masker bands can be lateralized to opposite sides of the head. While some studies report a large dichotic CMR for pairs of continuous maskers (e.g., Cohen and Schubert, 1987), others using a family of flanking bands have reported a dichotic CMR about 50% of the analogous monaural CMR (Hall *et al.*, 1990; Moore and Shailer, 1991). The finding of reduced dichotic CMR for multiple flanking maskers could reflect an increased tendency for stream segregation due to perceptual dissimilarity between the on-signal band and the family of contralateral flanking maskers. In the present study, it was hypothesized that distributing a family of flanking bands across ears would reduce the tendency to segregate the on-signal and flanking masker bands and thus increase the masking release produced by the contralateral flanking bands.

II. METHODS

A. Observers

Observers were ten normal hearing adults, four males and six females. All had pure tone thresholds of 20 dB HL or less at octave frequencies from 250 to 8000 Hz (ANSI, 1996) and no significant history of ear disease. Their ages ranged from 17 to 53, with a mean age of 29 years.

B. Stimuli

The signal was a 2 kHz pure tone, 400 ms in duration, including 50 ms raised-cosine ramps. The masker was one or more 20 Hz wide bands of Gaussian noise presented at 50 dB spectrum level. There was always an on-signal masker band centered on 2 kHz. When two flanking maskers were present they were centered on 1.6 and 2.4 kHz, the fourth and sixth harmonics of 400 Hz. When eight flankers were present the masker consisted of bands centered on the first to ninth harmonics of 400 Hz.

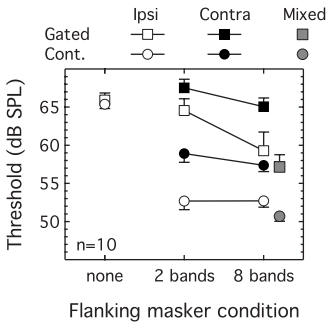


FIG. 1. Mean thresholds are plotted as a function of flanking masker condition, with error bars indicating 1 s.e. of the mean. Symbols reflect the mode of presentation. Symbol shapes distinguish gated (squares) from continuous (circles) masker presentation. Symbol shadings distinguish ipsilateral (open), contralateral (solid), and mixed (gray-filled) flanker presentation.

The signal and the on-signal masker band were always presented to the right ear. In the *ipsilateral* conditions, additional flanking bands were presented to the right ear, and in the *contralateral* conditions those flanking bands were presented to the left ear. The *mixed* condition included two flankers ipsilateral to the signal (the fourth and sixth harmonics of 400 Hz) and six flankers to the contralateral ear (first to third and seventh to ninth harmonics of 400 Hz). Maskers were gated on and off synchronously with the 400 ms signal or were presented continuously.

Maskers were generated in the frequency domain based on 2^{18} points, which when played at 24.4 kHz resulted in a 10.7 s sample that repeated seamlessly. The on-signal band was generated based on random Gaussian draws defining the real and imaginary components associated with bins within the range ± 10 Hz around 2 kHz. When flanking maskers were present, these same random draws were used to define the bands at different center frequencies.

C. Procedures

Stimuli were presented as a three-alternative forced choice, with the signal interval selected at random. Listening intervals were marked visually, with a 300 ms interstimulus interval. Correct answer feedback was provided after the observer response. The signal level was adjusted with a three-down one-up procedure (Levitt, 1971) estimating the 79% correct point. The initial step size was 4 dB prior to the second reversal and 2 dB thereafter. A track continued until eight reversals had been obtained, and the resulting threshold estimate was the average signal level at the last six reversals. Three such estimates were obtained in each condition, with a fourth collected in cases where the first three spanned a range of 3 dB or more. Thresholds were obtained in blocks by condition, and observers went through conditions in quasi-random order.

III. RESULTS AND DISCUSSION

Mean data are shown in Fig. 1. Thresholds are plotted as

a function of the flanking masker condition, with symbols indicating ear of presentation (shading) and whether the masker was gated (squares) or continuous (circles). Error bars indicate 1 standard error (s.e.) of the mean. Thresholds in the on-signal condition (labeled "none" in Fig. 1) were quite similar in the gated and continuous conditions ($t_9 = 0.84$, $p = 0.42$), with means of 65.9 and 65.4 dB SPL, respectively.

A. Ipsilateral and contralateral conditions

Gated conditions will be considered first. For ipsilateral flanking bands, introducing two flanking masker bands decreased thresholds by only 1.4 dB, and eight flanking bands decreased thresholds by 6.6 dB. In contrast, there was little evidence of masking release for either number of bands presented contralaterally, with mean thresholds rising 1.6 dB for two bands and falling 0.9 dB for eight bands; thresholds in both conditions were within the 95% confidence interval around the on-signal masker threshold. For continuous ipsilateral masker presentation there was no effect of masker number, with masking release of 12.7 dB for both two and eight bands. Continuous contralateral maskers were slightly less effective when there were two bands as compared to eight bands, with 6.4 and 8.0 dB masking release, respectively.

A repeated-measures analysis of variance (ANOVA) was performed on estimates of masking release, calculated as the change in threshold relative to the associated on-signal threshold (either gated or continuous). There were two levels of GATING (gated, continuous), NUMBER (two bands, eight bands) and EAR (ipsilateral, contralateral). All three main effects were significant: GATING ($F_{1,9} = 44.08$, $p < 0.0001$), NUMBER ($F_{1,9} = 14.29$, $p < 0.005$), and EAR ($F_{1,9} = 24.27$, $p < 0.001$). The interaction between GATING and NUMBER just failed to reach significance ($F_{1,9} = 4.5$, $p = 0.06$), and the other two-way interactions were not significant ($p >= 0.45$). The three-way interaction was significant ($F_{1,9} = 7.20$, $p < 0.05$). These results are consistent with the interpretation that the GATING × NUMBER interaction demonstrated in previous studies with monaural stimuli is absent in the dichotic data. This result could be attributed to fundamental differences between monaural and dichotic CMR or to the fact that there are insufficient cues to promote perceptual integration of the on-signal and flanking bands in the dichotic gated condition.

B. Mixed conditions

Attention now turns to the *mixed* conditions, wherein two flanking masker bands were presented to the same ear as the signal and the remaining six bands were presented contralaterally. Thresholds in these conditions were quite low, with average masking release of 8.7 and 14.7 dB in the gated and continuous conditions, respectively. These values can be compared to masking release of 6.6 and 12.7 dB obtained in comparable eight-band ipsilateral masker conditions. Masking release in the *mixed* conditions was assessed with a repeated-measures ANOVA, with two levels of EAR (eight-band ipsilateral, mixed) and two levels of GATING (gated, continuous). There was a significant main effect of EAR ($F_{1,9} = 59.02$, $p < 0.0001$) and of GATING (65.31, $p < 0.0001$), but no interaction ($F_{1,9} = 0.76$, $p = 0.41$). This outcome suggests that mixed presentation not only overcomes any reduction in CMR associated with contralateral flanking band presentation, but it may provide additional cues not present in the ipsilateral masker condition. It is not clear how this result comes about, but it could be related to previous findings of greater dichotic than monaural CMR in some conditions (Cohen and Schubert, 1987; Schooneveldt and Moore, 1989) or to previous reports of better diotic than monaural thresholds under conditions where performance is limited by internal noise (Langhans and Kohlrausch, 1992). It is possible that distributing flanking maskers across ears increases the number of independent auditory filters carrying information about the masker envelope; this possibility receives indirect support from the argument of Moore *et al.* (1993) that monaural CMR is reduced in hearing impaired listeners due to reduced frequency selectivity, but that this effect can be counteracted by presenting flanking bands contralateral to the signal and on-signal masker.

The most striking aspect of the *mixed* data is the 8.7 dB CMR obtained in the gated condition. If listeners were making use of cues present only in the signal ear, a modest CMR would be expected, similar to the 1.4 dB CMR for the two-band *ipsilateral* condition. The six contralateral flanking maskers in combination with these ipsilateral bands supported a robust CMR despite the fact that CMR was absent in the eight-band *contralateral* condition. This result is consistent with the hypothesis that CMR obtained with contralateral flanking bands may be strongly affected by auditory segregation. If bilaterally distributing masker bands improves performance by facilitating grouping, this result is consistent with the idea that little or no dichotic CMR is observed for gated stimuli due to a failure to group gated stimuli into a single auditory stream.

C. Central versus peripheral effects

Results presented here are consistent with the hypothesis that central effects related to auditory grouping can play a large role in the size of dichotic CMR. If grouping of comodulated bands is reduced by factors such as gating, a large disparity between numbers of bands presented to each ear, and differential lateralization, then the relative size of monaural and dichotic CMR should not be interpreted solely in terms of peripheral contributions to masking release. Results from the *mixed* conditions of this experiment are consistent with an interpretation that the poor CMR associated with the dichotic gated condition is related to a failure to process the on-signal and flanking bands as emanating from a single source. Gated masker bands in the nonsignal ear can be quite effective in contributing to making release when cues are provided that promote the fusion of bands across the ears.

IV. SUMMARY

(1) With all flanking masker bands contralateral to the signal, a dichotic CMR was obtained for continuous but not gated presentation.

(2) Distributing flanking maskers across ears resulted in a relatively large CMR. This result was most striking for the gated masker presentation mode: Contralateral maskers in the *mixed* condition improved thresholds by 7 dB compared to the ipsilateral condition with just two flanking masker bands.
(3) These results are consistent with the hypothesis that the absence of a dichotic CMR for gated maskers may be due to perceptual segregation of the stimulus components presented to the two ears, and that the *mixed* dichotic presentation promotes perceptual grouping of those components.
(4) Differences between monaural and dichotic CMR cannot be attributed solely to peripheral effects.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH NIDCD (Nos. R01 DC000418 and R01 DC007391). The authors thank Armin Kohlrausch, John Grose, Brian Moore, and two anonymous reviewers for contributing to this work.

ANSI (**1996**). *ANSI S3-1996, American National Standards Specification for Audiometers* (American National Standards Institute, New York).

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulator as a function of modulation rate," Hear. Res. **42**, 37–45.

Cohen, M. F., and Schubert, E. D. (**1987**). "Influence of place synchrony on detection of a sinusoid," J. Acoust. Soc. Am. **81**, 452–458.

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, Berlin), pp. 335–343.

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., Moore, B. C. J., and Schooneveldt, G. P. (**1993**). "Comodulation masking release as a function of type of signal, gated or continuous masking, monaural or dichotic presentation of flanking bands, and center frequency," J. Acoust. Soc. Am. **93**, 2106–2115.

Grose, J. H., and Hall, J. W. (**1993**). "Comodulation masking release: Is comodulation sufficient?," J. Acoust. Soc. Am. **93**, 2896–2902.

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.

Hatch, D. R., Arne, B. C., and Hall, J. W. (**1995**). "Comodulation masking release (CMR): Effects of gating as a function of number of flanking bands and masker bandwidth," J. Acoust. Soc. Am. **97**, 3768–3774.

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.

Langhans, A., and Kohlrausch, A. (**1992**). "Differences in auditory performance between monaural and dichotic conditions. I. Masking thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.

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

Moore, B. C. J., and Shailer, M. J. (**1991**). "Comodulation masking release as a function of level," J. Acoust. Soc. Am. **90**, 829–835.

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.

Oxenham, A. J., and Plack, C. J. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am. **104**, 3500–3510.

Schooneveldt, G. P., and Moore, B. C. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

Schooneveldt, G. P., and Moore, B. C. J. (**1989**). "Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands," J. Acoust. Soc. Am. **85**, 262–272.

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.

On the transient solutions of three acoustic wave equations: van Wijngaarden's equation, Stokes' equation and the time-dependent diffusion equation

Michael J. Buckingham[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California—San Diego, 9500 Gilman Drive, La Jolla, California 92093-0238

(Received 26 March 2008; revised 21 July 2008; accepted 23 July 2008)

Acoustic wave propagation in a dispersive medium may be described by a wave equation containing one or more dissipation terms. Three such equations are examined in this article: van Wijngaarden's equation (VWE) for sound propagating through a bubbly liquid; Stokes' equation for acoustic waves in a viscous fluid; and the time-dependent diffusion equation (TDDE) for waves in the interstitial gas in a porous solid. The impulse-response solution for each of the three equations is developed and all are shown to be strictly causal, with no arrivals prior to the activation of the source. However, the VWE is nonphysical in that it predicts instantaneous arrivals, which are associated with infinitely fast, propagating Fourier components in the Green's function. Stokes' equation and the TDDE are well behaved in that they do not predict instantaneous arrivals. Two of the equations, the VWE and Stokes' equation, satisfy the Kramers-Kronig dispersion relations, while the third, the TDDE, does not satisfy Kramers-Kronig, even though its impulse-response solution is causal and physically realizable. The Kramers-Kronig relations are predicated upon the (mathematical) existence of the complex compressibility, a condition which is not satisfied by the TDDE because the Fourier transform of the complex compressibility is not square-integrable.
 [DOI: 10.1121/1.2973231]

PACS number(s): 43.20.Bi, 43.20.Hq, 43.30.Es [JJM] Pages: 1909–1920

I. INTRODUCTION

Any partial differential equation describing acoustic-wave propagation in a dispersive medium contains at least one term that is representative of dissipation in the material. Three such equations are examined in this article, all of which are based on the assumption that the medium supporting the propagation is a homogeneous, time-invariant continuum: van Wijngaarden's equation[1] (VWE) for acoustic waves in an isothermal, viscous, bubbly liquid; Stokes' equation,[2] which may be thought of as a special case of the VWE, for acoustic waves in a bubble-free, viscous fluid; and the time-dependent diffusion equation[3] (TDDE) for waves in the interstitial gas filling a porous, statistically isotropic, perfectly rigid solid.[4] In all three cases, the solutions of these equations for harmonic waves have long been known; but the transient solutions are less amenable to analysis. As a result, a number of misconecptions have appeared in the literature. For instance, Jordan *et al.*[5] claim that Stokes' equation yields solutions that do not satisfy causality; and Blackstock[6] concludes that Stokes' equation is nonphysical, based on the appearance of an infinitely fast precursor in his solution. In fact, as demonstrated later, the transient solutions of Stokes' equation[7] are perfectly well-behaved: (1) they satisfy causality, with no arrivals present prior to the activation of the source; and (2) they are physically realizable, with no instantaneous arrivals predicted anywhere in the medium.

Much of the of the interest in the transient solutions of wave equations focuses on causality and the behavior of the predicted arrival around $t=0$, the time at which the source is activated. Although extensive discussions of causality may be found in the literature,[8–10] a simple definition holds true: a macroscopic system is said to be causal provided that the output does not precede the input. Similarly, a wave equation may be said to be causal provided that, everywhere in the medium, its solutions are zero prior to the time at which the source is activated.

Although causality requires that the effect must not precede the cause, this condition is not sufficient to ensure that the system, or predicted response in the case of an equation, is physically realizable. Since an infinitely fast propagating wave is impossible, at least in the realm of continuum mechanics, the first nonzero arrival at a receiver that is not coincident with the source must appear at some finite time after the activation of the source. In other words, instantaneous arrivals are nonphysical. Thus, to be physical, the first nonzero response must be delayed relative to the activation time of the source.

Assuming that they are accessible, the transient solutions reveal whether a wave equation is not only causal but also physical. Based on such solutions, it will be shown that the VWE is causal but not physical, while Stokes' equation and the TDDE are both causal and physical. As an alternative to the transient solutions, it is possible to investigate causality, but not physical realizability, through the Kramers-Kronig dispersion relations,[11,12] which were developed originally in connection with optical wave propagation in a

[a)]Also affiliated with the: Institute of Sound and Vibration Research, The University, Southampton SO17 1BJ, U.K. Electronic mail: mjb@mpl.ucsd.edu

dielectric material. Dispersion relations, which take the form of an integral-transform pair, express the real part of a complex function, $K(\omega)$, in terms of the Hilbert transform of the imaginary part and *vice versa*. In acoustics, the function $K(\omega)$ represents the complex compressibility, which scales inversely with the square of the complex sound speed; and the real variable ω represents angular frequency. The first investigation of dispersion relations for acoustic waves is attributed to Ginzberg[13] with subsequent contributions from Mangulis,[14] Weaver and Pao,[15] and Lee *et al.*[16]

The Kramers-Kronig dispersion relations hold provided that the physical system under consideration is both linear and causal (certain Fourier-transform existence conditions must also be satisfied). They do not depend on the detailed physical mechanisms underlying the wave propagation and thus they provide less information about the wave field than the transient solutions of the associated wave equation. Notwithstanding, the Kramers-Kronig dispersion relations have traditionally provided a useful test of causality. Thus, a wave equation that possesses causal transient solutions might be expected to satisfy Kramers-Kronig, whereas a wave equation that does not satisfy Kramers-Kronig might be expected to return noncausal transient solutions. It turns out that neither of these expectations is necessarily correct: all three of the wave equations investigated in this article are causal but only two, the VWE and Stokes' equation, satisfy Kramers-Kronig. The TDDE does not satisfy Kramers-Kronig but nevertheless is perfectly well behaved in that its impulse response is not only causal but also physically realizable. The VWE, on the other hand, satisfies Kramers-Kronig and is causal but returns an impulse response that is nonphysical. Stokes' equation also satisfies Kramers-Kronig, is causal and, unlike the VWE, yields an impulse response that is physically realizable.

Before pursuing the transient solutions of the three wave equations, it is instructive to develop the Kramers-Kronig relations for acoustic waves in a dispersive medium. This is achieved by following an argument similar to that of O'Donnell *et al.*,[17] which is predicated upon (1) linear acoustic propagation, implying that superposition holds, and (2) a causal relationship between the pressure and density fluctuations in the medium.

II. ACOUSTIC DISPERSION RELATIONS

A. Derivation of the Kramers-Kronig relations

As a sound wave propagates through a dispersive medium, the pressure fluctuation is accompanied by a change in the density. Provided both are sufficiently small, as is usually the case for acoustic waves, the pressure and density fluctuations may be regarded as the input and output, respectively, of a linear system whose impulse response is the time-domain compressibility $k(t)$. Thus, the equation of state may be expressed in the time domain as a convolution integral having the form

$$r(t) = \int_{-\infty}^{\infty} k(t-t')p(t')dt', \tag{1}$$

where $p(t)$ is the pressure fluctuation about the mean and $r(t)$ is the relative density fluctuation, or condensation,[18]

$$r(t) = \frac{\rho(t) - \rho_0}{\rho_0} \ll 1. \tag{2}$$

In this expression, $\rho(t)$ is the absolute density fluctuation and ρ_0 is the bulk density of the medium. Since $k(t)$ is a measurable quantitiy, it is a real function of time.

Assuming that the Fourier transforms of $r(t)$, $k(t)$, and $p(t)$ exist, that is, the time functions are square-integrable, then the convolution in Eq. (1) may be expressed in the frequency domain as the product

$$R(\omega) = K(\omega)P(\omega), \tag{3}$$

where the uppercase letters denote the Fourier transforms of the corresponding time functions

$$R(\omega) = \int_{-\infty}^{\infty} r(t)e^{-i\omega t}dt, \tag{4}$$

and similarly for $K(\omega)$ and $P(\omega)$. Now, on taking the pressure in Eq. (1) to be an impulse,

$$p(t) = q_0\delta(t), \tag{5}$$

where q_0 is a constant and $\delta(\cdots)$ is the Dirac delta function, and performing the appropriate inverse Fourier transform, the condensation becomes

$$r(t) = q_0k(t) = \frac{q_0}{2\pi}\int_{-\infty}^{\infty} K(\omega)e^{i\omega t}d\omega. \tag{6}$$

At this point in the argument, the concept of causality is introduced. The condensation must be zero prior to the onset of the pressure at $t=0$ and hence the Fourier integral in Eq. (6) has to be zero for negative times. This condition may be written as

$$\int_{-\infty}^{\infty} K(\omega')e^{-i\omega' t}d\omega' = 0 \quad \text{for } t > 0. \tag{7}$$

By allowing ω' to become complex, and assuming that the frequency-domain compressibility converges according to the condition

$$\lim_{|\omega'|\to\infty} |K(\omega')| = 0, \quad 0 \geqslant \arg\omega' > -\pi, \tag{8}$$

then it should be clear, by constructing a semicircle of infinite radius around the lower half-plane, that Eq. (7) will be satisfied provided that $K(\omega')$ is analytic both in the lower half of the complex ω'-plane and on the real axis. This causality requirement, that there be no singularities in $K(\omega')$ on or below the real axis, is central to the derivation of the dispersion relations.

To continue, the expression in Eq. (7) is now multiplied by $e^{i\omega t}$ and integrated over positive t, to obtain

$$\int_{-\infty}^{\infty} K(\omega')d\omega' \int_{0}^{\infty} e^{i(\omega-\omega')t}dt = 0. \tag{9}$$

By using the identity

$$\int_{0}^{\infty} e^{iat}dt = \pi\delta(a) - \frac{1}{ia}, \tag{10}$$

which holds for a real, Eq. (9) reduces to

$$\pi K_1(\omega) - i\pi K_2(\omega) - i\int_{-\infty}^{\infty} \frac{K_1(\omega') - iK_2(\omega')}{\omega' - \omega} d\omega' = 0, \tag{11}$$

where K_1 and $-K_2$ are the real and imaginary parts of $K(\omega)$:

$$K(\omega) = K_1(\omega) - iK(\omega), \tag{12}$$

which serves to define the sign convention used hereafter.

The Kramers-Kronig dispersion relations emerge from Eq. (11) by equating the real and the imaginary parts of the left side to zero, which yields

$$K_1(\omega) = \frac{1}{\pi}\mathrm{P}\int_{-\infty}^{\infty} \frac{K_2(\omega')}{\omega' - \omega} d\omega' \tag{13a}$$

and

$$K_2(\omega) = -\frac{1}{\pi}\mathrm{P}\int_{-\infty}^{\infty} \frac{K_1(\omega')}{\omega' - \omega} d\omega'. \tag{13b}$$

The integrals in Eqs. (13a) and (13b) are taken along the real axis, which contains a simple pole at $\omega' = \omega$. This is accommodated by interpreting each integral as the Cauchy principal value, as denoted by "P" in Eqs. (13a) and (13b). According to these expressions, the real and imaginary parts of the complex compressibility are not independent: the real part, $K_1(\omega)$, may be determined if the imaginary part, $-K_2(\omega)$, is known at all frequencies, and *vice versa*. Moreover, if the function $k(t)$ is causal, then the dispersion relations in Eqs. (13a) and (13b) hold, and conversely, if Eqs. (13a) and (13b) hold, then $k(t)$ is causal, which constitutes a partial statement of Titchmarsh's theorem.[9,19]

B. Conjugate poles of $K_1(\omega)$ and $K_2(\omega)$

While causality requires $K(\omega')$ to have no poles on or below the real axis, this is not so of its real and imaginary parts. Both $K_1(\omega')$ and $K_2(\omega')$ have poles in the lower half-plane, which are conjugates of the poles of $K(\omega')$ above the real axis, and these conjugate poles will contribute to the integrals in Eqs. (13a) and (13b). Their effect may be seen by constructing a closed contour, formed by the real axis and a semicircle of infinite radius in the lower half-plane. Since, from Eq. (8), the integral around the semicircle is zero, it follows from Cauchy's theorem that the integrals along the real axis in Eqs. (13a) and (13b) may be written as

$$K_1(\omega) = -i\left\{K_2(\omega) + 2\sum \text{ residues of } \frac{K_2(\omega')}{\omega' - \omega} \text{ from conjugate poles of } K_2(\omega)\right\}, \tag{14a}$$

$$K_2(\omega) = -i\left\{K_1(\omega) + 2\sum \text{ residues of } \frac{K_1(\omega')}{\omega' - \omega} \text{ from conjugate poles of } K_1(\omega)\right\}, \tag{14b}$$

The first term in each of the parentheses in Eqs. (14) is the residue of the pole on the real axis at $\omega' = \omega$, obtained in the usual way by indenting a small semicircle either above or below the pole. From Eqs. (14), it is readily shown that the following equality holds:

$$\sum \text{ residues of } \frac{K_2(\omega')}{\omega' - \omega} \text{ from conjugate poles of } K_2(\omega) = i\sum \text{ residues of } \frac{K_1(\omega')}{\omega' - \omega} \text{ from conjugate poles of } K_1(\omega). \tag{15}$$

Note that the pole on the real axis at $\omega' = \omega$ does not contribute to this identity.

In effect, Eqs. (14a) and (14b) may be regarded as an alternative formulation of the Kramers-Kronig dispersion relations. They state that $K_1(\omega)$ can be computed from $K_2(\omega)$, and *vice versa*, provided only that the poles of $K(\omega)$ are known. In this case, unlike the situation with the Kramers-Kronig integrals themselves, it would not be necessary to know the real (imaginary) part of $K(\omega)$ at all frequencies in order to compute the imaginary (real) part of $K(\omega)$. However, such an application of Eqs. (14a) and (14b) does present certain difficulties, at least in the context of an attempt to determine, say, $K_1(\omega)$ from experimental measurements of $K_2(\omega)$, since such measurements do not return the required poles. On the other hand, Eqs. (14a) and (14b) provide a useful tool for investigating the causal properties of partial differential wave equations of the type to be discussed later.

C. Wave speed and attenuation

The complex compressibility $K(\omega)$ is related to the complex sound speed $c(\omega)$ as follows:

$$\frac{1}{c^2(\omega)} = \rho_0 K(\omega), \tag{16}$$

which, in terms of the phase speed $c_p(\omega)$ and attenuation $\alpha_p(\omega)$ becomes

$$\left[\frac{1}{c_p(\omega)} - \frac{i\alpha_p(\omega)}{\omega}\right]^2 = \rho_0[K_1(\omega) - iK_2(\omega)]. \tag{17}$$

By equating real part to real part and imaginary part to imaginary part, this may be decomposed into

$$\frac{1}{c_p^2(\omega)} - \frac{\alpha_p^2}{\omega^2} = \rho_0 K_1(\omega) \tag{18a}$$

and

$$\frac{2\alpha_p(\omega)}{\omega c_p(\omega)} = \rho_0 K_2(\omega). \tag{18b}$$

In these expressions, the phase speed and attenuation are coupled, which raises difficulties in using the Kramers-Kronig dispersion relations to compute one from knowledge of the other, even though that knowledge may be exact and for all frequencies. The situation is mitigated if, as is often the case, the attenuation is small such that

$$\frac{\alpha_p(\omega)c_p(\omega)}{\omega^2} \ll 1, \tag{19}$$

for then the phase speed may be written explicitly in terms of $K_1(\omega)$:

$$c_p(\omega) \approx \frac{1}{\sqrt{\rho_0 K_1(\omega)}}. \tag{20a}$$

But $K_2(\omega)$ is still mixed in $c_p(\omega)$ and $\alpha_p(\omega)$:

$$\alpha_p(\omega) = \frac{\rho_0 c_p(\omega)}{2}\omega K_2(\omega). \tag{20b}$$

By substituting these expressions for $K_1(\omega)$ and $K_2(\omega)$ into the Kramers-Kronig dispersion relation in Eq. (13b), it is evident that the attenuation can be computed if the phase speed is known at all frequencies. The converse, however, is not true; the phase speed cannot be derived from knowledge of the attenuation at all frequencies, although O'Donnel *et al.*[17] suggest otherwise. The difficulty is apparent from Eq. (13a), which, with Eqs. (20a) and (20b), becomes

$$\frac{1}{c_p^2(\omega)} = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{\alpha_p(\omega')}{\omega' c_p(\omega')(\omega' - \omega)} d\omega'. \tag{21}$$

Thus, the unknown function $c_p(\omega)$ occurs both outside and inside the integral, creating an intractable integral equation. Certain integral equations can be solved, but not so in this case, which leads to the conclusion that the wave speed cannot be determined from full knowledge of the attenuation.

III. van WIJNGAARDEN'S EQUATION

In 1972, van Wijngaarden[1] developed a wave equation describing the propagation of acoustic waves in an isothermal, viscous, bubbly liquid. For one-dimensional propagation, the inhomogeneous form of the VWE for the velocity potential, $g = g(x,t)$, is

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c_0^2}\frac{\partial^2 g}{\partial t^2} + \gamma\frac{\partial^3 g}{\partial t \partial x^2} + \beta^2\frac{\partial^4 g}{\partial t^2 \partial x^2} = -Q\delta(x)\delta(t), \tag{22}$$

where the delta functions on the right represent a planar, impulsive source and Q is the source strength with dimensions of (volume per unit area)=length. On the left, c_0 is the sound speed in the limit of low frequency or, equivalently, in the absence of the dissipation terms; the third term represents viscous effects, with the coefficient γ proportional to the dynamic viscosity of the bubbly mixture; and bubble-size effects are represented by the fourth term with coefficient β^2, which scales as the square of the equilibrium bubble radius. The homogeneous form of Eq. (22) has been examined by Jordan and Feuillade[20] in connection with a boundary value problem.

A comment on semantics is perhaps in order. In the following discussions, when referring to "van Wijngaarden's equation" or the "VWE," it is implied that the bubble-radius coefficient β^2 is greater than zero. When β^2 is equal to zero, representing an absence of bubbles, the fourth term on the left vanishes and Eq. (22) reduces to Stokes' equation. Thus, the term "Stokes equation" hereinafter signifies that β^2 is identically zero. This distinction is important because it underlies the fundamentally different behavior exhibited by the VWE and Stokes' equation.

A. Impulse response

To investigate the properties of its transient solutions, the VWE in Eq. (22) is to be solved using the same technique as that developed by Buckingham[7] in connection with Stokes' equation. Two bilateral Fourier transforms are applied to Eq. (22), the first with respect to time t, and the second with respect to distance x. The respective transform variables, angular frequency ω and wavenumber s, are used as subscripts to denote the transformed field, which is a convenient notation when multiple transforms are employed. Bearing in mind that the transforms of the second derivatives take their usual forms, with radiation conditions ensuring no integrated contributions, the result of the double Fourier transformation is the algebraic equation

$$g_{\omega s} = -\frac{Qc_0^2}{[\omega^2(1+\beta^2 c_0^2 s^2) - i\omega s^2 c_0^2\gamma - c_0^2 s^2]} = -\frac{Qc_0^2}{(1+\beta^2 c_0^2 s^2)(\omega-\omega_+)(\omega-\omega_-)}, \tag{23}$$

where the roots of the quadratic in ω are

$$\omega_\pm = i\frac{s^2\gamma c_0^2}{2(1+\beta^2 c_0^2 s^2)} \pm \frac{sc_0\sqrt{1+(\beta^2-(\gamma c_0/2)^2)s^2}}{(1+\beta^2 c_0^2 s^2)}. \tag{24}$$

Both of these roots lie in the top half of the complex ω-plane and thus $g_{\omega s}$ is analytic in ω on and below the real ω axis. Taking the inverse Fourier transform of Eq. (23) with respect to ω, yields the wavenumber spectrum

$$g_s(t) = -\frac{Qc_0^2}{2\pi(1+\beta^2 c_0^2 s^2)} \int_{-\infty}^{\infty} \frac{e^{i\omega t}}{(\omega-\omega_+)(\omega-\omega_-)} d\omega. \quad (25)$$

For $t<0$, this integral may be evaluated by forming a closed contour bounded by the real axis and a semicircle of infinite radius around the lower half-plane. As this contour contains no poles, the result is zero, indicating that the field is causal since there are no arrivals before $t=0$. For $t>0$, the appropriate contour is the real axis and a semicircle of infinite radius around the upper half-plane, which contains the poles at $\omega=\omega_+$. With the aid of Jordan's lemma and Cauchy's theorem,[3] the complete expression for the wavenumber spectrum of the velocity potential may therefore be written as

$$g_s(t) = u(t)\frac{Qc_0}{sR} \exp\left(-\frac{sc_0\gamma q}{2}t\right)\sin(qRt), \quad (26)$$

where

$$R = \sqrt{1+\left(\beta^2 - \frac{\gamma^2}{4}\right)s^2c_0^2}, \quad (27)$$

$$q = \frac{sc_0}{(1+\beta^2 c_0^2 s^2)}, \quad (28)$$

and $u(t)$ is the Heaviside unit step function. When $\beta=0$, Eq. (26) reduces identically to the corresponding solution of Stokes' equation derived by Buckingham.[7]

As the velocity potential is not a physically measurable quantity, it is convenient at this point to convert to the pressure $p(x,t)$. The wavenumber spectrum of the pressure is the time derivative of Eq. (26):

$$p_s(t) = \rho_0\frac{\partial g_s}{\partial t} = u(t)\frac{\rho_0 Qc_0}{sR} q \exp\left(-\frac{sc_0\gamma q t}{2}\right) \times\left\{R\cos(qRt) - \frac{sc_0\gamma}{2}\sin(qRt)\right\}, \quad (29)$$

where ρ_0 is the bulk density of the medium. It is interesting to note that, for $t=0+$, the wavenumber spectrum of the pressure is nonzero, and this is true for both the VWE ($\beta>0$) and Stokes' equation ($\beta=0$). As will be shown shortly, this step discontinuity at the origin of time in the wavenumber spectrum leads to a nonphysical Green's function for the VWE; but for Stokes' equation, the Green's function is perfectly well behaved.

B. The pressure Green's function

To obtain the pressure Green's function for the VWE, an inverse Fourier transform with respect to s is applied to Eq. (29), which yields

$$p(x,t) = u(t)\frac{\rho_0 Qc_0}{2\pi}\int_{-\infty}^{\infty}\frac{q}{sR}\exp\left(-\frac{sc_0\gamma q}{2}t\right)\left\{R\cos(qRt) - \frac{sc_0\gamma}{2}\sin(qRt)\right\}e^{isx}ds. \quad (30)$$

Obviously, this expression satisfies causality, since the unit step function $u(t)$ ensures that the pressure is zero for $t<0$. It is worth noting that, if the pressure field had been derived by reversing the order of the inverse Fourier transforms applied to Eq. (23), a different formulation from Eq. (30), with no explicit step function at the origin of time, would have been obtained. To be explicit, by inverse transforming Eq. (23) first with respect to wavenumber s, followed by the inverse transform with respect to frequency ω, the pressure Green's function is found to be

$$p(x,t) = \frac{\rho_0 Qc_0}{4\pi}\int_{-\infty}^{\infty}(1+i\omega\gamma-\beta^2\omega^2)^{-1/2} \times\exp\left\{-\frac{i\omega|x|}{c_0\sqrt{1+i\omega\gamma-\beta^2\omega^2}}\right\}e^{i\omega t}d\omega. \quad (31)$$

Of course, Eqs. (30) and (31) are equivalent, since they represent the same impulse response function. In the case of Eq. (31), although there is no step function at the origin, it still returns zero for negative times. This is confirmed by the fact that the integrand has no singularities (poles or branch points) on or below the real axis of the complex ω-plane and hence from Jordan's lemma and Cauchy's theorem, the field from Eq. (31) is identically zero for $t<0$, consistent with Eq. (30).

Both solutions exhibit a step discontinuity everywhere in the medium at $t=0$. The size of the discontinuity can be expressed explicitly from the integral in Eq. (30), which, for the special case $t=0+$, reduces to

$$p(x,0+) = \frac{\rho_0 Qc_0^2}{2\pi}\int_{-\infty}^{\infty}\frac{e^{isx}}{(1+\beta^2c_0^2s^2)}ds = \frac{\rho_0 Qc_0}{2\beta}e^{-|x|/\beta c_0} \quad \text{for } \beta>0, \quad (32)$$

a result that may be obtained by elementary contour integration or from tables of integrals.[21] Either way, it is evident that the size of the step depends on the bubble-size coefficient β but not the viscous coefficient γ. The presence of such a step in the pressure at $t=0$ can be attributed only to instantaneous arrivals at the receiver, thus forcing the conclusion that the VWE is nonphysical: such arrivals imply infinitely high wave speeds, which are impossible under the constraints of continuum mechanics. Some insight into the origin of these infinitely fast waves will be gained later from the dispersion curves associated with the VWE.

From inspection of Eq. (32), it is evident that the size of the step discontinuity at the origin is not a monotonic function of β but instead shows a maximum when $\beta=t_0$, where t_0 is the arrival time of a pressure impulse in a lossless medium [see Eq. (36) below]:

$$t_0 = |x|/c_0. \quad (33)$$

Substituting this value of β into Eq. (32) yields the maximum step size as

$$p_{\max} = \frac{\rho_0 Qc_0}{2t_0}e^{-1}. \quad (34)$$

By normalizing the pressure step in Eq. (32) to the maximum in Eq. (34), the relative step size can be expressed as

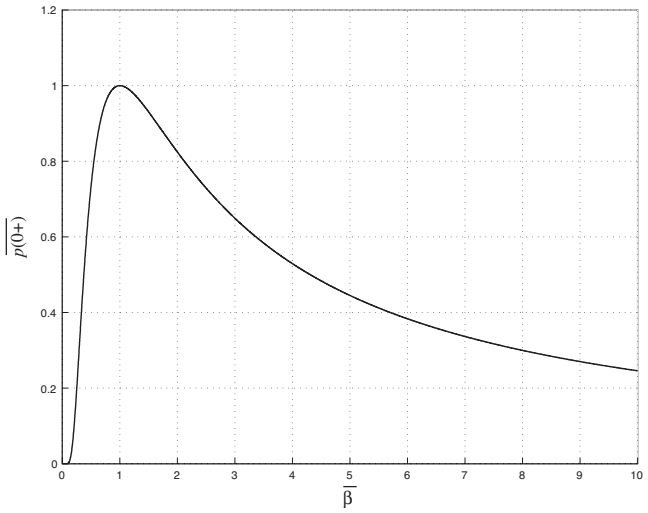


FIG. 1. Height of the VWE step discontinuity at $t=0$ as a function of the normalized bubble-radius coefficient $\bar{\beta}$, evaluated from Eq. (35).

$$\overline{p(0+)} = \frac{p(x,0+)}{p_{\max}} = \frac{1}{\beta}\exp\left(1-\frac{1}{\bar{\beta}}\right), \tag{35}$$

which is plotted in Fig. 1 as a function of the normalized coefficient $\bar{\beta}=\beta/t_0$.

Returning to the Green's function solution in Eq. (30), it is readily shown that, for the case of a lossless medium, represented by $\beta=\gamma=0$, the integral reduces to a delta function, allowing the pressure to be expressed as

$$p(x,t) = \frac{\rho_0 Q c_0}{2}\delta\left(t-\frac{|x|}{c_0}\right) \quad \text{for } \beta=\gamma=0. \tag{36}$$

This will be recognized as the correct form for the pressure impulse in an inviscid fluid; it is trivial to show that an identical result is also returned by Eq. (31).

Generally, when β and γ are finite, the integrals in Eqs. (30) and (31) cannot be evaluated explicitly but must be computed numerically. Taking the slightly simpler formulation in Eq. (31), it is convenient to work with the following normalized parameters and variables, denoted by overbars:

$$\bar{\beta}=\frac{\beta}{t_0};\quad \bar{\gamma}=\frac{\gamma}{t_0};\quad \bar{t}=\frac{t}{t_0};\quad \bar{\omega}=\omega t_0;\quad \overline{p(t)}=\frac{p(x,t)}{p_{\max}}. \tag{37}$$

The normalized form of Eq. (31) is then

$$\overline{p(t)} = \frac{e}{2\pi}\int_{-\infty}^{\infty}(1+i\bar{\omega}\bar{\gamma}-\bar{\beta}^2\bar{\omega}^2)^{-1/2} \times\exp\left\{-\frac{i\bar{\omega}}{\sqrt{1+i\bar{\omega}\bar{\gamma}-\bar{\beta}^2\bar{\omega}^2}}\right\}e^{i\bar{\omega}\bar{t}}d\bar{\omega}. \tag{38}$$

An example of the VWE pressure Green's function, computed from Eq. (38) using a Simpson's rule algorithm, is shown in Fig. 2(a), where the nonphysical discontinuity at the origin is clearly evident. It should be noted that the discontinuity illustrated in Fig. 2(a) is associated exclusively with the bubble-radius term in the VWE. A similar discontinuity will appear in any transient solution of the VWE whenever the source spectrum contains nonzero Fourier components at indefinitely high frequencies.

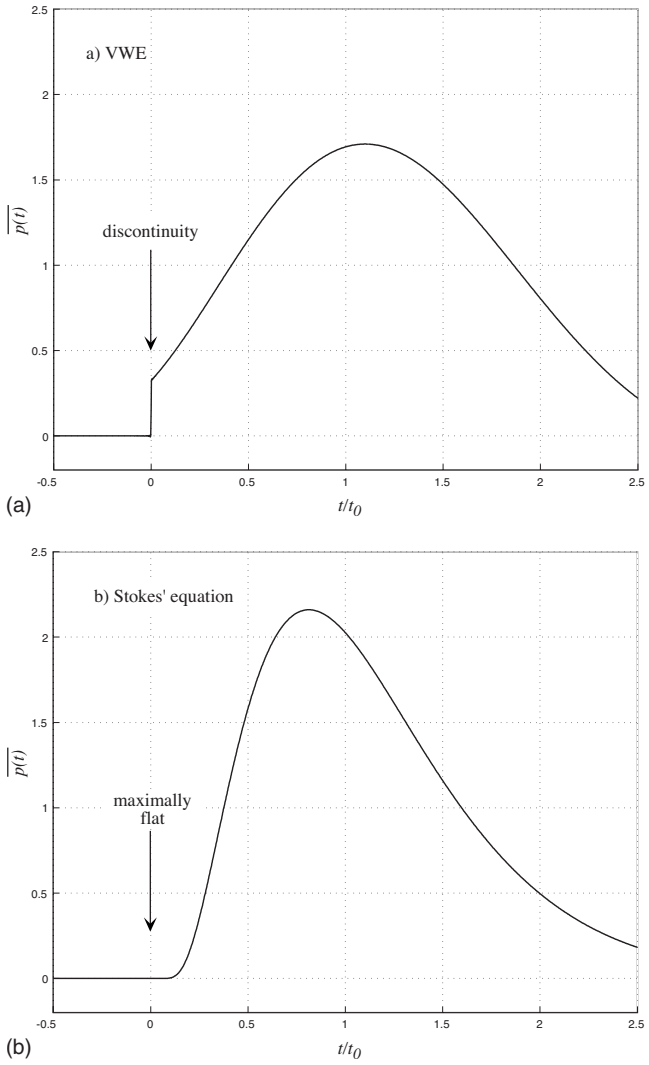


FIG. 2. Dimensionless pressure Green's function, computed from Eq. (38) using the normalization scheme described in the text. (a) The VWE with $\beta=\gamma$. (b) Stokes' equation, $\beta=0$.

C. An integral identity

The equivalence of the expressions in Eqs. (30) and (31) leads to an explicit expression for the integral in Eq. (31) for the special case $t=0+$. Under this condition, since Eq. (30) has the value given by Eq. (32), it follows that

$$\int_{-\infty}^{\infty}(1+i\omega\gamma-\beta^2\omega^2)^{-1/2}\exp\left\{-\frac{i\omega|x|}{c_0\sqrt{1+i\omega\gamma-\beta^2\omega^2}}\right\} \times e^{i\omega t}d\omega = \frac{2\pi}{\beta}e^{-|x|/\beta c_0}, \quad \text{for } \beta>0 \text{ and } t=0+. \tag{39}$$

Obviously, for $t<0$ the integral here is zero. Note that, as the integral in Eq. (39) is, in effect, a measure of the step size at the origin of time, it depends on β but not on γ, consistent with the statements made earlier in connection with Eq. (32).

D. Dispersion curves

To obtain the dispersion equations for the VWE, a Fourier transform with respect to time is applied to Eq. (22), which yields

$$\frac{\partial^2 g_\omega}{\partial x^2} + \frac{\omega^2}{c_0^2(1 + i\omega\gamma - \beta^2\omega^2)} g_\omega = -Q\frac{\delta(x)}{(1 + i\omega\gamma - \beta^2\omega^2)}. \tag{40}$$

The complex sound speed is therefore

$$c(\omega) = c_0\sqrt{1 + i\omega\gamma - \beta^2\omega^2}, \tag{41}$$

from which the phase speed is given by

$$\begin{aligned} c_p(\omega) &= \left[\mathrm{Re}\left(\frac{1}{c}\right)\right]^{-1} \\ &= \frac{\sqrt{2}c_0\sqrt{(1-\beta^2\omega^2)^2 + \omega^2\gamma^2}}{[(1-\beta^2\omega^2) + \sqrt{(1-\beta^2\omega^2)^2 + \omega^2\gamma^2}]^{1/2}} \\ &\to \begin{cases} c_0 & \text{as } \omega \to 0 \\ \dfrac{2c_0\beta^3\omega^2}{\gamma} & \text{as } \omega \to \infty \end{cases} \end{aligned} \tag{42a}$$

and the attenuation is

$$\begin{aligned} \alpha_p(\omega) &= -\frac{|\omega|}{c_0}\,\mathrm{Im}\left(\frac{1}{c}\right) \\ &= \frac{|\omega|}{\sqrt{2}c_0}\frac{[\sqrt{(1-\beta^2\omega^2)^2 + \omega^2\gamma^2} - (1-\beta^2\omega^2)]^{1/2}}{\sqrt{(1-\beta^2\omega^2)^2 + \omega^2\gamma^2}} \\ &\to \begin{cases} \dfrac{\gamma\omega^2}{2c_0} & \text{as } \omega \to 0 \\ \dfrac{1}{\sqrt{2}\beta c_0} & \text{as } \omega \to \infty \end{cases}. \end{aligned} \tag{42b}$$

In Fig. 3(a), normalized versions of Eqs. (42) are plotted as functions of the dimensionless frequency $\omega\gamma$. These plots reveal the origin of the step discontinuity in the pressure Green's function at $t=0$. Immediately after the source is activated, the behavior of the Green's function is dominated by its high-frequency Fourier components, which exhibit a phase speed that diverges as the square of the frequency but an attenuation that asymptotes to a constant. Thus, in the limit of high frequency, the VWE predicts Fourier components that travel infinitely fast through the medium while experiencing only finite attenuation. These infinitely fast, propagating waves gives rise to instantaneous arrivals everywhere in the medium and are responsible for the nonphysical step response in the Green's function at the instant the source is activated.

E. Kramers-Kronig relations

The relationship between complex compressibility and complex sound speed is given in Eq. (16). It follows from this expression, in conjunction with Eq. (41), that the complex compressibility for the VWE is

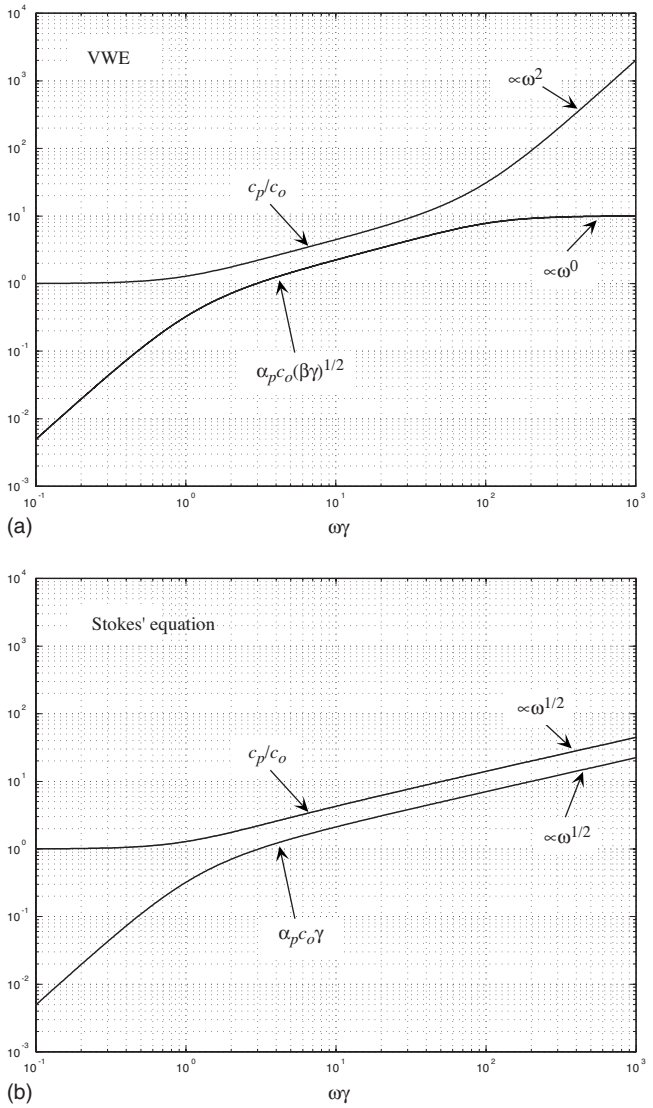


FIG. 3. Normalized dispersion curves as functions of the dimensionless frequency $\omega\gamma$ for (a) the VWE, computed from Eqs. (42a) and (42b) with $\beta=0.1\gamma$ and (b) Stokes' equation, computed from Eqs. (50a) and (50b).

$$K(\omega) = \frac{1}{\rho_0 c_0^2(1 + i\omega\gamma - \beta^2\omega^2)}, \tag{43a}$$

where the denominator has two roots, $a_\pm$, given by

$$a_\pm = \frac{i\gamma \pm \sqrt{4\beta^2 - \gamma^2}}{2\beta^2}. \tag{43b}$$

For $\beta>0$, both of these roots lie in the upper half of the complex ω-plane, irrespective of whether the radical is real or imaginary. Since both poles lie in the upper half-plane, the complex compressibility is analytic on and below the real axis, a corollary of which is that $k(t)$, the inverse Fourier transform of $K(\omega)$, is zero for $t<0$. By applying an inverse Fourier transform to Eq. (43a), the complete formulation of $k(t)$ is obtained:

$$k(t) = u(t)\frac{2}{\rho_0 c_0^2\sqrt{4\beta^2-\gamma^2}}e^{-\gamma t/2\beta^2}\sin\left(\frac{\sqrt{4\beta^2-\gamma^2}}{2\beta^2}t\right). \quad (44)$$

Thus, $k(t)$ is causal, and hence from Titchmarsh's theorem[9,19] the Kramers-Kronig dispersion relations in Eq. (13a) and (13b) must hold for the VWE.

Of course, it could be established that the VWE satisfies Kramers-Kronig by a direct evaluation of the residue expressions in Eqs. (14a) and (14b). To this end, the real and imaginary parts of the complex compressibility are, respectively,

$$K_1(\omega) = \frac{1-\omega^2\beta^2}{\rho_0 c_0^2[(1-\omega^2\beta^2)^2+\omega^2\gamma^2]} \quad (45a)$$

and

$$K_2(\omega) = \frac{\omega\gamma}{\rho_0 c_0^2[(1-\omega^2\beta^2)^2+\omega^2\gamma^2]}, \quad (45b)$$

both of which have poles above and also below the real axis in the complex ω-plane. In the upper half-plane, the two poles are located at $\omega=a_+$, where a_+ are the roots of the denominator of $K(\omega)$, as specified in Eq. (43b); in the lower half-plane, the two poles are at the conjugate positions, $\omega=a_+^*$. From the residues of these two conjugate poles, it is straightforward to show that the Kramers-Kronig dispersion relations, as formulated in Eqs. (14a) and (14b), are indeed satisfied by the VWE. Moreover, it is readily verified that the residues from the conjugate poles in the lower half-plane also satisfy the identity in Eq. (15).

IV. STOKES' WAVE EQUATION

Acoustic waves propagating in a viscous fluid are governed by a classical wave equation originally derived by Stokes.[2] For one-dimensional propagation along the x-axis, the inhomogeneous form of Stokes' equation for an impulsive source is

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c_0^2}\frac{\partial^2 g}{\partial t^2} + \gamma\frac{\partial^3 g}{\partial t\partial x^2} = -Q\delta(x)\delta(t), \quad (46)$$

which is identical to the VWE but with the bubble-radius coefficient β^2 set to zero. The absence of the bubble-radius term is responsible for a significant difference in behavior between Stokes' equation and the VWE. Whereas the VWE predicts nonphysical, instantaneous arrivals due to the presence of the bubble-radius term, the transient solutions of Stokes' equation[7] are perfectly well behaved at the instant the source is activated.

A. The pressure Green's function

The Green's function predicted by Stokes' equation may be obtained by setting $\beta=0$ in either of the VWE solutions, Eq. (30) or (31). In these equivalent solutions, there are no arrivals prior to the source being activated, and hence Stokes' equation, like the VWE, satisfies causality. With β set to zero, the normalized Green's function in Eq. (38) is plotted in Fig. 2(b), where it can be seen that no unphysical step discontinuity is present at the instant the source is activated. This is verified by setting $\beta=0$ under the integral in Eq. (32), to obtain the pressure at $t=0+$:

$$p(x,0+) = \frac{\rho_0 Q c_0^2}{2\pi}\int_{-\infty}^{\infty} e^{isx}ds = \rho_0 Q c_0^2\delta(x), \quad (47)$$

which indicates that, at the instant the source is activated, the pressure is identically zero everywhere in the medium apart from the source position itself. In fact, as discussed in detail by Buckingham,[7] the impulse response of Stokes' equation is maximally flat at the origin in time: not only the pulse itself but all of its time derivatives are zero at $t=0$, giving rise to a perfectly smooth transition from $t=0-$ to $t=0+$. Thus, as well as being causal, the transient solutions of Stokes' equation are physically realizable, implying a finite speed for all the propagating Fourier components and no instantaneous arrivals.

B. Two integral identities

The equivalence of the expressions in Eqs. (30) and (31) leads to an explicit expression for the integral in Eq. (31) for the special case when $\beta=0$ and $t=0+$. Under this condition, since Eq. (30) has the value given by Eq. (47), it follows that

$$\int_{-\infty}^{\infty}(1+i\omega\gamma)^{-1/2}\exp\left\{-\frac{i\omega|x|}{c_0\sqrt{1+i\omega\gamma}}\right\}e^{i\omega t}d\omega = 2\pi c_0\delta(x) \quad \text{for } t=0+. \quad (48)$$

Thus, the integral on the left is zero for all $|x|\neq 0$ when $t=0+$. The same integral is zero everywhere for $t\leqslant 0$, that is

$$\int_{-\infty}^{\infty}(1+i\omega\gamma)^{-1/2}\exp\left\{-\frac{i\omega|x|}{c_0\sqrt{1+i\omega\gamma}}\right\}e^{-i\omega t}d\omega = 0 \quad \text{for } |x|>0 \text{ and } t\geqslant 0. \quad (49)$$

C. Dispersion curves

The dispersion curves associated with Stokes' equation may be obtained directly from Eqs. (42a) and (42b) by setting $\beta=0$, under which condition the phase speed is

$$c_p(\omega) = \frac{\sqrt{2}c_0\sqrt{1+\omega^2\gamma^2}}{[1+\sqrt{1+\omega^2\gamma^2}]^{1/2}} \to \begin{cases} c_0 & \text{as } \omega\to 0 \\ c_0\sqrt{2\omega\gamma} & \text{as } \omega\to\infty \end{cases} \quad (50a)$$

and the attenuation is

$$\alpha_p(\omega) = \frac{|\omega|}{\sqrt{2}c_0}\frac{[\sqrt{1+\omega^2\gamma^2}-1]^{1/2}}{\sqrt{1+\omega^2\gamma^2}} \to \begin{cases} \dfrac{\gamma\omega^2}{2c_0} & \text{as } \omega\to 0 \\ \dfrac{1}{c_0}\sqrt{\dfrac{\omega}{2\gamma}} & \text{as } \omega\to\infty \end{cases}. \quad (50b)$$

Normalized versions of Eqs. (50a) and (50b) are plotted in Fig. 3(b) as functions of the dimensionless frequency $\omega\gamma$.

An important distinction between the VWE and Stokes' equation is illustrated in Fig. 3. At high frequencies, the VWE predicts a phase speed that increases as the square of the frequency and an attenuation that asymptotes to a finite, constant value, whereas Stokes' equation yields a phase speed and attenuation that both scale as the square-root of frequency. Thus, in the limit of high frequency, both equations predict infinitely fast Fourier components, which are infinitely attenuated in the case of Stokes' equation, but propagate finite distances according to the VWE. These infinitely fast, propagating waves are responsible for the unphysical discontinuity predicted by the VWE at the instant the source is activated. In contrast, the infinitely fast waves predicted by Stokes' equation do not propagate through the medium because they have zero amplitude, due to the infinite attenuation, and accordingly the solution is well behaved at the time the source is activated, showing no step discontinuity at $t=0$.

D. Kramers-Kronig

The real and imaginary parts of the complex compressibility from Stokes' equation are given by Eqs. (45a) and (42b) with β set to zero:

$$K_1(\omega) = \frac{1}{\rho_0 c_0^2(1+\omega^2\gamma^2)} \tag{51a}$$

and

$$K_2(\omega) = \frac{\omega\gamma}{\rho_0 c_0^2(1+\omega^2\gamma^2)}. \tag{51b}$$

Both of these functions have two simple poles symmetrically placed on the imaginary axis at $\omega = \pm i/\gamma$. The negative pole, of course, lies in the lower half-plane and therefore falls within the contour of integration used to derive the dispersion relations in Eqs. (14a) and (14b). It follows that, in both Eqs. (14a) and (14b), the residue of this pole is nonzero, thus making a finite contribution to the corresponding integrals in Eqs. (13a) and (13b).

It is almost trivial to verify that Eqs. (51a) and (51b) satisfy the Kramers-Kronig dispersion relationships, as expressed in Eqs. (14a) and (14b). Alternatively, the dispersion integrals in Eqs. (13a) and (13b) may be evaluated directly by factorizing the integrands and then making an elementary substitution, bearing in mind that the integral of an odd function taken over infinite limits is zero. Either way, it is found that Eqs. (51a) and (51b) do indeed satisfy Kramers-Kronig. Of course, this is only to be expected, since $K(\omega)$ has no poles on or below the real axis, hence $k(t)$ is causal and, from Titchmarsh's theorem, it follows that the Kramers-Kronig relations must be satisfied.

It is a straightforward matter to derive the full expression for the Stokes' equation compressibility $k(t)$, by taking the limit of Eq. (44) as β approaches zero. This procedure yields the expression

$$k(t) = u(t)\frac{1}{\rho_0 c_0^2 \gamma} e^{-t/\gamma}. \tag{52}$$

V. THE TIME-DEPENDENT DIFFUSION EQUATION (TDDE)

A commonly encountered partial differential equation of theoretical physics is the time-dependent diffusion equation (TDDE), the one-dimensional, inhomogeneous form of which is

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c_\infty^2}\frac{\partial^2 g}{\partial t^2} - \eta\frac{\partial g}{\partial t} = -Q\delta(x)\delta(t), \tag{53}$$

where η is a positive constant and c_∞ is the speed of sound in the medium in the limit of high frequency. Buckingham,[22] in the context of acoustic propagation through a porous medium with a rigid frame, developed the solution of Eq. (53) by applying two Fourier transforms, one temporal and the other spatial, to obtain an algebraic equation for the doubly transformed field. The corresponding inverse Fourier transforms were then applied, yielding the impulse response for the velocity potential:

$$g(x,t) = \frac{Qc_\infty}{2} u\left(t - \frac{|x|}{c_\infty}\right)\exp\left(-\frac{\eta c_\infty^2}{2}t\right) I_0\left(\frac{\eta c_\infty}{2}\sqrt{c_\infty^2 t^2 - x^2}\right), \tag{54}$$

where $I_0(..)$ is the modified Bessel function of the first kind of order zero. An equivalent solution can also be found in Morse and Feshbach,[23] derived in connection with the diffusion of heat in a gas.

It is clear from the presence of the step function that the solution in Eq. (54) is causal and physically realizable, since all the arrivals occur after a finite time: $t=t_\infty=|x|/c_\infty$. Unlike the case of the VWE, no instantaneous arrivals are predicted. As illustrated in Fig. 4, the onset of the pulse is abrupt, in contrast with the perfectly smooth impulse response predicted by Stokes' equation.

A Fourier transform of Eq. (53) with respect to time shows that the complex sound speed from the TDEE is given by

$$c(\omega) = \left(\frac{1}{c_\infty^2} - i\frac{\eta}{\omega}\right)^{-1/2}. \tag{55}$$

It follows that the phase speed is

$$c_p(\omega) = \sqrt{2}c_\infty\left[1 + \sqrt{1 + \frac{\eta^2 c_\infty^4}{\omega^2}}\right]^{-1/2} \rightarrow \begin{cases} \sqrt{\dfrac{2\omega}{\eta}} & \text{for } \omega \ll \eta c_\infty^2 \\ c_\infty & \text{for } \omega \gg \eta c_\infty^2 \end{cases} \tag{56a}$$

and the attenuation is

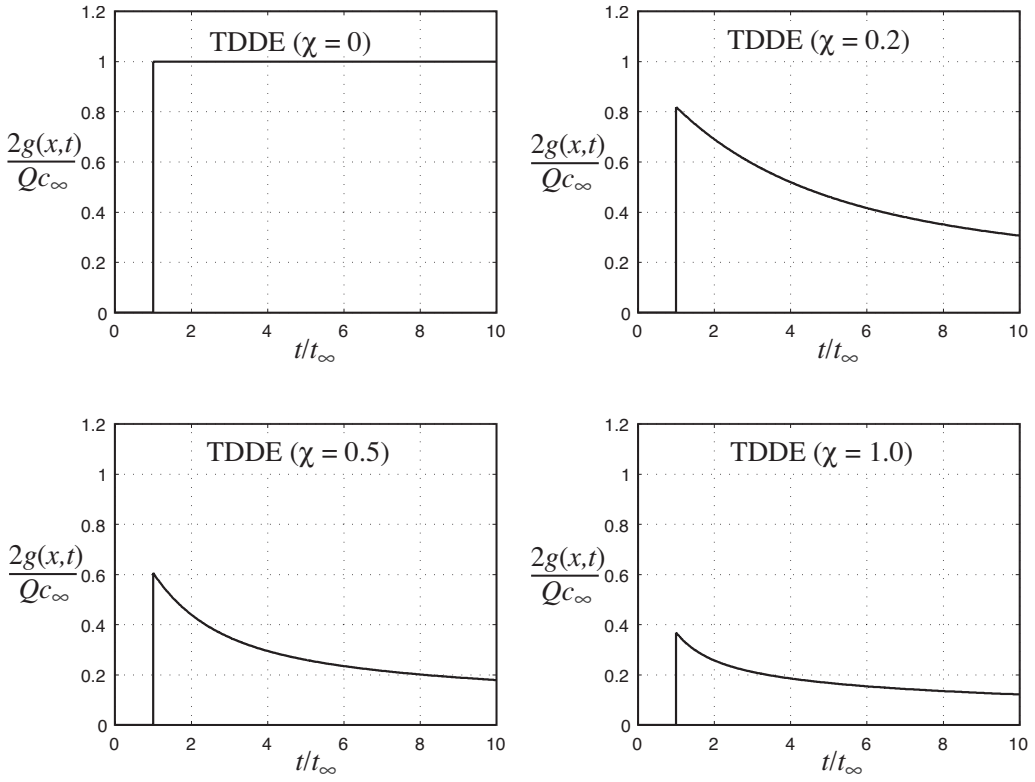


FIG. 4. Dimensionless velocity potential predicted by the TDDE, computed from Eq. (54), for four values of the parameter $\chi = \eta c_\infty |x|/2$.

$$\alpha_p(\omega) = \frac{\omega}{\sqrt{2}c_\infty}\left[\sqrt{1+\frac{\eta^2 c_\infty^4}{\omega^2}}-1\right]^{1/2} \rightarrow \begin{cases} \sqrt{\frac{\eta\omega}{2}} & \text{for } \omega \ll \eta c_\infty^2 \\ \frac{\eta c_\infty}{2} & \text{for } \omega \gg \eta c_\infty^2 \end{cases}. \tag{56b}$$

These dispersion equations are plotted in dimensionless form in Fig. 5, where it can be seen that the wave speed and attenuation are both finite in the limit of high frequency. Such behavior is consistent with the absence of instantaneous arrivals in the Green's function in Eq. (54).

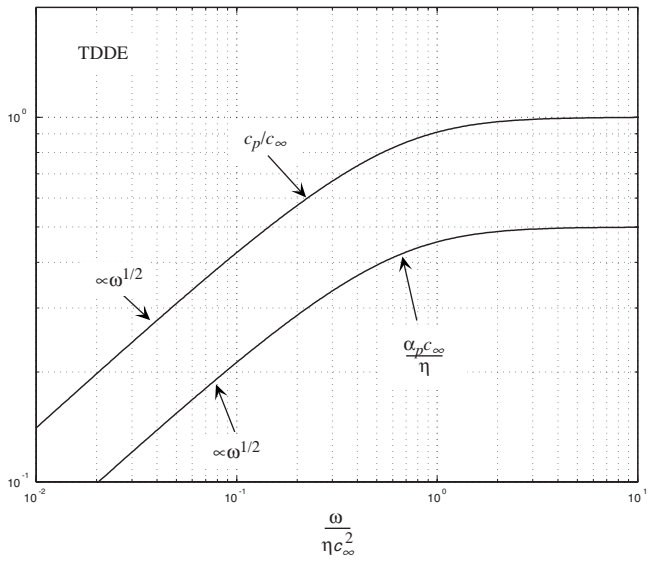


FIG. 5. Normalized dispersion curves for the TDDE, computed from Eqs. (56a) and (56b), as functions of the dimensionless frequency $\omega/\eta c_\infty^2$.

The complex compressibility is obtained directly from Eq. (55):

$$K(\omega) = \frac{1}{\rho_0 c^2(\omega)} = \frac{1}{\rho_0 c_\infty^2} - i\frac{\eta}{\rho_0\omega}, \tag{57}$$

from which

$$K_1(\omega) = \frac{1}{\rho_0 c_\infty^2} \tag{58a}$$

and

$$K_2(\omega) = \frac{\eta}{\rho_0\omega}. \tag{58b}$$

When substituted into the integrals in Eqs. (13a) and (13b), it is evident that these expressions fail to satisfy the Kramers-Kronig dispersion relations. In fact, both integrals return a value of zero for $\omega \neq 0$. The failure could, perhaps, have been anticipated, since the convergence condition in Eq. (8), which underpins the Kramers-Kronig relations, is not satisfied by the complex compressibility in Eqs. (56a) and (56b): the real part of $K(\omega)$ is independent of frequency and hence remains finite in the limit of high frequency.

A similar problem occurs with optical dispersion in a dielectric but is overcome by working with a function that is equivalent to $U(\omega) = K(\omega) - K_1(\infty)$, which, in the electromagnetic case, does satisfy the convergence condition in Eq. (8). Such a procedure does not solve the problem with the TDDE, however, even though $U(\omega)$ satisfies Eq. (8), as may

TABLE I. Causal properties of wave equations.

	Causal impulse response function?	Physically realizable impulse response function?	Kramers-Kronig satisfied?
van Wijngaarden's equation	Yes	No	Yes
Stokes' equation	Yes	Yes	Yes
Time-dependent diffusion equation	Yes	Yes	No

be appreciated from the fact that the real and imaginary parts of $U(\omega)$ fail to satisfy the Kramers-Kronig dispersion relation in Eq. (13b).

The fundamental difficulty with the TDDE is that neither $K(\omega)$ nor $U(\omega)$ is square-integrable and therefore the corresponding inverse Fourier transforms do not exist. But the causality argument that is central to the derivation of the Kramers-Kronig dispersion relations is predicated upon the existence of $k(t)$. Thus, the TDDE is an example of a wave equation that does not satisfy Kramers-Kronig but whose transient solutions are causal and physically realizable.

VI. CONCLUDING REMARKS

The impulse-response solutions of three wave equations, van Wijngaarden's equation (VWE), Stokes' equation, and the time-dependent diffusion equation (TDDE), have been developed in this article, and the causal properties of the equations themselves and their solutions examined in some depth. These properties are summarized in Table I, where it can be seen that all three equations return impulse-response functions that are strictly causal; that is, no acoustic arrivals are predicted anywhere in the medium prior to the activation of the source.

Stokes' equation for the propagation of acoustic waves in a viscous fluid is perfectly well behaved in that the impulse response is physically realizable, with no instantaneous arrivals predicted anywhere in the medium at a finite distance from the source; the complex compressibility satisfies the Kramers-Kronig dispersion relations. At first sight, the absence of instantaneous arrivals may be surprising since, at high frequencies, the phase speed increases without limit, scaling as frequency to the power of one-half. However, the attenuation follows the same scaling law, also increasing without limit as the square-root of frequency (Fig. 3(b)). Thus, the infinitely fast Fourier components are infinitely attenuated and hence do not propagate through the medium. In effect, because they have zero amplitude, such components make no contribution to the solution of Stokes' equation. Consequently, the impulse response is absolutely physical, and exhibits a particularly interesting feature at $t=0$, the time at which the source is activated: the response is maximally flat when $t=0$, that is, the impulse response itself and all its time derivatives are zero, giving rise to a perfectly smooth transition through the origin of time.

Taken in isolation, the fact that Stokes' equation predicts wave speeds that increase without limit could conceivably be construed as unphysical, but this would be a misinterpretation. The transient solutions of a wave equation are intimately connected to the frequency dependence of not just the wave speed but also the attenuation. A wave equation is physical provided it is causal and predicts no instantaneous arrivals, criteria which are both satisfied by Stokes' equation.

The same cannot be said of Van Wijngaarden's equation for the propagation of acoustic waves through a viscous, bubbly liquid. Although the VWE is causal and satisfies Kramers-Kronig, its impulse-response function is not physically realizable in that it predicts instantaneous acoustic arrivals everywhere in the medium. These nonphysical contributions to the solution are due solely to the presence of the bubble-radius term with nonzero coefficient β^2 in Eq. (22). This term has two effects on the high-frequency behavior of the dispersion curves [Fig. 3(a)]: above the transition frequency $\omega=\beta^{-1}$, the phase speed diverges as the square of frequency and the attenuation asymptotes to a constant, finite value. Thus, the bubble-radius term introduces indefinitely fast, finite-amplitude Fourier components into the solution for the impulse response, and these components propagate finite distances through the medium, arriving everywhere at the same instant that the source is activated. Such a phenomenon is physically impossible in any material that is subject to the constraints of continuum mechanics.

The time-dependent diffusion equation predicts an impulse-response function that is physically realizable, with no instantaneous arrivals, but its complex compressibility does not satisfy the Kramers-Kronig dispersion relations. This failure is interesting because Kramers-Kronig is generally taken as a definitive test of causality. So, here is a situation where a wave equation possesses a causal, physically realizable impulse response, yet, judged on the basis of Kramers-Kronig, would be said to be noncausal. The origin of the contradiction lies with the Kramers-Kronig relations themselves. They are predicated upon the causal behavior of the compressibility $k(t)$, which is the inverse Fourier transform of the complex compressibility $K(\omega)$. But, in this case, $K(\omega)$ is not square-integrable, therefore $k(t)$ does not exist (mathematically), and hence the argument leading to the Kramers-Kronig relations fails.

A few final comments are perhaps in order. An impulsive source, as considered in this article, has an infinite bandwidth and is thus an idealized representation of any actual physical source, which is always band-limited. Now, the solution for the acoustic field produced by an impulsive source is the Green's function $g(t)$. If a finite-bandwidth source had been considered, with pulse shape $f(t)$, then it is readily shown that the solution for the time-dependent field is the convolution $f(t)\otimes g(t)$, which, in the frequency domain, becomes the product $F(\omega)G(\omega)$, where the upper case letters are the Fourier transforms of their lower case counterparts.

Since $F(\omega)$ is zero beyond some upper-frequency limit, the solution for the field due to the band-limited source does not contain indefinitely high frequencies. Although $f(t)$ will affect shapes of the transient solutions of Stokes' equation and the TDDE, the band-limited source introduces no fundamentally new behavior into these solutions. In the case of the VWE, however, the absence of indefinitely high frequencies in the source function means that instantaneous arrivals do not appear in the solution for the band-limited transient field. In fact, if the source bandwidth were reduced below β^{-1}, the bubble-radius term would, in effect, become neutralized and the solution of the VWE would reduce to that of Stokes' equation.

All of which leads to the question: Is the VWE a valid descriptor of acoustic propagation in a viscous, bubbly liquid or, indeed, in any physically realizable medium? Arguably, the VWE is invalid because, in the limit of high frequency, it returns nonphysical solutions. Just because infinitely high frequencies cannot be produced in practice does not alter the fact that the predicted field must show the correct limiting behavior, which, of course, the VWE fails to do. The situation is much like that encountered with a radiation condition, where the field must go to zero infinitely far from the source. The fact that a receiver could never be placed at an infinite distance from the source does not invalidate the requirement that the solution for the field must satisfy the boundary condition at infinity. Similarly, any wave equation which admits a solution that is nonphysical in any part of the spectrum, including the region extending to indefinitely high frequencies, must be regarded as flawed. In the case of the VWE, the instantaneous arrivals are attributable directly to the presence of the bubble-radius term, suggesting either that this term is incorrect or that another term needs to be included which would modify the high-frequency response. Be that as it may, the instantaneous arrivals represent a deficiency of the VWE which, accordingly, should be treated with caution.

ACKNOWLEDGMENTS

The research was supported by Drs. Ellen Livingston and Robert Headrick, Ocean Acoustics Code, the Office of Naval Research, under Grant Number N00014-07-1-0109.

[1] L. van Wijngaarden, "One-dimensional flow of liquids containing small gas bubbles," Annu. Rev. Fluid Mech. **4**, 369–396 (1972).
[2] G. G. Stokes, "On the theories of the internal friction of fluids in motion and of the equilibrium and motion of elastic solids," Trans. Cambridge Philos. Soc. **8**, 287–319 (1845).
[3] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. (Academic Press, San Diego, 1985).
[4] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), p. 254.
[5] P. M. Jordan, M. R. Meyer, and A. Puri, "Causal implications of viscous damping in compressible fluid flows," Phys. Rev. E **62**, 7918–7926 (2000).
[6] D. T. Blackstock, "Transient solution for sound radiated into a viscous fluid," J. Acoust. Soc. Am. **41**, 1312–1319 (1967).
[7] M. J. Buckingham, "Causality, Stokes' wave equation and acoustic pulse propagation in a viscous fluid," Phys. Rev. E **72**, 026610 (2005).
[8] J. S. Toll, "Causality and the dispersion relation: logical foundations," Phys. Rev. **104**, 1760–1770 (1956).
[9] J. Hilgevoord, *Dispersion Relations and Causal Description* (North-Holland, Amsterdam, 1960).
[10] H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic Press, New York, 1972).
[11] H. A. Kramers, "Some remarks on the theory of absorption and refraction of x-rays," Nature (London) **117**, 775 (1926).
[12] R. de L. Kronig, "On the theory of the dispersion of x-rays," J. Opt. Soc. Am. **12**, 547–557 (1926).
[13] V. L. Ginzberg, "Concerning the general relationship between absorption and dispersion of sound waves," Sov. Phys. Acoust. **1**, 32–41 (1955).
[14] V. Mangulis, "Kramers-Kronig or dispersion relations in acoustics," J. Acoust. Soc. Am. **36**, 211–212 (1964).
[15] R. L. Weaver and Y.-H. Pao, "Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media," J. Math. Phys. **22**, 1909–1918 (1981).
[16] C. C. Lee, M. Lahham, and B. G. Martin, "Experimental verification of the Kramers-Kronig relationship for acoustic waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 286–294 (1990).
[17] M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[18] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (John Wiley, New York, 1982), p. 99.
[19] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. (Oxford University Press, London, 1937).
[20] P. M. Jordan and C. Feuillade, "On the propagation of transient acoustic waves in isothermal bubbly liquids," Phys. Lett. A **350**, 56–62 (2006).
[21] A. Erdélyi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. **1**, p. 8.
[22] M. J. Buckingham, "Acoustic pulse propagation in dispersive media," in *New Perspectives on Problems in Classical and Quantum Physics. Part II. Acoustic Propagation and Scattering-Electromagnetic Scattering*, edited by P. P. Delsanto and A. W. Sáenz (Gordon and Breach, Amsterdam, 1998), Vol. **2**, pp. 19–34.
[23] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics: Part 1* (McGraw-Hill, New York, 1953), p. 867.

On sound propagation from a slanted side branch into an infinitely long rectangular duct

S. K. Tang[a] and G. C. Y. Lam
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

(Received 6 September 2007; revised 4 March 2008; accepted 10 July 2008)

The transmission of sound from a slanted side branch into an infinitely long rectangular duct is studied numerically using the method of finite element with absorptive domain exit boundaries. The sound transmission coefficients associated with various acoustic modes are investigated in details. The results show that the plane wave assumption is only valid at very low frequency. It is also found that the intensities of the higher modes are stronger than that of the plane wave once they are excited. Besides, a critical side-branch slant angle is found over which a significant change of sound propagation mode takes place. This affects substantially the energy distribution between various acoustic modes inside the main duct. A simplified model is proposed to explain the phenomenon and the relationship of this critical angle with the width ratio between the side branch and the main duct is established. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2968671]

PACS number(s): 43.20.Mv [LT] Pages: 1921–1929

I. INTRODUCTION

Complicated duct works are very common in modernized air-conditioned buildings. Their primary functions are to convey the treated air with the right temperature and humidity into the occupied zones and to provide a controlled route for the extraction of the used air out of the same areas. However, noise from the ventilation systems will propagate along these duct works and thus affect the indoor acoustical environment.[1]

Duct junctions are very frequently found in the ventilation ducting network. Usually, they are found at locations where a smaller duct merged into a main duct or where a main air duct forks into smaller ducts. The noise propagation between these ducts depends on the junction configuration. Conventionally, duct junctions are right angled and the noise transmission across these junctions has attracted the attention of many researchers for improved prediction and duct noise control. Typical examples include the works of Miles,[2] Bruggeman,[3] Redmore and Mulholland,[4] Dubos *et al.*,[5] and more recently Tang.[6] Effects of coupled side branches on sound transmission along a main duct were examined experimentally by Tang and Li.[7] There are also related studies such as that of Davies[8] for connected corridors with rectangular cross sections, that of Lippert[9] for right-angled duct bends, and those related to waveguides (for instance, Razavy[10] and Lee *et al.*[11]). However, many of these studies are restricted in the low frequency range because the waves involved can be analytically handled and the plane waves can be measured readily using the conventional transfer function method.[12]

In a heavily serviced building, many services are located in the very congested ceiling voids and very often the right-angle junction configuration does result in additional constraints in the space utilization. Also, such junction will also be accompanied with flow bends, resulting in high static pressure losses in the duct work. A slanted junction appears to be able to help reduce the spatial constraint and to some extent smoothing the flows around the junctions, resulting in smaller pressure loss and weaker turbulence intensity, which in turn has an impact on the aeroacoustic radiation (for instance, Dequand *et al.*[13]). Slanted junctions may also help reduce the amount of material used.

The present study is an attempt to understand how sound propagates from a long slanted side branch into an infinitely long duct in the absence of a flow. The cross section of the duct and side branch are taken to be rectangular as rectangular ducts are the most commonly found duct type in commercial buildings with centralized air-conditioning and ventilation systems. The duct and the side branch are of the same spanwise length (very common case as well) such that no excitation in the spanwise direction is possible. The present adopted system is exactly a two-dimensional one. The interactions between the various duct modes are examined. The finite element method adopted in the recent studies of S.K.T. is used.[6,14] It is hoped that the present results can provide information for improved acoustical design of the air-conditioning and ventilation duct work. The present study also forms part of a comprehensive investigation on the sound propagation across duct junctions with and without flows.

II. THE COMPUTATIONAL DOMAINS

Figure 1 shows the junction configuration and the nomenclature adopted in the present study. In practice, the side branches are not wider than the main duct and thus $b \leqslant a$. One should note that the interfacial length between the slanted side branch and the main duct varies with b and the slant angle θ. In the present study, a plane wave is set to propagate toward the junction from the slanted side branch.

The finite element computational approach of Tang and Lau[14] and Lau and Tang[15] is adopted in the present study. The wave equation to be solved is the Helmholtz equation:

[a]Author to whom correspondence should be addressed. Electronic mail: besktang@polyu.edu.hk

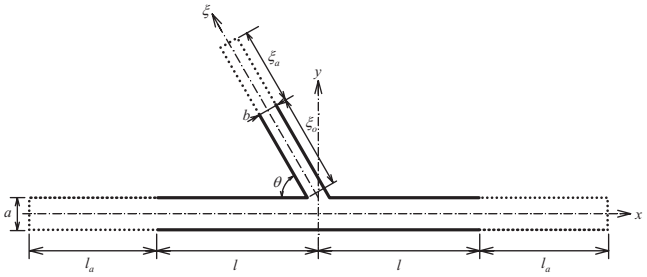


FIG. 1. Schematic of the slanted side-branch junction and nomenclature. (Dotted lines represent locations of absorptive boundaries.)

$$\nabla^2 p + k^2 p = q, \tag{1}$$

where q is the source, p is the acoustic pressure, and k the sound wave number. For the low frequency range where the frequency of sound is below the first eigenfrequency of the main duct, one can expect that only plane waves exist at locations far away from the junction. The boundary conditions can be analytically written as

$$\frac{\partial p}{\partial n} + ikp = \begin{cases} 2ik & \text{at } \xi = \xi_0 \\ 0 & \text{at } x = l \\ 0 & \text{at } x = -l, \end{cases} \tag{2}$$

where $i=\sqrt{-1}$, n represents the unit outward normal of a surface and $q=0$. The computational domain is bounded within $|x| \leqslant l$ and $0 \leqslant \xi \leqslant \xi_0$. Outgoing plane wave condition at $x=l$ and $x=-l$ are adopted, while the condition at $\xi=\xi_0$ represents a unit plane wave excitation together with a plane wave reflection in the positive ξ direction. For higher frequencies, higher modes exist and the plane wave excitation is, as depicted in Tang and Lau,[14]

$$q = 2ke^{i\gamma}\delta(\xi - \xi_0), \tag{3}$$

where δ is delta function and γ is an arbitrary phase angle of the excitation. The latter is not important in the present study and without loss of generality, it is set at $\pi/2$. The computation domain is extended to cover the regions with absorptive boundaries (dotted lines in Fig. 1) having

$$\frac{\partial p}{\partial n} + ik\alpha p = 0, \tag{4}$$

where α is an arbitrary absorption coefficient which varies with distance Δ from each interface between the normal domain and the absorptive domain. It vanishes at the location of the interface and increases very slowly in the beginning to ensure that no numerical reflection is produced at the interfaces. In the present study, $\alpha=0.01\Delta^2$.

The higher modes are completely damped within these regions before they reach the ends of the absorptive domains, where the outgoing plane wave conditions are again adopted. More details of these settings and their performance can be found in Tang and Lau[14] and Lau and Tang[15] and thus are not repeated here. In the present study, l is set at $20a$ and $\xi_0=20b$. The lengths of the absorptive regions are $30a$ for the main duct (l_a) and $30b$ for the slanted side branch (ξ_a).

The finite element computations are implemented using the software MATLAB[16] and the mesh generation is done through the Delaunay triangular algorithm.[17] Figure 2 illustrates an example of the meshes adopted in the higher frequency computations. For the meshes adopted in the present study, over 95% of the mesh triangles have a mesh quality index greater than the acceptable level of 0.6.[16] Also, the largest mesh has its length approximately 1/26 of the smallest wavelength involved in the present computation. It has been checked that any further refinement of the meshes will not produce noticeable variations in the results.

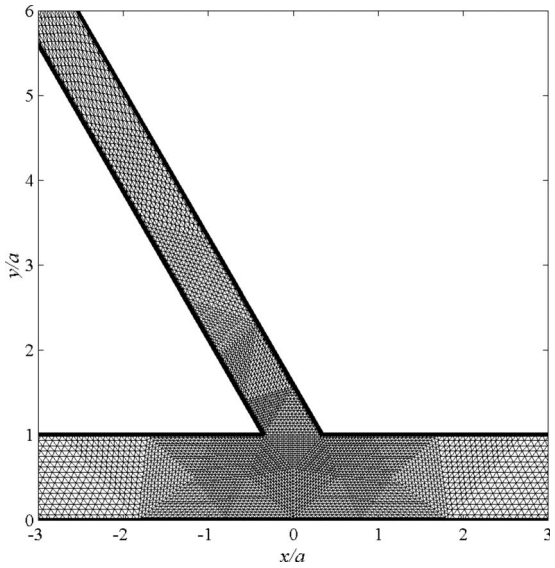


FIG. 2. Example of the finite element meshes adopted for $b/a=0.6$ and $\theta=60°$.

III. SOME THEORETICAL CONSIDERATIONS

In general, analytical solution for sound propagation across the slanted junction is very difficult if not impossible unless the sound frequency is very low. Peake[18] demonstrated the use of the Wiener–Hopf technique for the slanted nozzle exit, but an analytical solution is still not possible even at low frequencies.

When the wavelength of the sound is much larger than the physical dimension of the junction cross section, the acoustic pressure within a large region around the junction can be assumed to be uniform. Standard technique for plane wave calculation[19] suggests that the sound pressure magnitude in the main duct is, for a unit strength plane wave excitation in the slanted branch,

$$|p_d^+| = |p_d^-| = \frac{b}{a+0.5b} \Rightarrow \tau^+ = \tau^- = \frac{ab}{(a+0.5b)^2}, \tag{5}$$

where τ is sound power transmission coefficient, the suffix d denotes the quantity associated with the main duct, and the positive and negative prefixes denote downstream and upstream of the junction, respectively. The magnitude of the corresponding reflected wave in the side branch, $|p_r|$, is

$$|p_r| = \frac{2a-b}{2a+b}. \tag{6}$$

It should be noted that Eqs. (5) and (6) are independent of the slant angle θ at this low frequency situation.

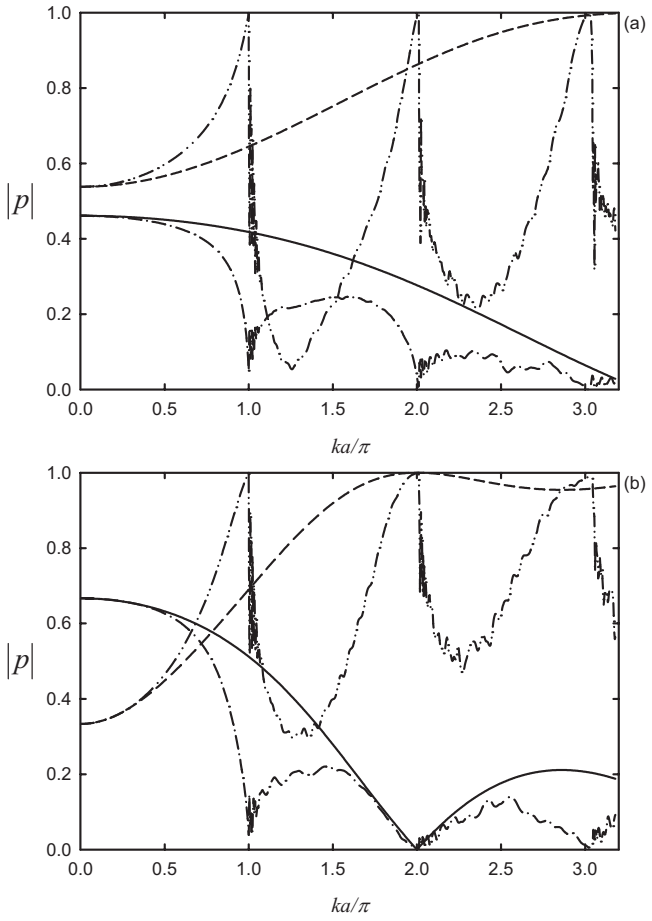


FIG. 3. Comparison between plane wave theory and numerical results for plane wave magnitudes for $\theta=90°$. (a) $b/a=0.6$ and (b) $b/a=1$. (—) In main duct from plane wave theory; (---) in side branch from plane wave theory; (–·–) in main duct from finite element computation; and (–···–) in side branch from finite element computation.

Numerical studies adopting the mode coupling technique or the finite element method have been done for $\theta=90°$ (for instance, Redmore and Mulholland[4] and Tang[6]). However, analytical solution for this slant angle is only possible when all the higher modes and the evanescent waves are ignored. Assuming the exit of the side branch acts as a massless piston exciting plane wave propagation inside the main duct, it is not difficult to show that for an unit strength plane wave excitation in the branch,

$$p = \frac{2\sin(kb/2)}{ka - i\left[1 - e^{-ikb/2}\dfrac{\sin(kb/2)}{kb/2}\right]} e^{-ik(|x|+\phi)}, \tag{7}$$

where ϕ is a constant related to the phase of the wave excitation inside the side branch and use has been made of the analytical formulas provided by Huang[20] for purely plane wave motion. Equation (7) converges to Eq. (5) when $ka \to 0$, and thus $kb \to 0$ as $a \geqslant b$. However, Eq. (7) overestimates the planar sound powers inside the main duct in the presence of higher modes or evanescent waves as shown in Fig. 3. The corresponding reflected wave is

$$p_r = \frac{1 - ika - e^{-ikb/2}\dfrac{\sin(kb/2)}{kb/2}}{1 + ika - e^{-ikb/2}\dfrac{\sin(kb/2)}{kb/2}} e^{-ik(y+\phi)}. \tag{8}$$

Again, Eq. (8) converges to Eq. (6) when $ka \to 0$. However, Eq. (8) does not model well the reflected wave in the side branch (Fig. 3). In conclusion, the plane wave assumption can only predict the wave magnitudes for $ka < 0.2\pi$ at $b/a=0.6$ [Fig. 3(a)] and for $ka < 0.4\pi$ at $b=a$ [Fig. 3(b)]. Though it seems that Eqs. (6) and (8) are able to give better predictions for $b=a$ than for $b=0.6a$, one should note that the controlling parameter in the plane wave assumption is b while the duct eigenfrequencies are more important in reality and they are related to a. One should also note that $\theta=90°$ is a very special case in the present study because asymmetric modes are not able to propagate inside the side branch under this right-angled junction situation.

IV. RESULTS AND DISCUSSIONS

In the foregoing discussions, the magnitudes of the plane waves and those of the higher modes inside the main duct are obtained at a distance $19a$ from the junction center (before any absorptive boundary) using the conventional orthogonal modal decomposition technique.[21] The corresponding location is $\xi=19b$ in the side branch. It should be noted that the magnitudes of the waves in the main duct and those of the reflected waves inside the side branch are independent of distance from the junction provided that the latter is larger than twice the duct/branch width.[22] It has been tested that the adopted long absorptive boundaries are able to eliminate backreflections into the computational region with rigid boundaries.

The magnitudes of the plane waves transmitted into the main duct and reflected back into the side branch are illustrated in Fig. 4. Those for $\theta=90°$, which have been shown in Fig. 3, are also presented here for the sake of easy comparison.

Concerning the plane wave propagating downstream of the junction in the main duct [Fig. 4(a)], the wave pressure magnitude (and thus the energy it carries) increases with decreasing θ when the sound frequency is low enough and then falls sharply at ka approaches π (the first main duct eigenfrequency). The larger the slant angle, the more rapid the fall of the wave magnitude. The wave magnitude also increases slightly as frequency increases for $\theta < 45°$ and $ka < \pi$, but the opposite is observed when $\theta \geqslant 52.5°$. The wave magnitude increases with the sound frequency for a relatively narrow ka region after $ka=\pi$ in general. For $\theta \leqslant 45°$, the variation of the wave magnitude with ka appears to be wavier and a local minimum is found at $ka \sim 1.57\pi$. The curves in Fig. 4(a), except that for $\theta=90°$, appear to converge at $ka=5\pi/3$, which corresponds to the first eigenfrequency of the side branch. It should be noted that there is no asymmetric mode propagation in the side branch for $\theta=90°$ as mentioned before.

The cut-on of the first asymmetric mode in the side branch also marks the beginning of a frequency region where

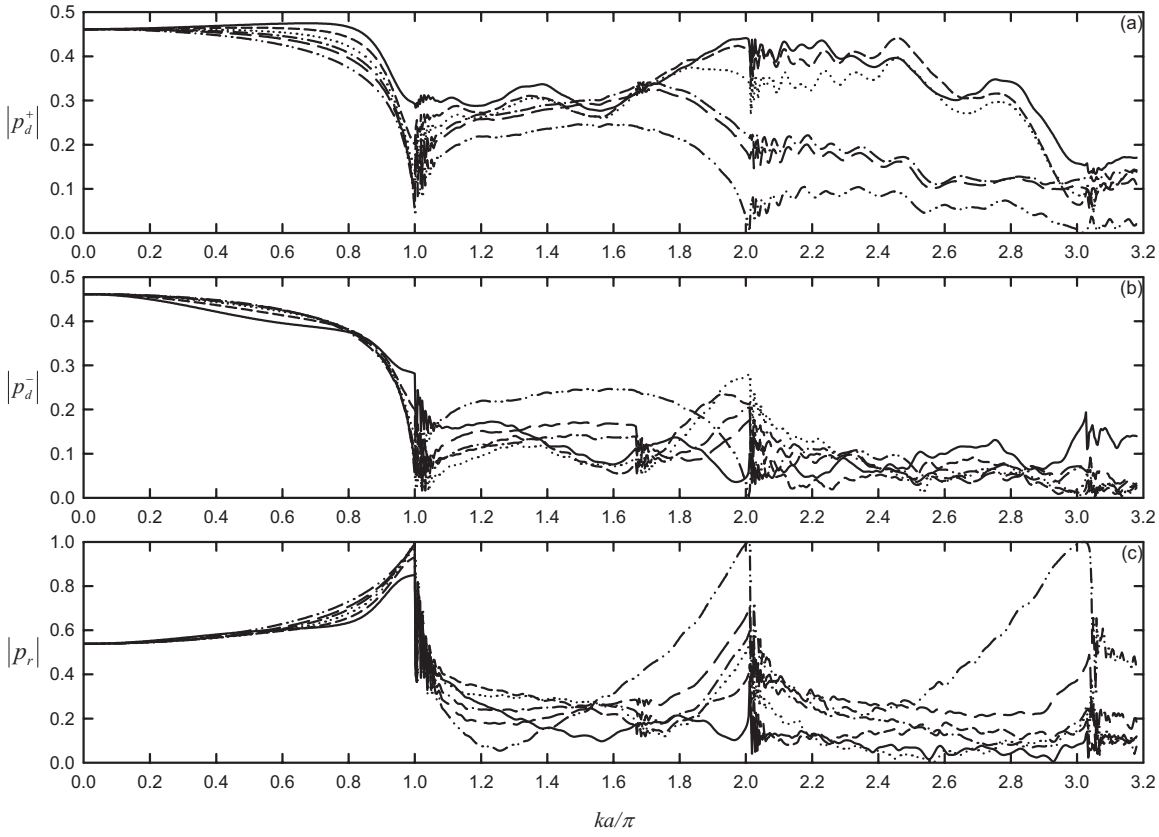


FIG. 4. Variations of plane wave magnitudes with frequency for $b/a=0.6$. (a) Downstream going wave in main duct; (b) upstream going wave in main duct; and (c) upstream going wave in the side branch. (—) $\theta=30°$; (---) $\theta=37.5°$; (···) $\theta=45°$; (–·–) $\theta=52.5°$; (——) $\theta=60°$; (–···–) $\theta=90°$.

the plane wave magnitude either continues to rise or starts to fall upon further increase of the sound frequency, and the transition between these two distinctive phenomena actually takes place within a very narrow range of θ [Fig. 4(a)]. For $b/a=0.6$, the transition takes place within $45° \leqslant \theta \leqslant 52.5°$ and the critical slant angle, θ_c, is expected to depend on b/a. This will be discussed and explained further later. A fall of the wave magnitude is again observed at $ka \sim 2\pi$, when the second eigenmode of the main duct is excited. It should be noticed that the plane wave magnitude after the excitation of the first eigenmode of the main duct for $45° < \theta < 60°$ is comparable to that at low frequency, but it does not mean that the acoustical energy contained in the higher modes will be insignificant within this slant angle range. Again, further discussions will be given later.

The plane wave propagating upstream of the junction carries less amount of acoustical energy than that propagating downstream for all slant angles studied as illustrated in Fig. 4(b). Before the first eigenfrequency of the main duct and for $ka < 0.8\pi$, the wave magnitude increases with increasing θ, because of the weaker diffraction round a more acute angle. This trend reverses at $ka > 0.8\pi$ followed by a rapid fall of the wave magnitude as $ka \to \pi$. The results in Fig. 4(b) are also indicating a change in the mode of wave propagation takes place within $45° \leqslant \theta \leqslant 52.5°$. One can notice that the cut-on of the first asymmetric higher mode in the side branch has a stronger effect on this upstream going plane wave than on the downstream going plane wave, especially for large slant angles (except $\theta=90°$).

At low frequency, the magnitude of the reflected wave increases with increasing slant angle and sound frequency as shown in Fig. 4(c). This is rather expected as stronger reflection back into the side branch takes place at larger slant angle. One can also notice from Fig. 4 that the acoustical energy carried by the plane wave is not high for $ka > 2\pi$. This energy is insignificant for small slant angles. The propagation of the higher modes will be discussed later.

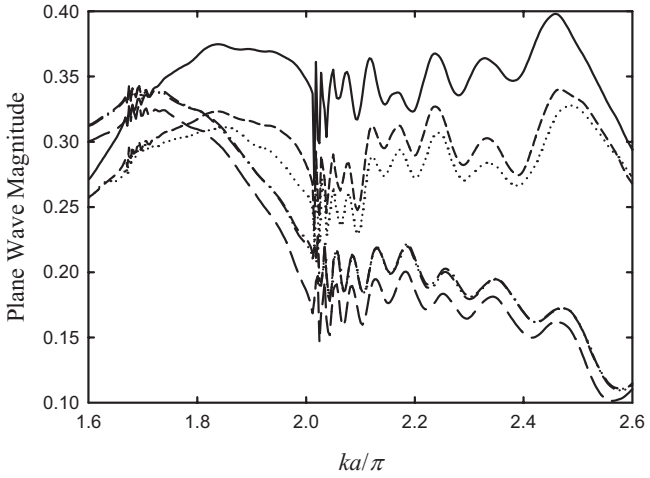


FIG. 5. Effects of slant angle on the downstream going plane wave magnitudes after the cut-on of the side-branch higher mode. (—) $\theta=45°$; (---) $\theta=49°$; (···) $\theta=52°$; (–·–) $\theta=52.5°$; (——) $\theta=53°$; (–···–) $\theta=60°$.

Figure 5 illustrates the plane wave magnitude in the main duct downstream of the junction for $45° < \theta < 60°$ and $1.6 < ka/\pi < 2.6$, which is the parameter range in which a change of the propagation mode of the plane wave is expected to occur. One can observe that the change takes place within a very narrow θ range at $\sim 52°$ to $52.5°$ and this change is abrupt. The results for $\theta = 53°$ almost overlap with those for $\theta = 52.5°$.

One can expect that a plane wave is more easily forced if there is a more uniform acoustical energy distribution at the region of the main duct just downstream of the junction. Plane wave passes through the slanted exit of the side branch, hits and reflected by the opposite main duct wall to set up the wave pattern at the junction. A very asymmetric pattern does not favor plane wave propagation and this happens when the slant angle is large as shown in Fig. 6(a) ($\theta = 90°$ and $ka = 1.8\pi$). At this condition, standing wave is setup inside the junction, making the acoustic pressure at the exit of the junction far from uniform. One should note that the acoustic pressure in the middle of the junction is out of phase with that at the opposite wall inside the junction in Fig. 6(a).

The corresponding acoustic pressure patterns for $\theta = 52.5°$ and $30°$ are given in Figs. 6(b) and 6(c), respectively. Though the acoustic pressures at the junction exit for these slant angles are still not uniform, that for $\theta = 30°$ appears to be more uniform and thus a stronger plane wave can propagate downstream. It is observed from Fig. 6 that the highest pressure point on the duct wall opposite the slanted side-branch exit moves from $x/a = 0$ to $x/a \sim 0.5$ when θ decreases from 90° to 30°. The plane wave appears to be bended toward the opposite duct wall as it leaves the side branch, probably because of the diffraction. Its energy is gradually spreading out within the junction at the same time. Very high wave pressure magnitude is localized near to the right-hand side of the side-branch exit in Fig. 6(b), suggesting that some sort of constructive interference is taking place there. This is probably due to the reflected wave from the opposite duct wall and thus a portion of the wave energy originally propagating inside the side branch is reflected back into the latter, reducing the acoustical energy downstream of the main duct. Such interference and backreflection are not observed when the slant angle θ is reduced to 30°. This explains the existence of a critical slant angle θ_c below which the sound energy can go easily into the main duct shown previously in Figs. 4 and 5. Such change of wave propagation mode should be abrupt as the mentioned interference will be much weakened at a slant angle slightly less than θ_c where most of the significant reflection from the opposite duct wall in the junction is reflected into the downstream duct. This effect also favors higher mode propagation, which will be discussed later.

A simplified schematic of the above-mentioned sound propagation mode at $\theta = \theta_c$ is summarized in Fig. 7. The arrow represents the direction of sound ray originated from the left-hand corner of the side-branch exit. The diffraction angle θ_d (angle ABC in Fig. 7) is theoretically unknown. The sound ray eventually hits the right-hand corner of the side-branch exit provided that

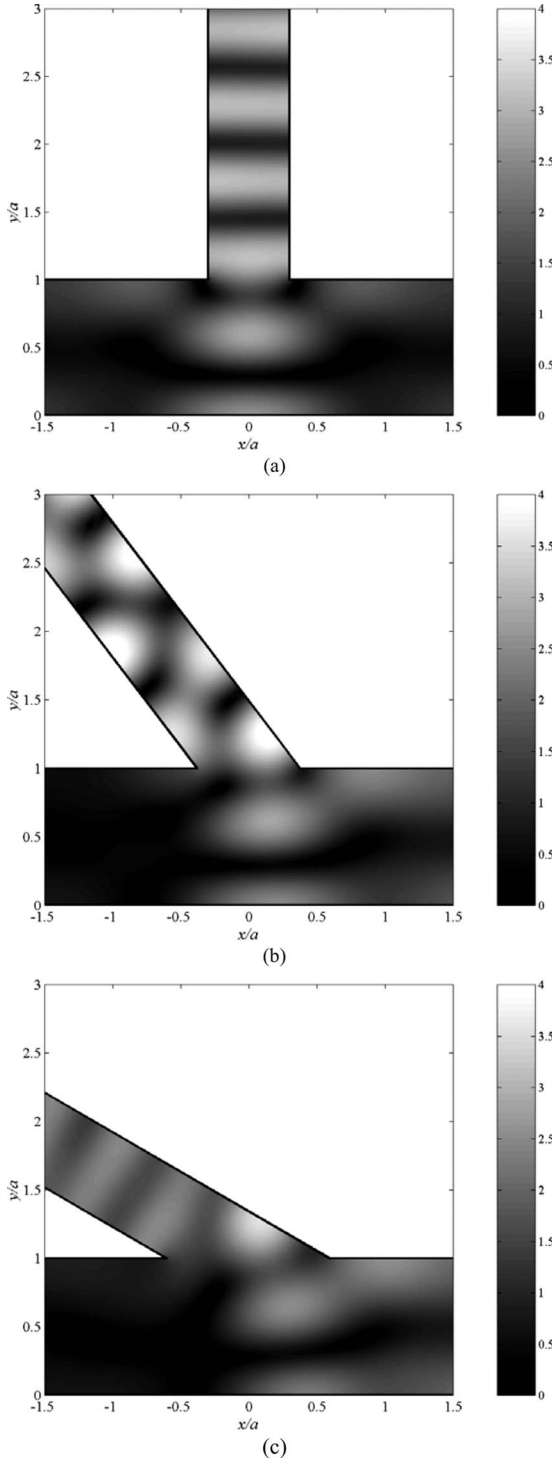


FIG. 6. Examples of sound magnitude distribution within the junction at $ka = 5.65$. (a) $\theta = 90°$; (b) $\theta = 52.5°$; and (c) $\theta = 30°$.

$$\alpha = \tan^{-1}(2a \sin \theta_c/b) = \theta_c + \theta_d. \tag{9}$$

The second part of Eq. (9) is obtained by trigonometry. Since the finite element computation suggests $\theta_c \sim 52.5°$ for $b/a = 0.6$, $\theta_d \sim 16.7°$. One can observe from Fig. 6(b) that the latter is roughly the direction beyond which the acoustical energy inside the main duct upstream of the junction center becomes insignificant. Similar θ_d is observed in Fig. 6(c) and even in Fig. 6(a). It will be shown later that both θ_d and θ_c decrease as b/a increases. The results of Figs. 4(a) and 6 also

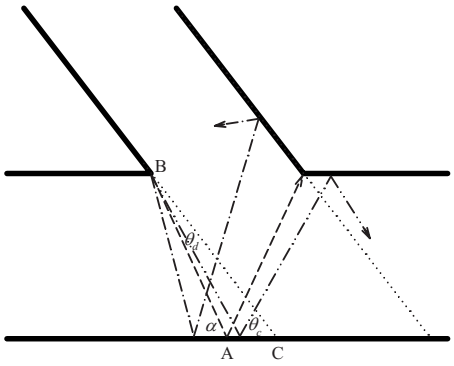


FIG. 7. Schematic for sound reflections within the junction. (---) $\theta=\theta_c$; (–·–) $\theta>\theta_c$; (–··–) $\theta<\theta_c$; (···) imaginary slanted side-branch extension.

tend to suggest that θ_d is not a strong function of frequency or side-branch slant angle.

For $ka>\pi$, the first asymmetric higher duct mode propagates inside the main duct. The larger the slant angle, the weaker the mode magnitude in the main duct section downstream of the junction before the cut-on of the first symmetrical higher duct mode ($m=1$) as shown in Fig. 8(a). The magnitude of the first symmetric higher duct mode ($m=2$) appears to be the highest at $\theta=\theta_c$ [Fig. 8(b)]. This is not surprising because the constructive interference that takes place at the left-hand corner of the side-branch exit tends to create an acoustic pressure pattern which matches with that of the $m=2$ duct mode at the entry of the downstream duct section. This phenomenon is illustrated in Fig. 9. The results shown in Fig. 8(b) also suggest that the $m=2$ symmetrical

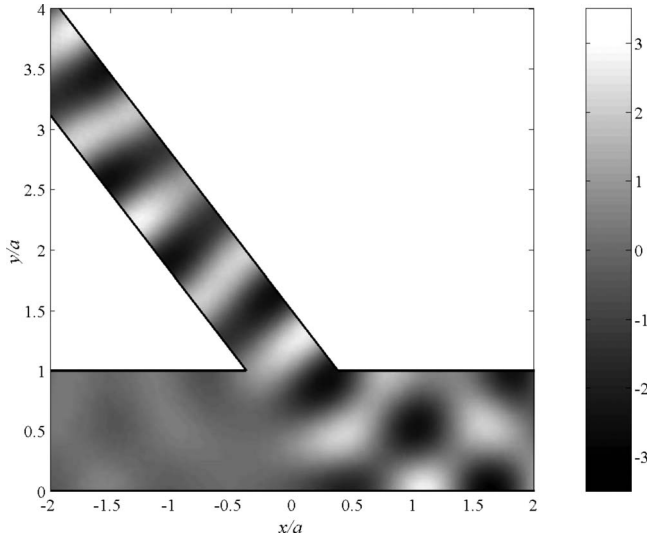


FIG. 9. Sound wave pattern within the junction for $\theta=52.5°$, $ka=8$.

mode is stronger than the $m=1$ asymmetric mode for $ka>2\pi$. Besides, one can notice that the acoustic powers carried along by the higher modes concerned (proportional to $|p_d^+|^2/2$) are stronger than that associated with the plane wave.

Figure 8(c) shows the magnitude of the asymmetric mode inside the side branch. After the cut-on of the side-branch higher mode and before the cut-on of the symmetrical mode of the main duct, the mode magnitude appears to decrease with increasing θ for $\theta<\theta_c$. This trend is opposite to

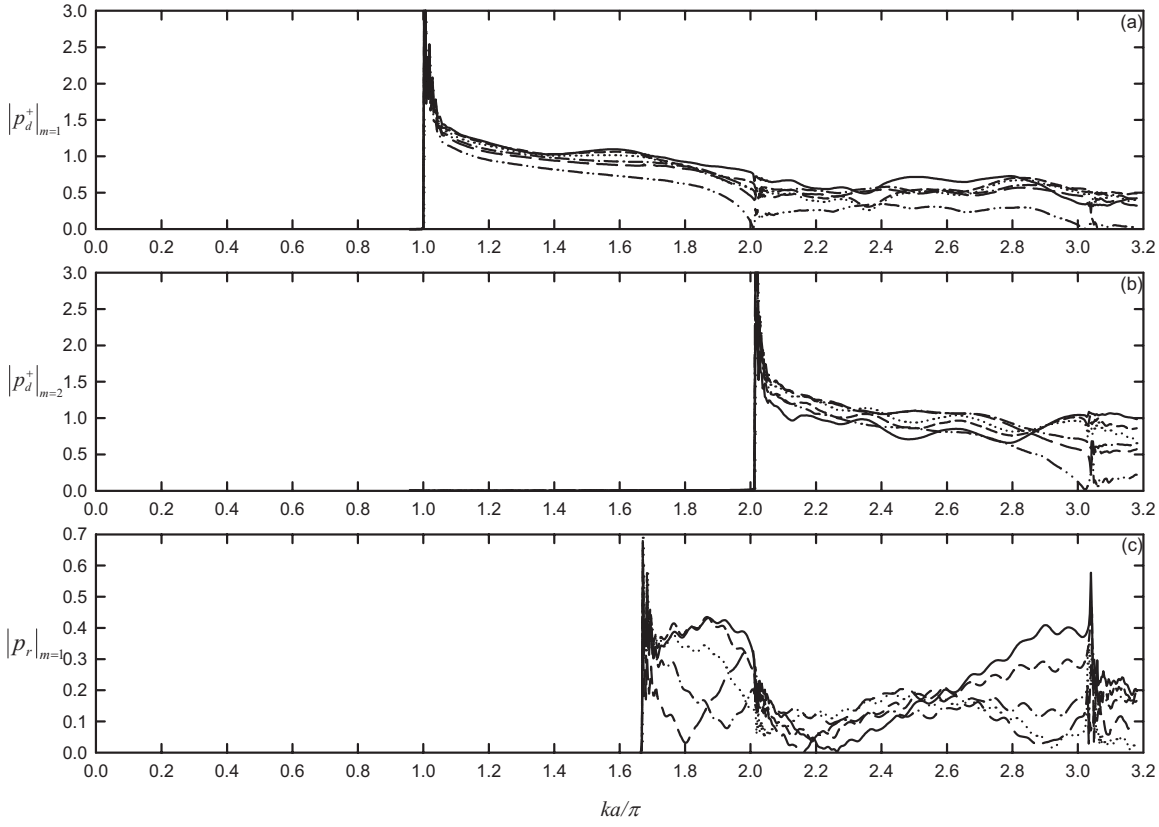


FIG. 8. Variations of higher mode magnitudes with frequency for $b/a=0.6$. (a) Downstream going $m=1$ mode in main duct; (b) downstream going $m=2$ mode in main duct; and (c) upstream going first asymmetric mode in the side branch. Legends: Same as those in Fig. 4.

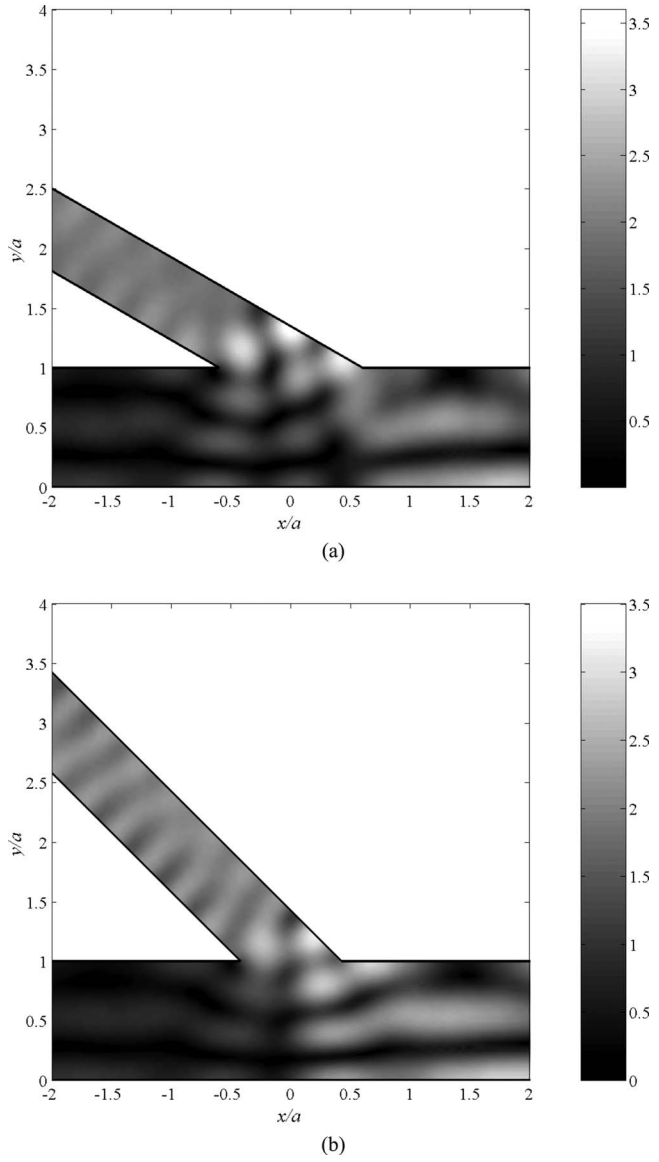


FIG. 10. Examples of sound magnitude pattern within the junction at $ka=9$. (a) $\theta=30°$ and (b) $\theta=45°$.

that of the plane wave shown in Fig. 4(c). Larger variation of the wave magnitude is found for larger slant angle. The mode magnitude rises up for $ka \to 2\pi$ near which the second asymmetric mode ($m=3$) starts to propagate along the main duct. Though the mode magnitudes at those frequencies are high for $\theta \leqslant 37.5°$, it does not increase monotonically with θ. The mode magnitude is particularly low for $\theta=45°$. Figures 10(a) and 10(b) illustrate the wave magnitude pattern within the junction for $\theta=30°$ and 45°, respectively, at $ka=9$. At higher frequencies and hence shorter wavelengths, the higher side-branch mode is well developed within the interfacing region between the side branch and the main duct at $\theta=30°$. At $\theta=45°$, the length of the interfacing region does not fit the proper development of the first asymmetric side-branch higher mode. However, one should note that such phenomenon depends on the length of the interfacing region relative to the sound wavelength and there can be a case where the higher mode is not so badly developed at $\theta>45°$.

Figure 11 illustrates the variation of the magnitude of the upstream propagating higher modes with the wave number. In general, the magnitudes of these higher modes are lower than those of the downstream going higher modes. For the asymmetric upstream going mode [Fig. 11(a)], its magnitude increases as the slant angle increases beyond 37.5°. This is rather expected as an acute θ does not favor upstream diffraction from the side-branch exit according to the well-known rectangular piston radiation characteristics.[23] The magnitude of this asymmetric mode does not depend much on the slant angle after the cut-on of the symmetrical higher mode. The magnitude of the symmetrical higher mode is also not a strong function of the slant angle. Its magnitude is also comparable to that of the asymmetric mode for $ka>2\pi$ [Fig. 11(b)].

The case of $b=a$ is a special case as the higher modes in the side branch and in the main duct will start to propagate at the same time. Figure 12 illustrates the magnitudes of the plane waves inside the system. It is again observed that there is a transition of propagation mode for the plane wave at $30° < \theta < 33°$. A more detailed investigation suggests that

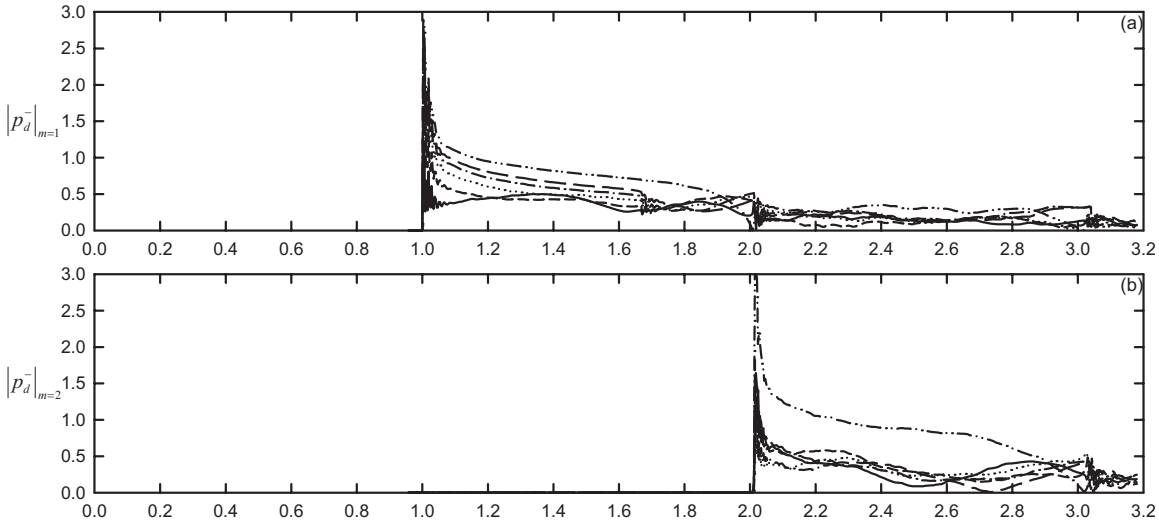


FIG. 11. Variations of upstream going higher mode magnitudes with frequency for $b/a=0.6$. (a) $m=1$ mode in main duct and (b) $m=2$ mode in main duct. Legends: Same as those in Fig. 4.

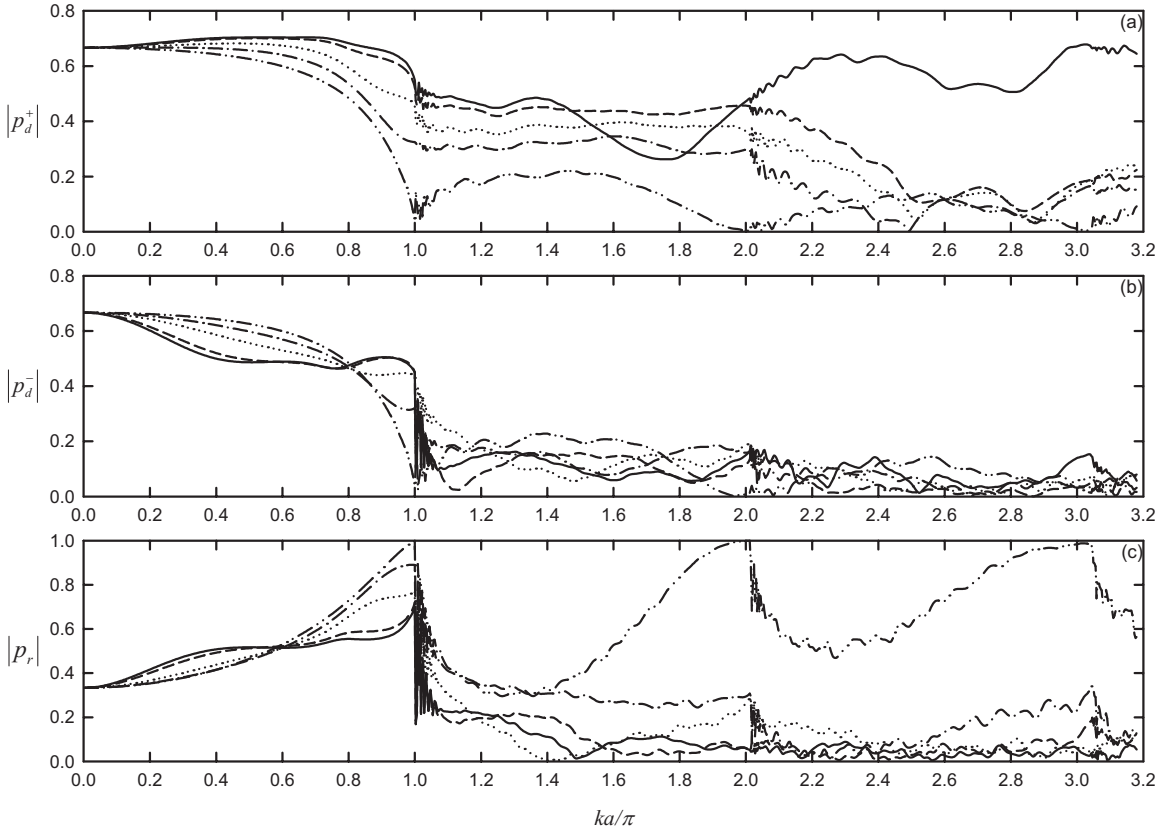


FIG. 12. Variations of plane wave magnitudes with frequency for $b/a=1$. (a) Downstream going wave in main duct; (b) upstream going wave in main duct; and (c) upstream going wave in the side branch. (—) $\theta=30°$; (---) $\theta=33°$; (···) $\theta=45°$; (–·–) $\theta=60°$; (–···–); $\theta=90°$.

the critical slant angle $\theta_c \sim 31.5°$ in this case and thus there is a slight reduction in the diffraction angle θ_d to 14.8°. For $b/a \to 0$, which is the case of a point source located on the duct wall, the diffraction angle should be large while θ_c should be small. According to Eq. (9) and the basic theory,[19,23] θ_d is capped below 90°. Figure 13 illustrates the variation of θ_d with b/a. One can observe that there is a small trough at $b/a=0.8$. The diffraction angle θ_d increases as b/a approaches 0, which is expected. However, the case for $b/a \to 0$ is not of practice interest and thus is not further discussed. It is also found that while the sudden drop of plane wave magnitude in the downstream duct due to the change in the propagation mode is significant for large b/a, those for small b/a are less pronounced (not shown here). In addition, one should note that Eq. (9) actually has two solutions for $0 < \theta_c < 90°$ when b/a is fixed. Those illustrated in Figs. 4 and 12 are the larger ones. Most of these smaller solutions are all less than 20°, which are basically of no practical significance. Nevertheless, they are presented in Fig. 13 for the sake of completeness.

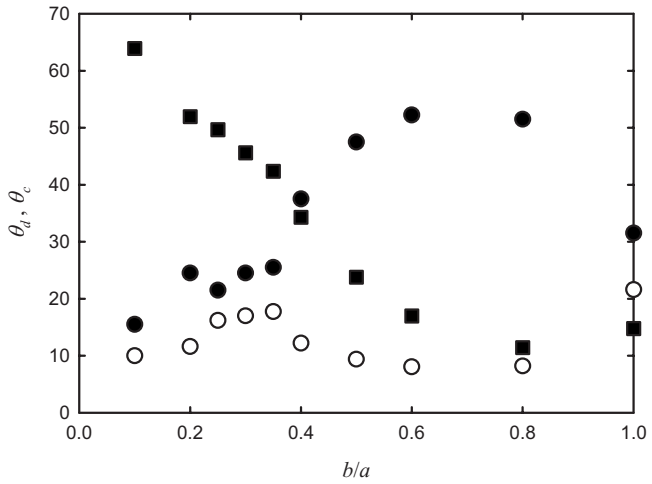


FIG. 13. Variation of θ_d and θ_c with b/a. (●) θ_c (larger solution); (○) θ_c (smaller solution); and (■) θ_d.

The behaviors of the higher mode propagation for $b/a=1$ case are very similar to those for $b/a=0.6$, except that the $m=2$ mode propagates along the slanted branch in the frequency range tested. Thus, these results are not discussed. Also, the above-mentioned wave interactions can be found at other values of b/a as far as $b < a$.

V. CONCLUSIONS

The sound propagation from a slanted rectangular side branch into an infinitely long rectangular duct of equal spanwise width and the corresponding interactions between the acoustic modes at the "side-branch-main duct" junction are investigated in the present study using the finite element method with absorptive edges at the computational domain boundaries. The relationship between the sound transmission effectiveness and the slant angle of the side branch relative to the main duct axis is examined in detail. Some theoretical deductions are also derived using the plane wave assumption. Owing to practice significance, side branches having width larger than that of the main duct are not considered.

Rectangular ducts and side branches are the most commonly found duct elements in modern buildings with centralized ventilation and air-conditioning systems.

The plane wave assumption is only valid when the forcing frequency is less than 20% of the first duct mode eigenfrequency. Before the cut-on of this duct mode frequency, the acoustical energy transmitted into the duct downstream of the junction increases with decreasing slant angle. However, the energy reflected back into the side branch decreases as the slant angle decreases. Within the frequency range studied, it is observed that the first symmetric duct mode possesses the strongest intensity, next is the first asymmetric duct mode, while the plane wave is the weakest when all these modes are coexisting.

It is also found that there exists a critical slant angle at which a drastic change in the propagation of sound will occur. A rapid drop of pressure transmission coefficient is observed when the slant angle increases beyond this critical slant angle. The angle varies with the width ratio between the side branch and the main duct. A hypothetical model incorporating a diffraction angle is set up to explain the phenomenon. It is ascribed to be due to the reflection of the diffracted sound ray at the upstream edge of the side-branch exit from the opposite duct wall, which causes significant reflection back into the side branch. The variation of this critical slant angle with the width ratio between the side branch and the main duct is determined. It is observed that this angle decreases as the width ratio decreases when the latter is less than 0.8. A slight increase of this angle is found otherwise. At or near to the critical slant angle, the magnitude and thus intensity of the first symmetric duct mode is found to be the highest.

ACKNOWLEDGMENTS

This work is supported by a grant from the Research Grant Council, the HKSAR Government (Project No. PolyU 5278/06E).

[1]L. L. Beranek, *Noise and Vibration Control Engineering, Principles and Applications* (Wiley, New York, 1992).
[2]J. W. Miles, "Diffraction of sound due to right angled joints in rectangular ducts," J. Acoust. Soc. Am. **19**, 572–579 (1947).
[3]J. C. Bruggeman, "The propagation of low-frequency sound in a two-dimensional duct system with T joints and right angle bends: Theory and experiment," J. Acoust. Soc. Am. **82**, 1045–1051 (1987).
[4]T. L. Redmore and K. A. Mulholland, "The application of mode coupling theory to the transmission of sound in the sidebranch of a rectangular duct system," J. Sound Vib. **85**, 323–331 (1982).
[5]V. Dubos, J. Kergomard, A. Khettabi, J.-P. Dalmont, D. H. Keefe, and C. J. Nederveen, "Theory of sound propagation in a duct with a branched tube using modal decomposition," Acust. Acta Acust. **85**, 153–169 (1999).
[6]S. K. Tang, "Sound transmission characteristics of Tee-junctions and the associated length corrections," J. Acoust. Soc. Am. **115**, 218–227 (2004).
[7]S. K. Tang and F. Y. C. Li, "On low frequency sound transmission loss of double side-branches: A comparison between theory and experiment," J. Acoust. Soc. Am. **113**, 3215–3225 (2003).
[8]H. G. Davies, "Noise propagation in corridors," J. Acoust. Soc. Am. **53**, 1253–1262 (1973).
[9]W. K. R. Lippert, "The measurement of sound reflection and transmission at right-angled bends in rectangular tubes," Acustica **4**, 313–319 (1954).
[10]M. Razavy, "An acoustic waveguide with variable cross section," J. Acoust. Soc. Am. **86**, 1155–1160 (1989).
[11]S. K. Lee, B. R. Mace, and M. J. Brennan, "Wave propagation, reflection and transmission in non-uniform one-dimensional waveguides," J. Sound Vib. **304**, 31–49 (2007).
[12]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[13]S. Dequand, S. Hulshoff, H. van Kuijk, J. Willems, and A. Hirschberg, "Helmholtz-like resonator self-sustained oscillations. 2. Detailed flow measurements and numerical simulations," AIAA J. **41**, 416–423 (2003).
[14]S. K. Tang and C. K. Lau, "Sound transmission across a smooth nonuniform section in an infinitely long duct," J. Acoust. Soc. Am. **112**, 2602–2611 (2002).
[15]C. K. Lau and S. K. Tang, "Sound transmission across duct constrictions with and without tapered sections," J. Acoust. Soc. Am. **117**, 3679–3685 (2005).
[16]L. Langemyr, A. Nordmark, M. Ringh, A. Ruhe, J. Oppelstrup, and M. Dorobantu, *Partial Differential Equation Toolbox User Guide* (The MathWorks Inc., Natick, MA, 1996).
[17]L. P. George, *Automatic Mesh Generation—Application to Finite Element Methods* (Wiley, New York, 1991).
[18]N. Peake, "On the radiation properties of an asymmetric cylinder," Wave Motion **22**, 371–385 (1995).
[19]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).
[20]L. Huang, "A theory of reactive control of low-frequency duct noise," J. Sound Vib. **238**, 575–594 (2000).
[21]M. Loéve, *Probability Theory* (Springer, New York, 1978).
[22]M. C. J. Trinder and P. A. Nelson, "Active noise control in finite length duct," J. Sound Vib. **89**, 95–105 (1983).
[23]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

Modeling sound propagation in acoustic waveguides using a hybrid numerical method

Ray Kirby[a)]
School of Engineering and Design, Mechanical Engineering, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom

(Received 3 September 2007; revised 2 May 2008; accepted 4 July 2008)

Sound propagation in an acoustic waveguide is examined using a hybrid numerical technique. Here, the waveguide is assumed to be infinite in length with an arbitrary but uniform cross section. Placed centrally within the guide is a short component section with an irregular nonuniform shape. The hybrid method utilizes a wave based modal solution for a uniform section of the guide and, using either a mode matching or point collocation approach, matches this to a standard finite element based solution for the component section. Thus, one needs only to generate a transverse finite element mesh in uniform sections of the waveguide and this significantly reduces the number of degrees of freedom required. Moreover, utilizing a wave based solution removes the need to numerically enforce a nonreflecting boundary condition at infinity using a necessarily finite mesh, which is often encountered in studies that use only the standard finite element method. Accordingly, the component transmission loss may readily be computed and predictions are presented here for three examples: an expansion chamber, a converging-diverging duct, and a circular cylinder. Good agreement with analytic models is observed, and transmission loss predictions are also presented for multimode incident and transmitted sound fields. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2967837]

PACS number(s): 43.20.Mv, 43.20.Bi, 43.50.Gf [SFW] Pages: 1930–1940

I. INTRODUCTION

The study of nonuniform obstructions in an otherwise uniform waveguide is of fundamental interest and it is not surprising that this subject has received extensive attention in literature. Applications are many, including underwater sound propagation, the transmission of elastic and electromagnetic waves, and sound propagation in ducts or pipes. It is common in these applications for relatively long uniform sections to be present, which are punctuated by relatively short area changes and/or nonuniform obstacles. This article is concerned with the application of a general numerical method suitable for examining sound propagation in uniform acoustic waveguides of arbitrary cross section that contain one or more nonuniform obstacles, or area changes, placed centrally within the guide.

The method presented here will focus on sound propagation in ductwork and so is applicable to rectangular ventilation ducts and also to circular and oval ductwork. The key challenge here is to model accurately the scattering of sound waves from nonuniform area discontinuities in a relatively large ductwork. Of course, the study of sound scattering from area discontinuities goes back to Rayleigh, although traditionally it has been possible only to examine relatively simple nonuniform geometries with modest dimensions. More recent examples include Miles[1] and Selamet and Easwaran,[2] who examined plane wave propagation in variable area ducts, and Boström[3] who used analytic techniques to study scattering by spherical and spheroidal obstacles in a duct. Boström examined ducts of arbitrary cross section but noted that tractable solutions were possible only for a limited range of geometries. The study of sound scattering from obstacles has also received extensive attention in the location of so-called trapped modes, see, for example, Refs. 4–7. Trapped modes are acoustic resonances near obstacles in ducts and have been found to exist for a number of geometries including parallel plates, rectangular obstructions, cylinders, and ball-type valves. The study of trapped modes is, however, largely based on analytic work and, while some relatively complicated obstructions have been examined, these techniques are not readily applicable to more complex nonuniform area changes and/or ducts of arbitrary cross section.

The study of sound propagation over a fully arbitrary area change will inevitably require numerical methods that can cope with irregular geometries and/or step changes in boundary conditions. Suitable numerical methods include the standard finite element method (FEM) and the boundary element method (BEM). For example, Tang and Lau,[8] and later Lau and Tang,[9] used the standard FEM to study tapered and convergent-divergent sections in rectangular ducts, although both studies required a large number of degrees of freedom in order to obtain a converged solution. Jeong *et al.*[10] reviewed the application of the BEM in rectangular ductwork and demonstrated that by discretising the duct into multiple domains one may generate an efficient BEM algorithm that may be applied to larger ductwork. However, area changes were not examined by Jeong *et al.*[10] and it is evident that, even for relatively low frequencies, a large number of degrees of freedom were still required. In principle, numerical methods may be applied to a wide range of sound propagation problems; however, a problem that plagues numerical

[a)]Electronic mail: ray.kirby@brunel.ac.uk

 0001-4966/2008/124(4)/1930/11/$23.00

methods is the computational time required to obtain solutions, especially if one wishes to cover a representative frequency range for large ducts such as those commonly found in ventilation systems. Here, the FEM normally finds favor over the BEM because discretization delivers a banded symmetric matrix,[10,11] although even the FEM requires a large number of degrees of freedom to study relatively simple problems. For example, Lau and Tang[9] used a very fine mesh to study a converging-diverging duct, which has the potential to severely limit the upper frequency of the analysis as the size of the problem becomes unmanageable. A further and arguably more fundamental problem with numerical methods is the specification of the terminating (downstream) axial boundary condition. When studying infinite waveguides, this boundary condition should be anechoic, but problems arise when attempting to specify a boundary condition at infinity using a necessarily finite grid. Of course, for plane wave propagation one may easily represent an anechoic termination by setting the impedance of the downstream boundary to be equal to the characteristic impedance of the fluid in the duct, and this has been applied by many authors (see Refs. 9–12). However, for larger ventilation ducts, higher order modes may propagate at relatively low frequencies and this plane wave boundary condition is no longer appropriate. To overcome this Lau and Tang[9] specified absorptive walls in the downstream section of their duct so that higher order modes are numerically damped down before reaching the (plane wave) terminating condition. Other methods may also be used, for example, a "perfectly matched layer" or a high order local nonreflecting boundary condition, see Givoli.[13] These various nonreflecting boundary conditions have been shown to work well under certain conditions; however, these methods can be computationally expensive and may only ever provide an approximation to the desired terminating boundary condition.

An alternative numerically efficient approach is to retain a modal analysis for the uniform section and to match, or join, this analysis to a standard finite element representation for the component section. In this way, complex nonuniform sections may be studied as efficiently as possible, and so it is not surprising that this type of "hybrid" approach has been applied to a wide range of problems in literature. For example, in the study of elastic waveguides, Mal and Chang[14] used a hybrid method to enforce continuity conditions over discrete nodal locations that are common to the boundary between the uniform and nonuniform regions. Liu *et al.*[15] also studied elastic waveguides but matched a finite element representation of an interior region to a Green's function integral representation of the exterior region. Other examples include Refs. 16–18, whereby the continuity conditions over the interface between two regions in electromagnetic waveguide are enforced using mode matching (MM). In acoustics, the application of the method is arguably less widespread and examples tend to favor exterior problems. For example, Kagawa *et al.*[19] used the FEM to analyze an interior problem and then applied MM to "join" this solution to a Green's function representation of the exterior acoustic far field. Astley and Cummings[20] also used MM but they analyzed sound radiation from a vibrating ventilation duct wall, using finite elements to discretize the exterior acoustic near field before coupling this to an eigenexpansion of the acoustic far field. A similar approach was also adopted by Imhof,[21] who studied both acoustic and elastic waveguides. A review of the application of a hybrid method to acoustic problems is discussed by Astley,[22] and here a close relationship between the finite element Dirichlet-to-Neumann (FE-DtN) and hybrid methods is demonstrated, provided one uses MM to enforce continuity conditions between the uniform and nonuniform sections. Furthermore, Astley[22] applied this hybrid method to the study of interior as well as exterior acoustic problems, and for the interior problem sound propagation is modeled in a simple diverging duct, although results were reported only for the exterior problem. Results for a similar interior problem were, however, reported in an earlier paper by Astley and Eversman,[23] who analyzed sound attenuation in a cosine tapered converging duct with mean flow and a locally reacting liner.

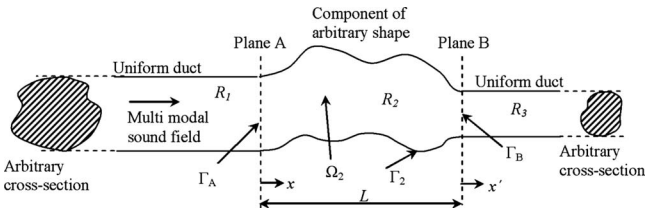


FIG. 1. Geometry of the duct.

Clearly, a hybrid numerical method is well suited in analyzing sound propagation in interior problems such as ventilation ducts. Accordingly, this article will focus on applying the method of Astley[22] to the analysis of uniform ducts of arbitrary cross section that include relatively complex component sections. Of interest here will be the relative convergence and accuracy of the method when compared to analytic results. This paper will also compare two different methods of enforcing the continuity conditions between the uniform and nonuniform sections, namely, the MM method of Astley[22] and a point collocation (PC) approach similar to the one described by Kirby and Lawrie.[24] Accordingly, this article begins by outlining the governing equations for the problem before describing the MM and PC methods; results are then presented for sound propagation in a duct containing an expansion chamber, a converging-diverging duct, and a cylinder.

II. THEORY

The duct geometry to be analyzed consists of three separate regions, as shown in Fig. 1. The inlet duct is denoted region R_1, and this is assumed to have a uniform but arbitrary cross section. The inlet duct abuts onto a general arbitrarily shaped "component" section, which is denoted region R_2. Finally, the component section abuts onto the outlet duct, region R_3, which is assumed to have a uniform but arbitrary cross section. In principle each region may be bounded by acoustically soft walls and/or wave-bearing walls; moreover, each section may contain more than one (equivalent) fluid, for example, a porous material. Separating the three regions are two planes, A and B, which lie perpendicular to regions

R_1 and R_3, respectively. Here, the convective effects of mean flow may readily be added to the analysis, although the component section is likely to generate free shear layers downstream of an area discontinuity. Hydrodynamic modes may then be present and it is likely that this will introduce further numerical complications when matching over plane B downstream of an area discontinuity. Furthermore, accommodating nonuniform geometries in the component section will also require a detailed knowledge of the flow patterns. As these issues are likely to complicate the analysis and detract from the main focus of this paper, mean flow is omitted from the analysis that follows.

Sound propagation in region q ($q=1$, 2, or 3) of the duct shown in Fig. 1 is governed by the acoustic wave equation

$$\frac{1}{c_q^2}\frac{\partial^2 p_q'}{\partial t^2} - \nabla^2 p_q' = 0, \tag{1}$$

where c_q is the speed of sound, p_q' is the acoustic pressure, and t is time. The solution of this equation proceeds by using a method that is the most appropriate for the geometry in each region. Accordingly, the uniform geometry of regions R_1 and R_3 encourages a modal analysis and so the acoustic pressure is expanded as an infinite sum over the eigenmodes in each region. For the duct component, provision is made for a complicated nonuniform shape. Here, the wave equation is solved using the FEM.

For regions R_1 and R_3, expansion of the sound pressure field, assuming a time dependence of $e^{i\omega t}$ (where $i=\sqrt{-1}$ and ω is the radian frequency), yields

$$p_1'(x,y,z) = \sum_{n=0}^{\infty} F_n \Phi_n(y,z) e^{-ik_1 \lambda_n x} + \sum_{n=0}^{\infty} A_n \Phi_n(y,z) e^{ik_1 \lambda_n x} \tag{2}$$

and

$$p_3'(x,y,z) = \sum_{j=0}^{\infty} B_j \Psi_j(y,z) e^{-ik_3 \gamma_j x'} + \sum_{j=0}^{\infty} C_j \Psi_j(y,z) e^{ik_3 \gamma_j x'}. \tag{3}$$

Here, F_n, A_n, B_j, and C_j are the modal amplitudes, $k_q = \omega/c_q$, λ is the (dimensionless) wavenumber in region R_1, and γ is the (dimensionless) wavenumber in region R_3. The quantities $\Phi_n(y,z)$ and $\Psi_j(y,z)$ are the transverse duct eigenfunctions in regions R_1 and R_3, respectively. In the analysis that follows it is assumed that regions R_1 and R_3 have an arbitrary cross section and so the eigenvalues and eigenvectors in regions R_1 and R_3 are found using the FEM.[25] The incident sound pressure field in region R_1 is assumed to be multimodal, which requires knowledge of the modal amplitudes in the incident sound field. Here, equal modal energy density (EMED) is assumed for all propagating modes, as this is thought to best represent the sound field emanating from a fan in a ventilation system (see Kirby and Lawrie[24]). This choice may, however, readily be changed to include other relationships between the modal amplitude such as those discussed by Kirby and Lawrie or by utilizing Green's functions to replicate a point source. For EMED, this yields

$$\left|\frac{F_n}{p_0}\right|^2 = \frac{I_0}{I_n \Sigma_{m=0}^{n_F} \lambda_m}, \tag{4}$$

where p_0 is a reference pressure chosen here, arbitrarily, to be equal to unity. The number of modes propagating in region R_1 is denoted by n_F (for modes that are "cutoff," $F=0$) and $I_n = \int_{\Gamma_A} |\Phi_n(y,z)|^2 dy dz$, where Γ_A denotes the surface of R_2 that lies on plane A.

For region R_2 the acoustic pressure is approximated by

$$p_2'(x,y,z) = \sum_{j=1}^{n_2} N_j(x,y,z) p_{2_j}, \tag{5}$$

where N_j is a global trial (or shape) function for the finite element mesh, p_{2_j} is the value of the acoustic pressure at node j, and n_2 is the number of nodes (or degrees of freedom) in region R_2. Expressing Eq. (5) in vector form yields

$$p_2'(x,y,z) = [N_1(x,y,z), N_2(x,y,z), \ldots, N_{n_2}(x,y,z)] \begin{bmatrix} p_{2_1} \\ p_{2_2} \\ p_{2_{n_2}} \end{bmatrix} = \mathbf{N}\mathbf{p}_2. \tag{6}$$

A weighted residual statement of the wave equation may now be formulated and, after application of Green's theorem, this yields

$$\left[\int_{\Omega_2} [\nabla \mathbf{N}^T \cdot \nabla \mathbf{N} - k_2^2 \mathbf{N}^T \mathbf{N}] d\Omega_2\right] \{\mathbf{p}_2\} = \int_{\Gamma_2} [\mathbf{N}^T \nabla p_2' \cdot \mathbf{n}_2] d\Gamma_2. \tag{7}$$

Here, Ω_2 and Γ_2 denote the volume and the outer surface of region R_2, respectively, and $\mathbf{n}_2$ is the outward unit normal vector for the surface of region 2. In order to simplify the presentation of the method, Eq. (7) assumes that a single homogenous fluid is present in region R_2; however, more than one fluid may readily be added simply by changing the wavenumber and writing a separate equation for each fluid. It is convenient here to separate planes A and B from the surface integral on the right hand side of Eq. (7) and to write

$$\int_{\Gamma_2} [\mathbf{N}^T \nabla p_2' \cdot \mathbf{n}_2] d\Gamma_2 = \int_{\Gamma_A} [\mathbf{N}^T \nabla p_2' \cdot \mathbf{n}_A] d\Gamma_A + \int_{\Gamma_B} [\mathbf{N}^T \nabla p_2' \cdot \mathbf{n}_B] d\Gamma_B + \int_{\Gamma_e} [\mathbf{N}^T \nabla p_2' \cdot \mathbf{n}_e] d\Gamma_e, \tag{8}$$

where Γ_A and Γ_B denote the surface of R_2 that lies on planes A and B, respectively, and Γ_e is the surface of region R_2 that does not lie on Γ_A and Γ_B. Here, $\mathbf{n}_q$ is the outward unit normal vector over surface q.

The analysis proceeds by enforcing continuity of acoustic pressure and axial velocity over surfaces Γ_A and Γ_B. Two different approaches to enforce these continuity conditions are examined here: the MM method of Astley[22] and the PC

method of Kirby and Lawrie.[24] For the MM method the number of modes utilized in regions R_1 and R_3 does not necessarily have to equal the number of nodes on Γ_A and Γ_B, whereas the number of collocation points should equal the number of nodes on Γ_A and Γ_B. Therefore, one would expect MM to be the more efficient of the two methods, although for interior problems it is likely that the number of nodes on Γ_A and Γ_B will be significantly smaller than the total number of nodes used to mesh region R_2. Hence, for interior problems the potential reduction in problem size afforded by the MM method is likely to be limited when compared to the PC method. However, Astley[22] noted that, provided one uses an appropriate weighting function, the MM method is capable of delivering a symmetric stiffness matrix, which is not the case for PC. Thus, the MM method will be faster than the PC method; however, the PC method is retained here in order to provide an alternative method that may prove useful, at least as a benchmark for the MM method.

The application of the continuity conditions and the solution of the problem will be covered in the next two sections. On solving the problem, the sound transmission loss (TL) of the component section is readily calculated from the ratio of the transmitted to incident sound powers, which yields

$$\mathrm{TL} = -10\log_{10}\frac{\rho_1 k_3}{\rho_3 k_1}\sum_{m=0}^{n_B}\frac{\gamma_m H_m |B_m|^2}{I_0}, \tag{9}$$

where $H_m = \int_{\Gamma_B} |\Psi_m(y,z)|^2 dydz$, n_B is the number of modes cut on in region R_3, and ρ_q is the density of the fluid in region q.

A. Mode matching

The MM method relies on weighting the appropriate continuity conditions and then integrating over Γ_A and Γ_B. Accordingly, making use of Eqs. (2) and (3) allows continuity of pressure and velocity over Γ_A to be written as

$$p_2'(0,y,z) = \sum_{n=0}^{\infty} F_n \Phi_n(y,z) + \sum_{n=0}^{\infty} A_n \Phi_n(y,z) \tag{10}$$

and

$$\frac{\partial p_2'}{\partial x}(0,y,z) = -ik_1\beta_1\left\{\sum_{n=0}^{\infty} F_n \lambda_n \Phi_n(y,z) - \sum_{n=0}^{\infty} A_n \lambda_n \Phi_n(y,z)\right\}. \tag{11}$$

Similarly, over Γ_B

$$p_2'(L,y,z) = \sum_{j=0}^{\infty} B_j \Psi_j(y,z) \tag{12}$$

and

$$\frac{\partial p_2'}{\partial x}(L,y,z) = -ik_3\beta_3\left\{\sum_{j=0}^{\infty} B_j \gamma_j \Psi_j(y,z) - \sum_{j=0}^{\infty} C_j \gamma_j \Psi_j(y,z)\right\}. \tag{13}$$

Here, L is the axial distance between Γ_A and Γ_B and $\beta_q = \rho_2/\rho_q$. To obtain a solution, it is necessary to ascribe boundary conditions to the outer surface of the problem domain. Here, complex boundary conditions such as acoustically soft or wave-bearing walls may be applied; however, in order to validate the method against analytic results the analysis that follows will assume that all outer duct surfaces are acoustically hard. Thus, for Γ_e, $\nabla p_2' \cdot \mathbf{n}_e = 0$, and for regions R_1 and R_3 the problem reduces to computing the eigenvalues and eigenvectors in a hard walled duct of arbitrary cross section. In addition, a nonreflecting boundary condition is specified in region R_3 by setting C_j equal to zero. The velocity matching conditions, given by Eqs. (11) and (13), may now substituted into Eq. (8), which yields (after dropping the summations for clarity)

$$\int_{\Gamma_2} [\mathbf{N}^T \nabla p_2' \cdot \mathbf{n}_2] d\Gamma_2 = ik_1\beta_1\lambda_n \int_{\Gamma_A} \mathbf{N}^T [F_n\Phi_n - A_n\Phi_n] d\Gamma_A - ik_3\beta_3\gamma_j \int_{\Gamma_B} \mathbf{N}^T [B_j\Psi_j] d\Gamma_B. \tag{14}$$

The MM method of Astley[22] proceeds by weighting each pressure condition using the incident velocity in the axial direction in region R_1 (or region R_3) and then integrating over Γ_A (or Γ_B). Thus, for Γ_A Eq. (10) yields

$$ik_1\beta_1\left\{\lambda_m \int_{\Gamma_A} \Phi_m \Phi_n d\Gamma_A\right\}\{F_n\} + ik_1\beta_1\left\{\lambda_m \int_{\Gamma_A} \Phi_m \Phi_n d\Gamma_A\right\}\{A_n\} = ik_1\beta_1\left\{\lambda_m \int_{\Gamma_A} \Phi_m \mathbf{N} d\Gamma_A\right\}\{\mathbf{p}_{2A}\}, \tag{15}$$

and for Γ_B, Eq. (12) yields

$$ik_3\beta_3\left\{\gamma_m \int_{\Gamma_B} \Psi_m \Psi_n d\Gamma_B\right\}\{B_n\} = ik_3\beta_3\left\{\gamma_m \int_{\Gamma_B} \Psi_m \mathbf{N} d\Gamma_B\right\}\{\mathbf{p}_{2B}\}. \tag{16}$$

Here, the summation signs have again been dropped for clarity, and on use of Eq. (6), vectors $\mathbf{p}_{2A}$ and $\mathbf{p}_{2B}$ hold values of the finite element solution in region R_2 at the nodal locations on the surfaces Γ_A and Γ_B, respectively. Finally, before solving it is necessary first to truncate the infinite sums, and here m_1 and m_3 will denote the number of modes assumed to be present in regions R_1 and R_3, respectively. It is then convenient to rewrite the problem in matrix form, where

$$[\mathbf{G}] = \int_{\Omega_2} [\nabla \mathbf{N}^T \cdot \nabla \mathbf{N} - k_2^2 \mathbf{N}^T \mathbf{N}] d\Omega_2, \tag{17}$$

$$[\mathbf{Q}] = ik_1\beta_1 \sum_{m=0}^{m_1} \lambda_m \int_{\Gamma_A} \Phi_m \mathbf{N} d\Gamma_A, \tag{18}$$

$$[\mathbf{R}] = ik_3\beta_3 \sum_{m=0}^{m_3} \gamma_m \int_{\Gamma_B} \Psi_m \mathbf{N} d\Gamma_B, \tag{19}$$

$$[\mathbf{M}_1] = ik_1\beta_1 \sum_{m=0}^{m_1} \sum_{n=0}^{m_1} \lambda_m \int_{\Gamma_A} \Phi_m \Phi_n d\Gamma_A, \tag{20}$$

$$[\mathbf{M}_2] = ik_3\beta_3 \sum_{m=0}^{m_3} \sum_{n=0}^{m_3} \gamma_m \int_{\Gamma_B} \Psi_m \Psi_n d\Gamma_B. \tag{21}$$

In addition,

$$[\tilde{\mathbf{Q}}] = ik_1\beta_1 \sum_{m=0}^{m_1} \sum_{n=0}^{n_F} F_n \lambda_m \int_{\Gamma_A} \Phi_m N d\Gamma_A \tag{22}$$

and

$$[\tilde{\mathbf{M}}_1] = ik_1\beta_1 \sum_{m=0}^{m_1} \sum_{n=0}^{n_F} F_n \lambda_m \int_{\Gamma_A} \Phi_m \Phi_n d\Gamma_A. \tag{23}$$

Equations (7), (15), and (16) may now be written in matrix form:

$$[\mathbf{G}]\{\mathbf{p}_2\} + [\mathbf{Q}]^T\{\mathbf{A}\} + [\mathbf{R}]^T\{\mathbf{B}\} = [\tilde{\mathbf{Q}}]^T, \tag{24}$$

$$-[\mathbf{M}_1]\{\mathbf{A}\} + [\mathbf{Q}]\{\mathbf{p}_{2A}\} = [\tilde{\mathbf{M}}_1], \tag{25}$$

$$-[\mathbf{M}_2]\{\mathbf{B}\} + [\mathbf{R}]\{\mathbf{p}_{2B}\} = \mathbf{0}. \tag{26}$$

To combine all three equations, matrix $\mathbf{G}$ is decomposed into separate elements to give

$$[\mathbf{G}]\{\mathbf{p}_2\} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{1e} & \mathbf{G}_{13} \\ \mathbf{G}_{e1} & \mathbf{G}_{ee} & \mathbf{G}_{e3} \\ \mathbf{G}_{31} & \mathbf{G}_{3e} & \mathbf{G}_{33} \end{bmatrix} \begin{Bmatrix} \mathbf{p}_{2A} \\ \mathbf{p}_{2e} \\ \mathbf{p}_{2B} \end{Bmatrix}, \tag{27}$$

where matrix $\mathbf{G}_{mn}$ has order $n_m \times n_n$. Here, n_1 and n_3 denote the number of nodes on Γ_A and Γ_B, respectively (where $m_1 \leqslant n_1$ and $m_3 \leqslant n_3$); n_2 is the number of nodes in region R_2; and n_e is the number of nodes that lie in region 2 but do not lie on Γ_A and Γ_B (so that $n_e = n_2 - n_1 - n_3$). The values for pressure at those nodes in region R_2 that do not lie on Γ_A and Γ_B are held in matrix $\mathbf{p}_{2e}$. The problem may now be written as

$$\begin{bmatrix} -\mathbf{M}_1 & \mathbf{Q} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}^T & \mathbf{G}_{11} & \mathbf{G}_{1e} & \mathbf{G}_{13} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{e1} & \mathbf{G}_{ee} & \mathbf{G}_{e3} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{31} & \mathbf{G}_{3e} & \mathbf{G}_{33} & \mathbf{R}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} & -\mathbf{M}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{p}_{2A} \\ \mathbf{p}_{2e} \\ \mathbf{p}_{2B} \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{M}}_1 \\ \tilde{\mathbf{Q}}^T \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{28}$$

Equation (28) consists of $n_T (= m_1 + n_2 + m_3)$ simultaneous equations, which are solved to give the unknown modal amplitudes and pressures. The values for $\mathbf{B}$ may then be substituted into Eq. (9) in order to calculate the TL of the component section. Note that, provided the nodes in the finite element mesh are numbered correctly, Eq. (28) delivers a banded matrix that is also symmetrical (a result of the weighting function chosen when enforcing the pressure conditions). A further advantage of this MM approach is that one may choose $m_1 \ll n_1$ (and/or $m_3 \ll n_3$) so that the size of the problem is reduced, although for interior problems it is unlikely that this will significantly reduce the value for n_T, since normally $n_2 \gg n_1, n_3$.

B. Point collocation

In this section, matching conditions are enforced using the PC method described by Kirby and Lawrie.[24] This involves matching over discrete points on Γ_A and Γ_B. Here, the points chosen in adjacent regions must be identical, and so the number and location of the nodes generated in the eigensolution for the uniform duct sections must coincide with the location and number of the nodes generated in the finite element discretization process for Γ_A and Γ_B. This also requires that the number of modes used in the expansion of the sound pressure fields in regions R_1 and R_3 must also equal the number of collocation points on Γ_A and Γ_B, respectively. Moreover, satisfying the velocity matching conditions at discrete points, rather than in the integral sense, means that the surface integrals in Eq. (8) are not carried out in the normal way; instead they are removed from Eq. (8) and enforced separately. Thus, after applying the boundary conditions listed in the MM approach, the velocity matching conditions may be written as

$$\sum_{n=0}^{n_1} A_n \lambda_n \mathbf{\Phi}_n + \frac{i}{k_1\beta_1} \frac{\partial \mathbf{p}_{2A}}{\partial x} = \sum_{n=0}^{n_F} F_n \lambda_n \mathbf{\Phi}_n \tag{29}$$

and

$$\sum_{j=0}^{n_3} B_j \gamma_j \mathbf{\Psi}_j - \frac{i}{k_3\beta_3} \frac{\partial \mathbf{p}_{2B}}{\partial x} = \mathbf{0}. \tag{30}$$

Here, vectors $\mathbf{\Phi}$ and $\mathbf{\Psi}$ hold values of the duct eigenfunctions at nodal locations equivalent to those chosen for the finite element mesh in region R_2, but in regions R_1 and R_3, respectively. Similarly, the pressure matching conditions give

$$-\sum_{n=0}^{n_1} A_n \mathbf{\Phi}_n + \mathbf{p}_{2A} = \sum_{n=0}^{n_F} F_n \mathbf{\Phi}_n \tag{31}$$

and

$$-\sum_{j=0}^{n_3} B_j \Psi_j + \mathbf{p}_{2B} = \mathbf{0}. \tag{32}$$

A requirement of the collocation approach is that all of the eigensolutions obtained for regions R_1 and R_3 are used when enforcing Eqs. (29)–(32). These eigensolutions deliver a set of eigenmodes for which only about 20% are accurate; however, this does not cause any difficulties provided one uses a sufficient number of accurate eigenmodes to obtain a converged solution for the sound pressure field at each node. This issue is discussed in detail by Kirby and Lawrie[24] but will also be reviewed here in the following section, in which PC predictions will be compared with MM predictions (that may use only "accurate" eigensolutions). The problem may finally be written as

$$\begin{bmatrix} \lambda_n \Phi_n & \frac{i}{k_1\beta_1}\frac{\partial \mathbf{N}}{\partial x} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\Phi_n & \mathbf{G}_{11} & \mathbf{G}_{1e} & \mathbf{G}_{13} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{e1} & \mathbf{G}_{ee} & \mathbf{G}_{e3} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{31} & \mathbf{G}_{3e} & \mathbf{G}_{33} & -\Psi_j \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\frac{i}{k_3\beta_3}\frac{\partial \mathbf{N}}{\partial x} & \gamma\Psi_j \end{bmatrix} \begin{bmatrix} A_n \\ \mathbf{p}_{2A} \\ \mathbf{p}_{2e} \\ \mathbf{p}_{2B} \\ B_j \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{n=0}^{n_F} F_n \lambda_n \Phi_n \\ \sum_{n=0}^{n_F} F_n \Phi_n \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{33}$$

Equation (33) consists of $n_T(=n_1+n_2+n_3)$ simultaneous equations, which are solved to give the unknown modal amplitudes and pressures. Note that Eq. (33) delivers a banded matrix, but this matrix is not symmetric. Note also that, when solving Eq. (33), it is necessary to multiply each of the terms that appear in Eqs. (31) and (32) by a scaling factor in order to enforce these equations in the final matrix (in the examples that follow 1×10^{10} is used).

III. RESULTS AND DISCUSSION

The methodology outlined in the previous section is validated here by comparison, where possible, with analytic predictions. Accordingly, this section begins by comparing predictions against analytic solutions for a simple expansion chamber and then proceeds to examine a converging-diverging duct. Finally, predictions are presented for a cylinder with a circular cross section. In each case, a regular (circular or rectangular) geometry is chosen for regions R_1 and R_3, and a symmetric component section is also chosen. This allows for a reduction in the problem to two dimensions, which facilitates comparison with analytic theory and also results presented in literature. In addition, where air is present, the speed of sound is taken to be $c_0=343.2244$ m/s, $k_1=k_3=\omega/c_0$, and $\beta_1=\beta_3=1$. For the finite element discretization, a mesh consisting of either eight noded quadrilateral or six noded triangular isoparametric elements is used. Furthermore, it is normally accepted that at least seven to ten nodes per wavelength are required in order to achieve reasonable accuracy when using the FEM. Accordingly, for all the results that follow, at least seven nodes per wavelength have been used in the axial and radial directions.

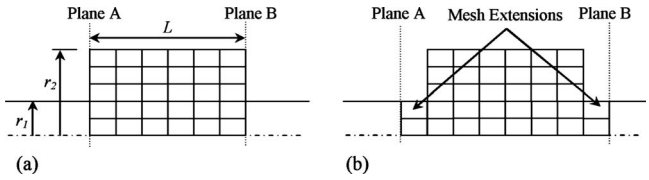


FIG. 2. (a) Mesh for expansion chamber and (b) extended mesh for expansion chamber.

A. Expansion chamber

The acoustic performance of an empty expansion chamber is well understood, and so provides a convenient example with which to begin validation of both methods. It is assumed here that the expansion chamber contains only air and has a circular cross section of radius r_2; the inlet/outlet ducts are also assumed to be circular, both with radius r_1. After drawing a line of symmetry, the problem may be represented, as shown in Figs. 2(a) and 2(b). Here, two separate figures have been drawn in order to show that the position of planes A and B is arbitrary, provided that they pass through the inlet and outlet pipes. Clearly, one may extend region R_2 well into the inlet and outlet ducts, although this will be computationally more expensive since extra elements are required. In view of the arbitrary position of planes A and B, predictions are presented here for different locations and the convergence of each model is investigated in order first to arrive at an optimum approach. Accordingly, a representative expansion chamber is chosen here, with $r_1=37$ cm, $r_2=76.2$ cm, and $L=315$ cm. In Fig. 3, TL predictions are presented for the MM (with $m_1=n_1$ and $m_3=n_3$) and PC methods, with and without mesh extensions. Here, the mesh ex-

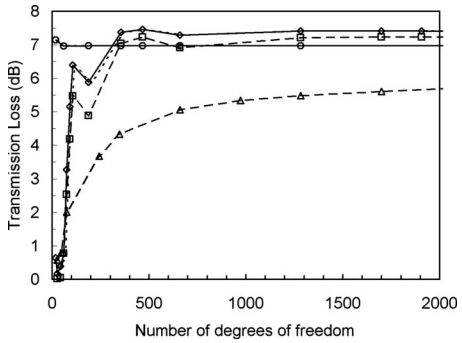


FIG. 3. Convergence of TL for expansion chamber: —○—○—, MM, 300 Hz; —◇—◇—, MM, 3240 Hz; —△—△—, PC no mesh extension, 300 Hz; —□—□—, PC no mesh extension, 3240 Hz; and ------, PC mesh extensions (overlays MM for 300 Hz and at higher values of n_T for 3240 Hz).

tensions in the inlet and outlet ducts are assumed to be one element deep (in the x direction) but may contain more than one element in the radial direction, see Fig. 2(b). Note that, for a simple area discontinuity, a modal expansion of the sound pressure field away from the discontinuity must deliver the same solution as that obtained using additional finite elements. Accordingly, there is nothing to be gained from adding additional elements to the mesh extensions, at least for the simple area discontinuities such as those found in this example.

In Fig. 3, TL predictions using MM and PC are shown at frequencies of 300 and 3240 Hz. Here, problems with convergence are clearly evident in the PC approach, but only when no mesh extensions are in place (note that the MM solutions with and without extensions are identical to at least 10 decimal places). The discrepancy between the two methods is likely to be caused by difficulties in enforcing the matching conditions at individual nodes over Γ_A and Γ_B when no extensions are present in the collocation method. At this corner, acoustic scattering will be pronounced and it is likely that a relatively large number of evanescent modes will be required when using a Fourier series to represent the pressure and velocity fields in R_1 and R_3. This is likely to cause relatively slow rates of convergence, as seen for the PC method in Fig. 3 when no mesh extensions are present. This trend has also been observed at other frequencies and for other chamber geometries (not shown here). In contrast, a more convergent system of equations is apparent when using mesh extensions since those evanescent modes at the silencer edge do not influence the sound pressure field away from the edge. It is noticeable, however, that if one uses average values by integrating over the cross section, these problems disappear and the MM approach is seen to converge very quickly, even when no mesh extensions are present. Clearly, the MM approach is the more convergent of the two methods, at least in terms of computing component TL.

In view of the improved efficiency of the MM method all future results reported here will be obtained using this method. It is of interest then to examine the potential savings that may be realized by setting $m_1 < n_1$ and $m_3 < n_3$. In Table I, TL values obtained using the MM method are compared for the expansion chamber at frequencies of 300 and 3240 Hz. Here, the value of $m_1(=m_3)$ is systematically increased up to $m_1=n_1=n_3$ for a relatively dense finite element mesh ($n_2=2735$). It is evident in Table I that at low frequencies convergence is very fast and one needs only to include one or perhaps two modes in the calculations. It is not surprising, however, that as the frequency is increased more modes are necessary and at 3240 Hz at least seven modes are required in order to achieve comparable levels of accuracy. Therefore, when using this hybrid approach it would appear to be sensible, at least at higher frequencies, to properly examine solution convergence before reducing values of m_1 and m_3. Alternatively, given that m_1 is normally a lot less than n_2, the potential savings from setting $m_1 < n_1$ are likely to be small, and so it is convenient simply to set $m_1=n_1$ and $m_3=n_3$ when solving the problem. In view of this, all the TL computations that follow are computed with $m_1=n_1$ and $m_3=n_3$, and convergence is examined by modifying n_2.

TABLE I. Convergence of mode matching approach for expansion chamber with plane wave forcing.

	TL (dB)	
$m_1(=m_3)$	$F=300$ Hz	$F=3240$ Hz
1	6.983 130	5.859 270
2	6.978 823	7.057 272
3	6.977 517	7.237 816
4	6.976 890	7.299 525
5	6.976 516	7.328 743
6	6.976 292	7.343 553
7	6.976 051	7.357 351
8	6.975 850	7.368 384
9	6.975 679	7.377 397
10	6.975 526	7.385 279
11	6.975 378	7.392 681
12	6.975 201	7.401 415
$n_1=13$	6.975 171	7.402 905

TL predictions for the expansion chamber are presented in Fig. 4. These predictions are compared with a plane wave analytic solution and an analytic MM solution that includes higher order modes. For purposes of comparison, the TL is plotted against a modified Helmholtz number k_c, where $k_c=2\pi r_1 f/3.8317c_0$, so that a value of $k_c=1$ represents the value at which higher order modes are "cuton" in regions R_1 and R_3. For this expansion chamber, $k_c=1$ equates to a frequency of 5657 Hz and a wavelength of $\lambda=1.64r_1$ m. In Fig. 4 the hybrid method overlays the analytic MM solution over virtually the entire frequency range, and very good agreement with the plane wave model is also observed at lower frequencies, as one would expect. Moreover, the results plotted here have been obtained with a relatively modest number of degrees of freedom, especially at lower frequencies ($n_T=75$ up to $k_c=0.1$, rising to $n_T=1127$ at $k_c=1$). Of course, one would not consider using this technique to design expansion chambers, as analytic techniques would be much faster; however, the results for the expansion chamber clearly demonstrate the efficacy of the hybrid method for interior problems and this will now be investigated further by examining nonuniform geometries.

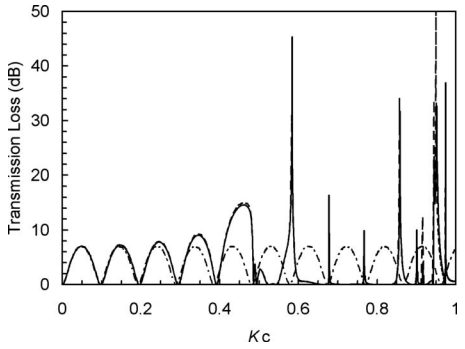


FIG. 4. TL predictions for expansion chamber: —, hybrid method (MM); – – –, analytic mode matching (overlays hybrid method); and –·–·–, plane wave.

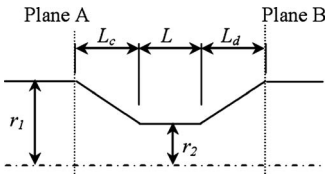


FIG. 5. Geometry of converging-diverging duct.

B. Converging-diverging duct

A simple nonuniform geometry that is often found in the ductwork is the converging-diverging duct. For example, Selamet and Easwaran[2] studied the so-called Herschel–Venturi tube, which has a circular cross section and is used in flow measurements; Lau and Tang[9] studied constrictions that are often found in a rectangular ventilation work. To accommodate both rectangular and circular geometries a representative convergent-divergent duct geometry is shown in Fig. 5. This assumes that a line of symmetry may be drawn through both geometries, remembering that for the circular geometry the solution must be integrated over the circular cross section in the normal way. A well-known analytic solution exists for the circular Herschel–Venturi tube,[2] which permits direct comparison with the current method. Accordingly, in Fig. 6 the hybrid method is compared with analytic predictions for values of the modified Helmholtz number k_c up to unity. Here, two of the geometries studied by Selamet and Easwaran[2] are examined, and for each example $2r_1=4.859$ cm and $L=r_2$; the other dimensions are (a) $r_2=0.5r_1$, $L_c=6.554$ cm, and $L_d=18.534$ cm and (b) $r_2=0.25r_1$, $L_c=9.831$ cm, and $L_d=27.8$ cm. In addition, when $k_c\leqslant 0.2$, $n_T=243$; $0.2<k_c\leqslant 0.5$, $n_T=655$; and $0.5<k_c\leqslant 1$, $n_T=1021$. Here, $k_c=1$ equates to an upper frequency limit of 8614 Hz and $\lambda=1.64r_1$ m. Good agreement between the numerical and analytic solutions is observed in Fig. 6 for values up to $k_c=1$, further validating the approach.

Lau and Tang[9] examined a converging-diverging restriction in a rectangular ventilation duct, which is much larger than the Herschel–Venturi tube studied by Selamet and Easwaran.[2] Accordingly, to cover a relevant frequency range, say, up to 1 kHz, it is necessary to go to a much higher modified Helmholtz number and this requires the inclusion of higher order modes in the inlet and outlet ducts. A significant advantage of the hybrid method is that specifying a nonreflecting boundary is trivial when higher order modes are present, and it is also straightforward to quantify the true sound power propagating in region R_3. In Fig. 7, TL predictions for three of the rectangular convergent-divergent sections studied by Lau and Tang are plotted against a modified Helmholtz number, $k_r=2r_1f/c_0$ (so that higher order modes in R_3 cut on at nk_r, $n=1,2,\ldots,\infty$). Here, when $k_r\leqslant 1$, $n_T=179$; $1<k_r\leqslant 2$, $n_T=429$; and $2<k_r\leqslant 3$, $n_T=667$. Also, a value of $k_r=3$ equates to an upper frequency limit of 1030 Hz and $\lambda=0.67r_1$ m. Note that r_1 now represents the half width of the rectangular section in regions R_1 and R_3 and that only those modes symmetric about the center line of the duct are considered in the analysis. The TL values presented in Fig. 7 generally compare well with Lau and Tang's predictions[9] up to $k_r=1$, although for case (c) some discrepancies are evident as $k_r\rightarrow 1$. TL values for $k_r>1$ are, however, very different from those reported by Lau and Tang and it is noticeable that negative values no longer appear. This is because the hybrid approach includes all modes in the TL calculations rather than examining individual modes. The TL predictions in Fig. 7 assume that plane wave propagation is present in the inlet duct; however, for $k_r>1$ multimode sound propagation is also possible in the inlet duct, especially if the noise source is a fan. Thus, in Fig. 8 TL predictions are presented for an incident sound field containing EMED,[24] with the same value for n_T as used in Fig. 7. It is evident in Fig. 8 that the TL of the converging-diverging section changes significantly if the incident sound field contains higher order modes. For these examples, the TL appears to depend strongly on the respective geometries and frequency of excitation and so no characteristic trends are evident. These results do show, however, that a system is likely to behave very differently if driven by a sound source that contains higher order modes.

In addition to predicting TL, Lau and Tang[9] also plotted the sound pressure distribution. This provides a further opportunity for validating the hybrid method and so in Fig. 9 three different plots of absolute sound pressure level are pre-

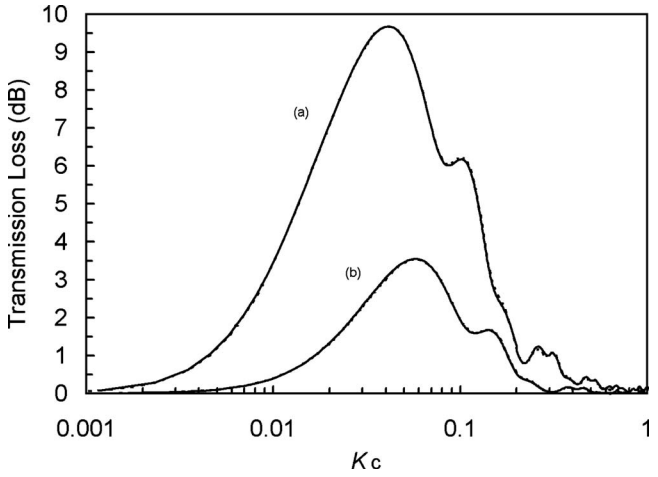


FIG. 6. TL predictions for Selamet and Easwaran's converging-diverging duct (Ref. 2): —, hybrid method (MM) and - - -, analytic method (Ref. 2) (overlays hybrid method). (a) $r_2=0.5r_1$, $L_c=6.554$ cm, and $L_d=18.534$ cm and (b) $r_2=0.25r_1$, $L_c=9.831$ cm, and $L_d=27.8$ cm.

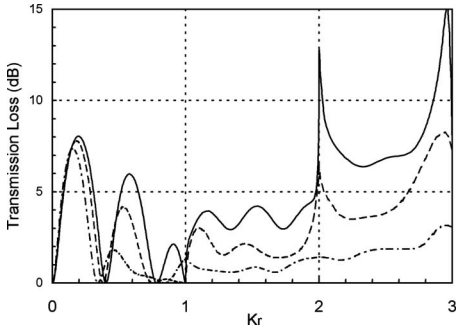


FIG. 7. Hybrid method (MM) predictions for Lau and Tang's converging-diverging duct (Ref. 9) with plane wave excitation. For each plot, $r_1=0.5$ m, $r_2=0.2r_1$, and $L=2r_1$. (a) —, $L_c=L_d=0.2309$ m; (b) – – –, $L_c=L_d=0.4$ m; and (c) –·–·–, $L_c=L_d=0.6928$ m.

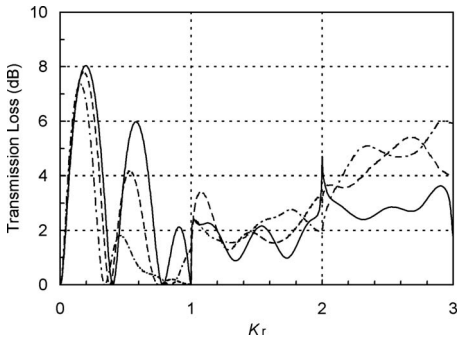


FIG. 8. Hybrid method (MM) predictions for Lau and Tang's converging-diverging duct (Ref. 9) with multimode (EMED) excitation. For each plot, $r_1=0.5$ m, $r_2=0.2r_1$, and $L=2r_1$. (a) —, $L_c=L_d=0.2309$ m; (b) – – –, $L_c=L_d=0.4$ m; and (c) –·–·–, $L_c=L_d=0.6928$ m.

sented, which have been chosen to match the Helmholtz number and contraction geometries presented in Fig. 6 of Ref. 9. A careful comparison with Lau and Tang's results indicates generally good agreement, although it is noticeable that for $k_r=1$ a more resonant response is observed in Fig. 9. Here, it is possible that any discrepancies between the two sets of results are caused by reflections from Lau and Tang's downstream boundary contaminating the pressure field when higher order modes are present.

C. Cylinder

A classical problem in duct acoustics is sound propagation over a cylinder placed on the centerline of a rectangular duct, see, for example, Duan *et al.*[7] A circular cylinder is examined here, which has the geometry shown in Fig. 10. The MM analysis carried out in the previous two sections sought to minimize, as far as possible, the number of degrees of freedom required and so planes A and B were located at duct discontinuities. However, when more complicated geometries are present, it is important to ensure that significant distortion within the finite elements used to discretize region R_2 is avoided in order to minimize numerical errors. One must, therefore, be careful when locating planes A and B, and here these planes are moved a distance L_e from the surface of the cylinder in order to minimize element distortion

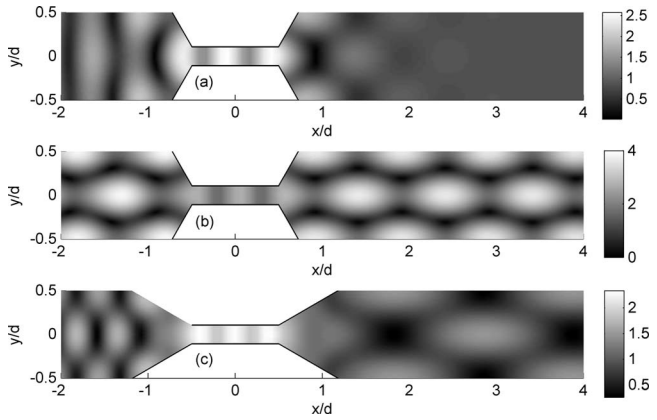


FIG. 9. Magnitude of sound pressure for Lau and Tang's converging-diverging duct (Ref. 9) with plane wave excitation. For each plot, $r_1=0.5$ m, $r_2=0.2r_1$, and $L=2r_1$. (a) $k_r=3/\pi$, $L_c=L_d=0.2309$ m; (b) $k_r=1$, $L_c=L_d=0.2309$ m; and (c) $k_r=4/\pi$, $L_c=L_d=0.6928$ m.

TABLE II. Convergence of mode matching approach for cylinder with EMED forcing.

	TL (dB) at $F=1000$ Hz ($k_r=2.914$) for $r_2=0.8r_1$			
n_T	$L_e=r_2$	$L_e=0.5r_2$	$L_e=0.25r_2$	$L_e=0.1r_2$
53	24.0202	25.3502	29.3648	30.6231
79	21.8790	29.1131	36.8324	33.7745
115	16.1962	21.4286	24.2204	27.3093
151	16.3297	21.6613	22.2208	23.1645
251	19.4340	21.8583	22.4775	23.2662
345	20.4249	21.3426	21.5664	21.7407
519	20.8249	21.3660	21.5725	21.7412
687	21.0402	21.2821	21.3552	21.4128
855	21.1031	21.2574	21.2896	21.3145
1099	21.1654	21.2598	21.2906	21.3146
1607	21.2065	21.2511	21.2663	21.2780
2523	21.2290	21.2454	21.2503	21.2541

close to the cylinder. The convergence of the TL is then investigated in the same way as before, noting that it is essential to use eight noded quadrilateral and/or six noded triangular isoparametric elements in order to accurately reproduce the geometry of a semicircle. The convergence of the TL for $r_2=0.8r_1$ is reviewed in Table II for different values of L_e. Here, the effect of moving planes A and B away from the cylinder can be seen in the rate at which the solution converges, although if the value of L_e is relatively small then numerical errors appear to inhibit convergence. From Table II, a value of $L_e=0.5r_2$ is generally seen to provide an optimum balance between competing requirements and it is this value for L_e that is used in the TL calculations that follow. It should be noted, however, that Table II represents a worst case scenario for those cylinders studied here; at lower frequencies (and for plane wave excitation) convergence is normally much faster and there is little difference between the TL values when L_e is altered. Nevertheless, it is good practice here to carefully investigate convergence for all complex geometries, noting that the rate of convergence will depend on the length of the mesh extension, the geometry of the mesh chosen, and the frequency of excitation. After establishing convergence over a range of frequencies, the TL for a circular cylinder with $r_2=0.4r_1$, $0.6r_1$, and $0.8r_1$ is shown in Fig. 11 for multimode (EMED) forcing, with $n_T=79$ for $k_r \leqslant 1$, $n_T=251$ for $1<k_r\leqslant 2$, and $n_T=687$ for $2<k_r\leqslant 3$. Here, a value of $k_r=3$ equates to an upper frequency limit of 1030 Hz and $\lambda=0.67r_1$ m. In Fig. 11 it is evident that for $k_r<1$ the TL of the cylinder is relatively small, even when 80% of the duct is blocked, although the TL still exhibits the domelike behavior seen for the converging-diverging ducts. However, when higher order modes propagate, the TL is seen

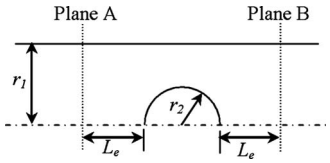


FIG. 10. Geometry of the circular cylinder.

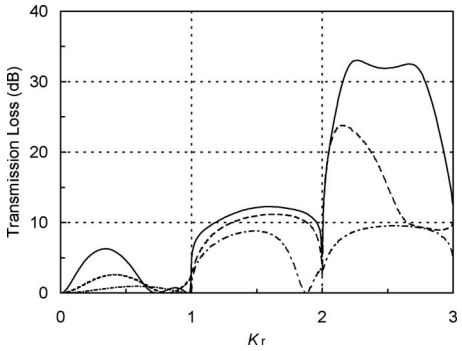


FIG. 11. Hybrid method (MM) predictions for circular cylinder with multimode (EMED) excitation. For each plot, $r_1=0.5$ m and $L_e=0.5r_2$. (a)—, $r_2=0.8r_1$; (b)– – –, $r_2=0.6r_1$; and (c)–·–·–, $r_2=0.4r_1$.

to increase significantly and those modes that cut on at $k_r=2$ and $k_r=3$ strongly influence the sound TL.

In the previous two sections, the TL predictions were partially validated by comparing the hybrid method against analytic predictions. An alternative method for validating the predictions is to examine the relative error in the energy balance over the cylinder. Therefore, if the normalized reflected and transmitted sound power are denoted by W_{ref} and W_{trans}, respectively, then the percentage error ΔE in the sound power is given by $\Delta E=100\times|1-W_{\text{ref}}-W_{\text{trans}}|$, provided that W_{ref} and W_{trans} are normalized against the incident sound power. Values for ΔE may then be computed over the entire frequency range for each cylinder. Here, for $k_r<1$, $\Delta E<10^{-12}\%$; for $1\leqslant k_r<2$, $\Delta E<0.001\%$; and for $2\leqslant k_r<3$, $\Delta E<0.03\%$. These values are representative of each cylinder studied and also apply (and in most cases are much lower for $k_r>1$) to the expansion chamber and converging diverging ducts studied previously. Accordingly, the values quoted for ΔE are sufficiently small to provide confidence in the accuracy of the hybrid method and in the results presented here.

IV. CONCLUSIONS

A hybrid numerical method that combines a wave based modal solution with a standard finite element solution has been reviewed here. The main advantage of this method is that it avoids the need to mesh long uniform sections of ductwork, but at the same time retains the flexibility of the FEM so that relatively short but geometrically complex component sections may be modeled accurately. Two different approaches in enforcing the continuity conditions over the interface between the uniform and nonuniform sections of the ductwork are analyzed: MM and PC. It is observed that the TL predictions obtained using the MM method converge faster than those obtained using PC. Moreover, to ensure good rates of convergence, the PC method requires continuity conditions to be enforced away from duct discontinuities in order to avoid corner nodes. Both the PC and MM methods retain a banded matrix, although only the MM technique will deliver a symmetrical matrix (provided one chooses an appropriate weighting function when enforcing continuity of pressure). Accordingly, the MM version of the hybrid method is capable of retaining all of the benefits of the traditional FEM. Furthermore, the hybrid method delivers a very straightforward way of calculating the component TL, which includes a computationally efficient specification of a nonreflecting boundary condition as well as accommodating higher order modes in the inlet and outlet ducts. This then allows sound propagation to be analyzed at frequencies above the cuton frequency of the first higher order mode in the outlet duct as well as allows for the introduction of a multimodal incident sound field.

The hybrid method is validated here by comparing numerical predictions against analytic solutions as well as other results reported in literature, and good agreement is observed in each case. In order to facilitate the validation of the method, the examples chosen here have been deliberately restricted to relatively simple two-dimensional geometries and so this remains to be demonstrated for more complex three-dimensional component geometries that include inlet/outlet ducts of arbitrary cross section. Of course, for fully three-dimensional models one may expect the number of degrees of freedom required to achieve a converged solution to grow rapidly, but this is true for any numerical model and for most fully three-dimensional shapes one is left with little alternative. The results presented here have been restricted to one component section only, although it is relatively straightforward to add further components since the modal solution in the outlet duct may readily be used as the incident sound pressure field for another component section further downstream. In this way one can build up a number of (multimode) transfer matrices and so examine the interaction between different components in, say, a ventilation system without worrying about the distance between each component. Thus, the hybrid method has the potential to provide an efficient method with which to study sound propagation over complex nonuniform components in a duct and also to study multiple components in order to build up a picture of how sound interacts throughout a ducting system.

[1]J. H. Miles, "Acoustic transmission matrix of a variable area duct or nozzle carrying a compressible subsonic flow," J. Acoust. Soc. Am. **69**, 1577–1586 (1981).
[2]A. Selamet and V. Easwaran, "Wave propagation and attenuation in Herschel-Venturi tubes," J. Acoust. Soc. Am. **101**, 936–942 (1997).
[3]A. Boström, "Transmission and reflection of acoustic waves by an obstacle in a waveguide," Wave Motion **2**, 167–184 (1980).
[4]D. V. Evans, M. Levitin, and D. Vassiliev, "Existence theorems for trapped modes," J. Fluid Mech. **261**, 21–31 (1994).
[5]C. M. Linton and M. McIver, "Trapped modes in cylindrical waveguides," Q. J. Mech. Appl. Math. **51**, 389–412 (1998).
[6]C. M. Linton, M. McIver, P. McIver, K. Ratcliffe, and J. Zhang, "Trapped modes for off-centre structures in guides," Wave Motion **36**, 67–85 (2002).
[7]Y. Duan, W. Koch, C. M. Linton, and M. McIver, "Complex resonances and trapped modes in ducted domains," J. Fluid Mech. **571**, 119–147 (2007).
[8]S. K. Tang and C. K. Lau, "Sound transmission across a smooth nonuniform section in an infinitely long duct," J. Acoust. Soc. Am. **112**, 2602–2611 (2002).
[9]C. K. Lau and S. K. Tang, "Sound transmission across duct constrictions with and without tapered sections," J. Acoust. Soc. Am. **117**, 3679–3685 (2005).
[10]J.-H. Jeong, J.-G. Ih, and B.-C. Lee, "A guideline for using the multidomain BEM for analysing the interior acoustic field," J. Comput. Acoust.1 **3**, 403–424 (2003).
[11]S. Bilawchuk and K. R. Fyfe, "Comparison and implementation of the various numerical methods used for calculating transmission loss in si-

lencer systems," Appl. Acoust. **64**, 903–916 (2003).
[12]O. Z. Mehdizadeh and M. Paraschivoiu, "A three-dimensional finite element approach for predicting the transmission loss in mufflers and silencers with no mean flow," Appl. Acoust. **66**, 902–918 (2005).
[13]D. Givoli, "High-order local non-reflecting boundary conditions: a review," Wave Motion **39**, 319–326 (2004).
[14]A. Mal and Z. Chang, "A semi-numerical method for elastic wave scattering calculations," Geophys. J. Int. **143**, 328–334 (2000).
[15]S. W. Liu, S. K. Datta, and T. H. Ju, "Transient scattering of Rayleigh-Lamb waves by a surface-breaking crack: comparison of numerical simulation and experiment," J. Nondestruct. Eval. **10**, 111–126 (1991).
[16]G. C. Chinn, L. W. Epp, and D. J. Hoppe, "A hybrid finite element method for axisymmetric waveguide-fed horns," IEEE Trans. Antennas Propag. **44**, 280–285 (1996).
[17]H. Esteban, S. Cogollos, V. E. Boria, A. San Blas, and M. Ferrando, "A new hybrid mode-matching/numerical method for the analysis of arbitrarily shaped inductive obstacles and discontinuities in rectangular waveguides," IEEE Trans. Microwave Theory Tech. **50**, 1219–1224 (2002).
[18]A. Monorchio, G. Manara, P. Grassi, and D. Arena, "Analysis of waveguide components and inductive frequency-selective surfaces by using a hybrid mode matching-finite element technique with edge elements," Electromagnetics **24**, 49–67 (2004).
[19]Y. Kagawa, T. Yamabuchi, T. Yoshikawa, S. Ooie, N. Kyouno, and T. Shindou, "Finite element approach to acoustic transmission-radiation systems and application to horn and silencer design," J. Sound Vib. **69**, 207–228 (1980).
[20]R. J. Astley and A. Cummings, "A finite element scheme for acoustic transmission through the walls of rectangular ducts: comparison with experiment," J. Sound Vib. **92**, 387–409 (1984).
[21]M. G. Imhof, "Scattering of acoustic and elastic waves using hybrid multipole expansions—finite element technique," J. Acoust. Soc. Am. **100**, 1325–1338 (1996).
[22]R. J. Astley, "FE mode-matching schemes for the exterior Helmholtz problem and their relationship to the FE-DtN approach," Commun. Numer. Methods Eng. **12**, 257–267 (1996).
[23]R. J. Astley and W. Eversman, "Acoustic transmission in non-uniform ducts with mean flow, Part II: The finite element method," J. Sound Vib. **74**, 103–121 (1981).
[24]R. Kirby and J. B. Lawrie, "A point collocation approach to modelling large dissipative silencers," J. Sound Vib. **286**, 313–339 (2005).
[25]R. J. Astley and A. Cummings, "A finite element scheme for attenuation in ducts lined with porous material: Comparison with experiment," J. Sound Vib. **116**, 239–263 (1987).

Radiation force calculations for ultrasonic fields from rectangular weakly focusing transducers

K. Beissner
Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany

(Received 12 June 2008; revised 11 July 2008; accepted 11 July 2008)

Knowledge of the theoretical relation between acoustic radiation force and ultrasonic power is important for measuring the emitted power, particularly from medical ultrasonic equipment, using a radiation force balance. A far-field method is applied to calculate the radiation force exerted by rectangular field types on a large absorbing target in a nonabsorbing fluid. This is an extension of previous work on fields with circular symmetry. Pistonlike, apodized, and weakly focusing transducer behavior including diffraction effects is dealt with. If diffraction is neglected, i.e., in the geometric or high-frequency limit, the result can be given as a closed-form expression.
 [DOI: 10.1121/1.2968687]

PACS number(s): 43.25.Qp, 43.80.Vj [CCC] Pages: 1941–1949

I. INTRODUCTION

The total time-averaged ultrasonic power P emitted by an ultrasonic transducer is one of the key parameters characterizing the field. Its determination in water is required by international standards such as IEC 61689 (Ref. 1) and IEC 61157 (Ref. 2) for therapeutic and diagnostic ultrasound devices, respectively. The recommended measurement method uses a radiation force balance [IEC 61161 (Ref. 3)]. An object ("target") is inserted in the field and the radiation force, the time-averaged force F acting on it is measured, mostly by means of a balance. F is to be understood as the measured forward component of the radiation force vector $\mathbf{F}$ throughout this paper. Experimental details have repeatedly been described in literature[3,4] and the method has successfully been tested in several international comparisons.[5]

Theory plays an important role in this determination method. The measured radiation force that represents the momentum flux in the field needs to be converted into the ultrasonic power, which expresses the energy flux in the field. The relation between these two values is to be provided by theory, and it depends on the field structure. The most simple field is the plane-progressive wave, and in this case the theoretical result has often been found (e.g., Ref. 6) to be

$$P = cF \tag{1}$$

for a perfectly absorbing target, where c is the speed of sound in the sound-propagating fluid (water) and where, as usual, the open-vessel situation (Langevin condition[6]) has been assumed.

Equation (1), which is tantamount to $cF/P=1$, is widely used in practice. However, real fields usually are nonplane, at least to a certain extent, and then the question arises as to how good the approximation of Eq. (1) is. Deviations from Eq. (1) will occur mainly due to diffraction and focusing. It will not generally be possible to find the exact power/force relation in each practical case, but the relation should be known at least for some representative nonplane field types so that estimates of correction factors or uncertainties can be applied to Eq. (1).

The general three-dimensional radiation force theory is based on the radiation stress tensor concept.[7] In this way the radiation force formula for a large absorber in a circular piston field, including diffraction, has been obtained.[8] It has turned out that in a lossless fluid the result is independent of the distance from the transducer. This is due to the momentum conservation law as expressed by the fact that the radiation stress tensor is a zero-divergence quantity.[7] This can be utilized for devising a simplified far-field calculation method (first applied in Ref. 9) which, however, yields the exact result. In this way numerical radiation force results for circular, focusing, and nonfocusing transducers with an amplitude distribution (apodization) have been obtained.[10] In addition, an approximate closed-form result neglecting diffraction (quasi in the "geometric" or "high-frequency" limit) has been obtained,[10] namely,

$$cF/P = (1 + \cos\ \gamma)/2 \tag{2}$$

for a circular focusing transducer with focus (half-) angle γ. This formula is of increasing importance for measurements in the field of high-intensity focused ultrasound or high-intensity therapeutic ultrasound applications (e.g., Ref. 11).

In all calculations mentioned, circular symmetry was assumed and in the comparisons mentioned,[5] circular transducers were used. However, rectangular transducers and fields, too, are of relevance, particularly in ultrasonic diagnosis with its rectangular linear-array transducers. It appears highly desirable to derive the radiation force formula for rectangular cases, at least for the basic field types. It may be possible to approximately adapt Eq. (1) or Eq. (2) with some analogy arguments to rectangular situations, but explicit results for the fields from rectangular transducers will form a better basis for the evaluation and interpretation of power measurements by means of radiation force in the rectangular case.

The same calculation method as in Ref. 10 will be used here and the results will be presented in a similar fashion. However, the situation is characterized by a lower degree of symmetry than the rotational symmetry assumed in Ref. 10 and, therefore, much more calculational effort is required and there are more characteristic parameters that can be varied.

The following terminology will be used: The origin is at the center of the transducer face. The axial coordinate away from the transducer is z and the coordinates perpendicular to the axis are x and y. The distance from the origin is referred to as r with $r^2=x^2+y^2+z^2$. The transducer is assumed to be rectangular and its length is $2a$ in the x direction and $2b$ in the y direction. The transducer is assumed to be baffled so that the field can be calculated using the well-known Rayleigh integral. Continuous-wave vibration with circular frequency ω and with wavelength λ and circular wavenumber $k=2\pi/\lambda$ in the sound-propagating fluid is assumed. A linear quantity h is used to characterize the transducer size and is chosen to be a mean of a and b. Various choices are possible and would be appropriate for various aspects of the paper, but in general the geometric mean $h=(ab)^{1/2}$ appears to be the best one and is used throughout. The transducer is characterized by h and b/a. Often kh will appear as parameter and sometimes $(kh)^2$ which can simply be understood as $kakb$. The terminology may differ from that in Ref. 10 and in other references.

The principal lateral directions in rectangular fields are in practice often verbally distinguished as, e.g., "azimuth" and "elevation."[12] This distinction is not necessary here. In most of the following cases, there is no difference in the meaning of x and y and, accordingly, a and b, and a radiation force result obtained for a certain b/a value also applies to the reciprocal value of this parameter.[13]

It should be noted that only one aspect of the power/force relation is dealt with here, namely, its dependence on the field structure. Other influences are disregarded or, in other words, the following idealized assumptions are made: (a) the fluid is lossless, (b) the target is a perfect absorber, and (c) the target in its lateral extension covers the entire field and thus the entire momentum flux. In practice, these other influences, too, have to be considered, of course,[3,4] but they are beyond the scope of the present paper. Reflecting targets are not dealt with here. A strict and general radiation force theory for a reflector with curved surfaces (such as the conical reflector often used) so far does not exist.

II. CALCULATION METHOD

A planar, baffled, and rectangular harmonically vibrating transducer is assumed. The general field point Q has the Cartesian coordinates x, y, z but is regarded in spherical coordinates r, θ, φ throughout the paper.

The method of calculating the radiation force exerted on a large perfect absorber is adopted from Ref. 10. Due to the momentum conservation law,[7,8] the (Langevin) radiation force acting on a perfect absorber in a lossless fluid is independent of the shape and distance of the absorber, provided the absorber is large enough to intercept the entire momentum flux leaving the transducer. Contrary to practical measurements, the absorbing target is assumed here to be a hollow half-sphere (hemispherical shell) of inner radius r (Fig. 1) and concentric to the transducer center. The radius r is assumed to be large enough for geometric far-field conditions to prevail, which means that the field at distance r resembles an outgoing spherical wave. In other words, r is

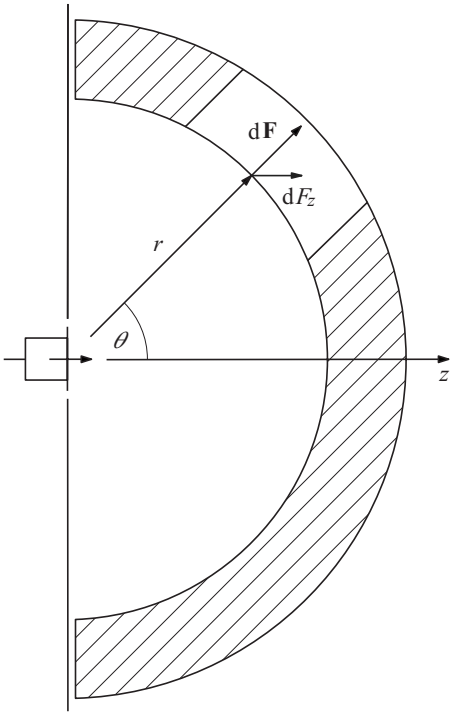


FIG. 1. Schematic section through the arrangement considered with baffled ultrasonic transducer (left) and absorbing target (hatched).

assumed to be large in comparison with all other dimensional parameters, or, the quantities are expanded in power series of $1/r$ and only the lowest order is retained. This can be regarded as the limit of $r\to\infty$. It has been shown in Ref. 10, in comparison with Ref. 8, that in spite of the simplifications, the method yields the exact radiation force result.

The time-averaged intensity vector in the geometric far-field limit is everywhere normal to the spherical target front surface and is characterized by its radial component $I(r,\theta,\varphi)$ in the direction of the radius vector $\mathbf{r}$. The ultrasonic power is the surface integral of I over the inner target surface whereby the infinitesimal surface element in spherical coordinates is $r^2 \sin\theta d\theta d\varphi$, i.e.,

$$P=\iint r^2 I(r,\theta,\varphi)\sin\theta d\theta d\varphi. \tag{3}$$

At every surface point, an almost plane wave normally runs against the absorbing target and the local normal radiation force, $d\mathbf{F}$, per surface element is given by the intensity divided by the speed of sound. However, only its component dF_z in the z direction contributes to the measured radiation force and this means an extra factor $\cos\theta$ so that the radiation force F as measured in the z direction is given by

$$cF=\iint r^2 I(r,\theta,\varphi)\sin\theta\cos\theta d\theta d\varphi. \tag{4}$$

The characteristic ratio cF/P finally is obtained by dividing expression (4) by expression (3).

Note the following three remarks. (a) The integration limits are as follows: for φ from 0 to 2π or, for reasons of symmetry, to $\pi/2$, with an extra factor of 4 that cancels out in the ratio; for θ from 0 to $\pi/2$ minus a certain small ε in

order to physically allow for the (Langevin) pressure relaxation between front and rear target sides, but mathematically ε is set to zero, i.e., neglected. (b) The far-field intensity is proportional to $1/r^2$ and r will disappear from both expressions above. This shows that the result is not bound to a certain r value but is a general and fundamental one. (c) The crucial factor $\cos\theta$ is ultimately due to the fact that energy is a scalar and momentum is a vector.

The far-field intensity I is linked with the real pressure amplitude $\hat{p}$ and the real velocity potential amplitude $\hat{\phi}$ by

$$I = \hat{p}^2/(2\rho c) = k^2 \rho c \hat{\phi}^2/2, \tag{5}$$

where ρ is the density of the fluid. Finally the velocity potential ϕ in the general far-field point Q needs to be derived. This is done using the well-known Rayleigh integral (Ref. 14, for example). A point Q' (with coordinates x', y') on the transducer surface is considered in addition. The distance from Q' to Q is given by

$$\begin{aligned}|\mathbf{r}-\mathbf{r}'| &= \{(x-x')^2 + (y-y')^2 + z^2\}^{1/2} \\ &= r\left\{1 - 2\frac{x'}{r}\sin\theta\cos\varphi - 2\frac{y'}{r}\sin\theta\sin\varphi + \left(\frac{x'}{r}\right)^2 \right. \\ &\quad \left. + \left(\frac{y'}{r}\right)^2\right\}^{1/2}. \end{aligned} \tag{6}$$

x' and y' are bounded by the transducer dimensions and, as already discussed, the field point is assumed to be in the geometric far field so that $r \gg |x'|$ and $r \gg |y'|$. This is not the Fresnel approximation in the classical sense[15] but has recently been considered a special form of the Fresnel approximation.[16] If the square root in Eq. (6) is expanded in a power series of x'/r and y'/r, and only terms up to the first order are retained, the result is

$$\begin{aligned}|\mathbf{r}-\mathbf{r}'| &= r\left\{1 - \frac{x'}{r}\sin\theta\cos\varphi - \frac{y'}{r}\sin\theta\sin\varphi\right\} \\ &= r - x'\sin\theta\cos\varphi - y'\sin\theta\sin\varphi. \end{aligned} \tag{7}$$

When this asymptotic expression is inserted in the Rayleigh integral,[14] the far-field velocity potential in Q is given by a double integral over the transducer surface, namely,

$$\begin{aligned}\phi(r,\theta,\varphi) = \frac{1}{2\pi r}\iint v_n(x',y')\exp(jk[x'\sin\theta\cos\varphi \\ + y'\sin\theta\sin\varphi])dx'dy', \end{aligned} \tag{8}$$

where j is the imaginary unit and the general time-dependent factor $\exp(j[\omega t - kr])$ has been suppressed. The real amplitude appearing in Eq. (5) is the magnitude (modulus) of expression (8). v_n is the normal velocity of the transducer surface. Its distribution over the transducer face is the main input information for this algorithm. Three cases will be dealt with, namely, (a) v_n being a constant=piston behavior, (b) v_n being a real function=amplitude distribution (apodization), and (c) v_n being a complex function=focusing.

The double integral in Eq. (8) is equivalent to a two-dimensional Fourier transform. The full algorithm as described by Eqs. (3)–(5) and (8) goes over four stages of integration, i.e., it is tantamount to a quadruple integral. The problem can be reduced to the level of a triple or even double integration if $v_n(x',y')$ in Eq. (8) either is simple enough so that the integral in Eq. (8) can be reduced or even solved analytically or the influences of x' and y' are separable, at least approximately, so that Eq. (8) reduces to the product of two single integrals. The latter approach can also be found in Ref. 16 where various methods including fast Fourier transform are discussed. Equation (8), however, is only one step here, and anyway, the additional integrations of Eqs. (3) and (4) still remain. The solution procedure applied here is numerical integration throughout using Gaussian quadrature formulas.[17] As shown in Sections III–VI, this means double or triple integration depending on the particular circumstances.

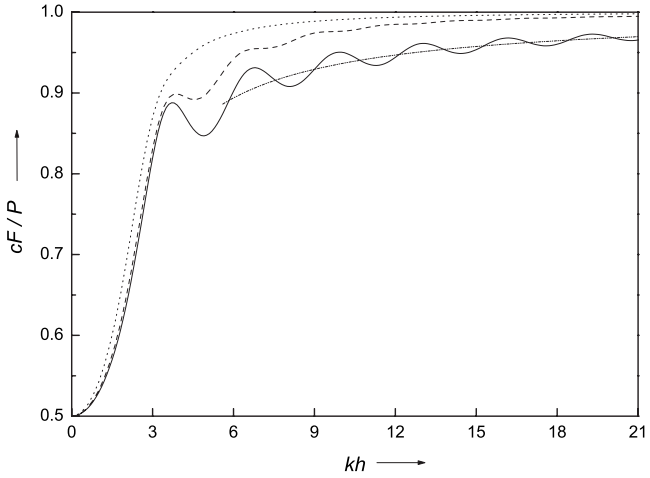


FIG. 2. Radiation force ratio cF/P as a function of the transducer's kh value; unfocused, $b/a=1$. Solid line: AD1 (piston); dash: AD2; dot: AD3; and dash-dot: average piston behavior.

III. RECTANGULAR PISTON

Here the function $v_n(x',y')$ is simply a constant v_0 and Eq. (8) can be analytically solved to yield

$$\begin{aligned}\phi(r,\theta,\varphi) = \frac{2abv_0}{\pi r}\mathrm{fct}_1(ka\sin\theta\cos\varphi) \\ \times \mathrm{fct}_1(kb\sin\theta\sin\varphi), \end{aligned} \tag{9}$$

where fct_1 is the well-known *sinc* function,[16] namely, $\mathrm{fct}_1(u)=\sin u/u$. In this case ϕ is purely real and its square can be directly inserted into Eq. (5) and then into Eqs. (3) and (4). They are evaluated by means of numerical double integration. The final result for cF/P is presented as the solid line in Figs. 2 and 3 for $b/a=1$ and $b/a=1.8$, respectively, as a function of kh. The curves show diffraction oscillations but generally start at $cF/P=0.5$ and tend toward the plane-wave value $cF/P=1$ for kh going to infinity. The range of high kh values will be covered in figures that are shown later.

The faint dash-dot curve in Figs. 2 and 3 represents the "average" behavior and has been calculated as follows. In Ref. 10 the average behavior of the circular piston was found to be given by $1-2/(\pi ka)$, where a was the piston radius and the expression resulted from Hankel's asymptotic expansion of the Bessel functions for large arguments. Bessel functions do not appear in the rectangular case, but the same

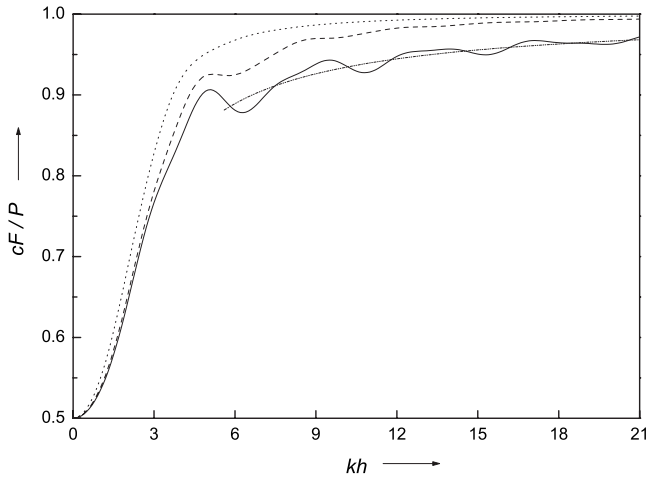


FIG. 3. Same as Fig. 2, but $b/a=1.8$.

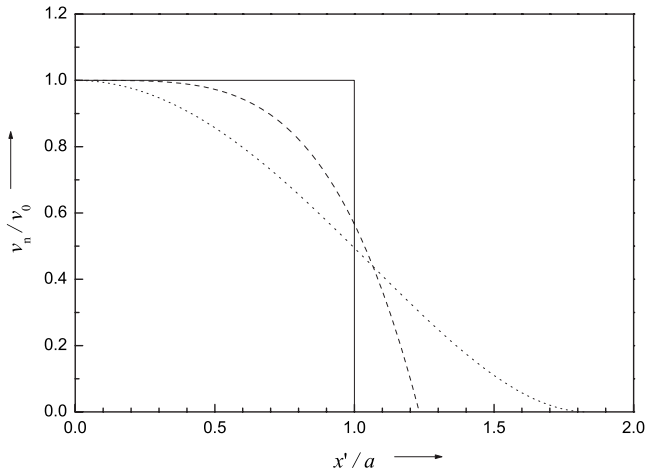


FIG. 4. Distribution of the transducer's relative normal vibration velocity amplitude along the x' axis. Solid line: AD1 (piston); dash: AD2; and dot: AD3.

formula type was tried here, with the radius being replaced with h. This was successful but only with a different definition of the mean, namely, with the harmonic mean h_h instead of the geometric mean h, i.e., $h_h=2/(1/a+1/b)=2h\sqrt{b/a}/(1+b/a)$. This means that the dash-dot curves follow the formula $1-2/(\pi k h_h)$. This approximation was numerically found to work well not only in the cases of Figs. 2 and 3 but also for various other b/a values.

IV. APODIZATION

The velocity amplitude goes through a Heaviside step function at the piston rim. Other amplitude distributions, namely, those with a continuous amplitude decrease ("apodization"), are also of interest; this is simulated using a real function for $v_n(x',y')$. Various approaches exist in literature (Ref. 16, and citations therein). The aim here is not to simulate a particular real transducer type but to demonstrate the general behavior. Straightforward and convenient amplitude distributions are chosen for this purpose and it is assumed that v_n is a (separable) product of functions of x' and y' and that these functions are polynomials so that the integral (8), again, can be solved analytically.

It is now advisable to distinguish between two kinds of transducer dimensions, namely, "true" and "effective" ones. The true values a_t and b_t are those values of x' and y' where the vibration amplitude becomes zero and beyond which it remains zero. The effective values a and b are those of an equivalent rectangular piston, equivalent with respect to a certain transducer property that is to be defined in each case. Various approaches are possible; the definition chosen here will be given later. The linear quantity h is understood to be associated with the effective values. For the piston itself, true and effective dimensions are identical, of course.

For simplicity, the form of the apodization curve is assumed to be the same in both lateral directions. The piston case already discussed will be referred to as amplitude distribution 1 (AD1). Amplitude distribution 2 (AD2) is

$$v_n = v_0\{1-(x'/a_t)^4\}\{1-(y'/b_t)^4\}. \tag{10}$$

Amplitude distribution 3 (AD3) is

$$v_n = v_0\{1-2(x'/a_t)^2+(x'/a_t)^4\}\{1-2(y'/b_t)^2+(y'/b_t)^4\}. \tag{11}$$

The integration in Eq. (8) extends for x' from $-a_t$ to a_t and for y' from $-b_t$ to b_t. The exact solution of Eq. (8) for AD2 is

$$\phi(r,\theta,\varphi) = \sigma_2 \frac{a_t b_t v_0}{\pi r} \mathrm{fct}_2(ka_t \sin\theta\cos\varphi) \times \mathrm{fct}_2(kb_t \sin\theta\sin\varphi), \tag{12}$$

with the factor $\sigma_2=32/25$ and the function $\mathrm{fct}_2(u)=5\{u(6-u^2)\cos u-3(2-u^2)\sin u\}/u^5$. The exact solution of Eq. (8) for AD3 follows the same formula as Eq. (12), but now with $\sigma_3=128/225$ replacing σ_2 and $\mathrm{fct}_3(u)=15\{(3-u^2)\sin u-3u\cos u\}/u^5$ replacing fct_2. The functions fct_2 and fct_3 have been scaled so that they tend toward 1 for u going to zero, like fct_1.

The respective function fct in each case gives the far-field directivity pattern as a function of θ in both lateral directions, which means (a) for $\varphi=0$ as a function of $ka_t \sin\theta$ in the x-z plane and (b) for $\varphi=\pi/2$ as a function of $kb_t \sin\theta$ in the y-z plane. This far-field directivity behavior is used here for the definition of the effective transducer dimensions, namely, on the basis of the first zero of the relevant function. The condition for a_t/a and b_t/b is that the first zero of ϕ as a function of θ occurs at the same θ value as for the equivalent (rectangular) piston. This leads to $a_t/a=b_t/b=1.232$ for AD2, $a_t/a=b_t/b=1.835$ for AD3, and $a_t/a=b_t/b=1$ for AD1, of course.

Figure 4 shows the three amplitude distributions, here, as an example, along the x' axis, i.e., for $y'=0$, and only for positive x' values which means that the curves should be understood to be symmetrically continued on the other side. The coordinate x' is normalized to the effective value a just discussed. The same curves could be depicted as a function of y'/b.

Equation (12) gives the explicit analytical result of Eq. (8) in this case. Again, ϕ is purely real and its square can be directly inserted into Eq. (5), and then into Eqs. (3) and (4)

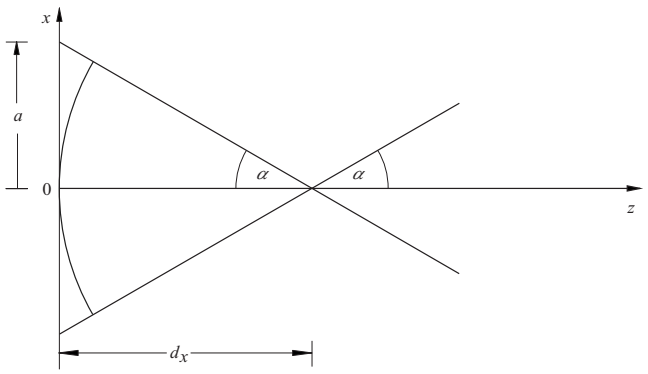


FIG. 5. Schematic representation of the quantities used for the description of the focused field in the x-z plane.

with numerical double integration. The cF/P results are included in Figs. 2 and 3 as dash and dot curves. The diffraction oscillations are much lower than in the piston case, and the values approach the plane-wave result $cF/P=1$ much faster for increasing kh. The behavior is completely similar to the results for circular transducers shown in Ref. 10, particularly there in the logarithmic plot of Fig. 4 which need not be repeated for the present case. The fundamental difference between the amplitude distributions is that the deviation of cF/P from the plane-wave value 1 is proportional to $1/(kh)$ for the (rectangular) piston but is proportional to $1/(kh)^2$ for the apodized distributions. This behavior has been discussed in detail in Ref. 10 for the circular case.

Again, the range of high kh values will be covered in figures that are shown later.

V. POINT FOCUS

We keep to the plane-rectangular transducer; focusing is brought about by assuming an additional phase distribution in v_n in accordance with

$$v_n(x',y') = \hat{v}_n(x',y')\exp(j\delta(x',y')). \tag{13}$$

Here $\hat{v}_n$ is the real function describing the amplitude distribution which again is given by the constant v_0 for AD1 or by Eq. (10) or Eq. (11) in the case of apodization. The function δ will be referred to as the phase function and it expresses the phase difference (phase advancement) of a Huygens elementary wavelet at the transducer point (x',y') in comparison with that at the transducer center required to obtain the focusing pattern desired. It should be noted (a) that in order to be mathematically straightforward and in conformity with the theoretical literature, continuous phase functions are considered here, whereas a stepwise phase function is often the practical solution for plane-focusing transducers and (b) that the phase function is chosen disregarding the diffractive edge wave, i.e., with regard to the "geometric" focusing pattern desired, but when it is then inserted in the Rayleigh integral, the field including diffraction is obtained. In this paper, the term focus is understood in the geometric sense only.

New quantities and symbols are used as described in Fig. 5 which shows the x-z plane. A geometrically perfect focusing at an axial distance d_x from the transducer is desired. To this end, the phase function needs to be such that the sections of the surfaces of constant phase with the x-z plane are circles concentric to the focus. The surface of constant phase touching the transducer center is referred to here as the principal surface of constant phase and is depicted as a circle in Fig. 5. The focus (half-) angle is $\alpha=\arctan(a/d_x)$. A similar figure could be drawn for the y-z plane, with the symbols b replacing a, d_y replacing d_x, and $\beta=\arctan(b/d_y)$ replacing α.

The focal lengths d_x and d_y can be equal or unequal. In this section, the case $d_x=d_y=d$ is dealt with. The mean quantity h defined and discussed earlier is associated with another angle, namely, with $\gamma=\arctan(h/d)$. Unless otherwise stated, all new symbols refer to the effective values of the dimensional quantities. The true values of the dimensional quantities are still required, e.g., as the integration boundaries in Eq. (8), but they will not be mentioned explicitly in the text. The quantitative relation between true and effective values is more difficult here than in the unfocused case as the far-field directivity pattern is not calculated explicitly and is likely to have no zeros. For reasons of simplicity and consistency, the same ratio values as in Sec. IV are used.

With $d_x=d_y=d$, the geometric focus is a point and the surfaces of constant phase are spheres. The principal surface of constant phase could be thought of as being replaced with a spherically curved transducer vibrating with uniform phase. This is the situation dealt with by O'Neil,[18] however, for circular, not for rectangular beam shape.

The exact (but inseparable) form of the phase function now is

$$\delta/k = (d^2 + x'^2 + y'^2)^{1/2} - d = d\{[1 + (x'/d)^2 + (y'/d)^2]^{1/2} - 1\}. \tag{14}$$

For weak focusing, i.e., for $(x'/d)^2 \ll 1$ and $(y'/d)^2 \ll 1$, it can be developed. The approximation of order 1 is

$$\delta/k = x'^2/(2d) + y'^2/(2d). \tag{15}$$

The approximation of order 2 is

$$\delta/k = x'^2/(2d) + y'^2/(2d) - (x'^4 + 2x'^2y'^2 + y'^4)/(8d^3). \tag{16}$$

Formula (15) leads to the integral (8) being separable, but formula (16) not, as it contains a mixed term with $x'y'$. However, generally it is $2x'^2y'^2 \le x'^4+y'^4$ which suggests omitting the mixed term and ending up with, so to speak, a separable approximation of order 1.5, namely,

$$\delta/k = x'^2/(2d) - x'^4/(8d^3) + y'^2/(2d) - y'^4/(8d^3). \tag{17}$$

Approach (15) can be found in literature.[16] Expression (17) is an improvement and will be used in the following. It is inserted into Eq. (13), and this into Eq. (8). The double integral (8) is now the product of two single integrals, each consisting of a real and an imaginary part, to be evaluated numerically. The resulting ϕ is complex and its magnitude (modulus) is inserted into Eq. (5) which then leads to the numerical double integration in accordance with Eqs. (3) and (4).

Typical results obtained are displayed as follows: Fig. 6 shows the radiation force relation for $b/a=1.8$ and for a

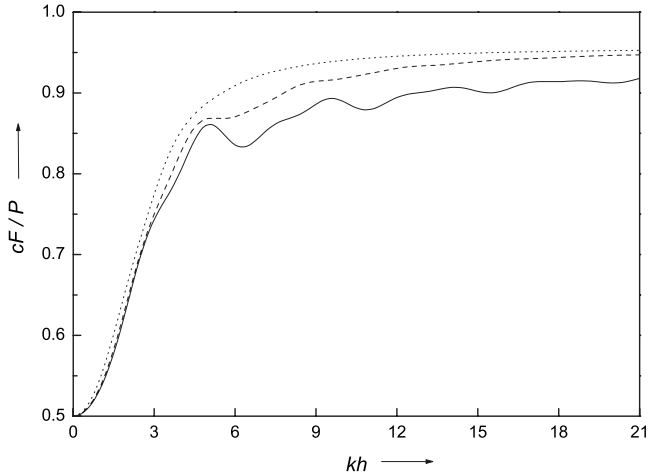


FIG. 6. Radiation force ratio cF/P as a function of the transducer's kh value; focused (point focus), $b/a=1.8$, $\gamma=20°$. Solid line: AD1 (piston); dash: AD2; and dot: AD3.

focus (half-) angle of $\gamma=20°$. The three amplitude distributions already discussed have been taken into account. The figure should be compared with Fig. 3 for the unfocused case. The behavior is very similar but with the difference that the curves obviously do not tend toward the plane-wave value 1 but remain lower. This is just the same effect as was found in Ref. 10 for circular transducers.

Next, the results obtained so far will be shown for high kh values. First we deal with AD1 (pistonlike amplitude distribution). As already mentioned, the overall behavior of cF/P is obviously characterized by a $1/(kh)$ dependence. Figures 7 and 8, therefore, present the numerical cF/P results as a function of $15/(kh)$ in the range $100 \geq kh \geq 15$. Figure 7 is for $b/a=1$ and Fig. 8 is for $b/a=1.8$. Each figure includes three curves for values of the focus (half-) angle γ equal to 0° (=unfocused) and to 10° and 20°. The upper curve in both figures is the continuation of the solid line in Fig. 2 or Fig. 3, respectively. The lowest curve in Fig. 8 is the continuation of the solid line in Fig. 6 ("continuation" in the sense that increasing kh here means a motion from the right to the left).

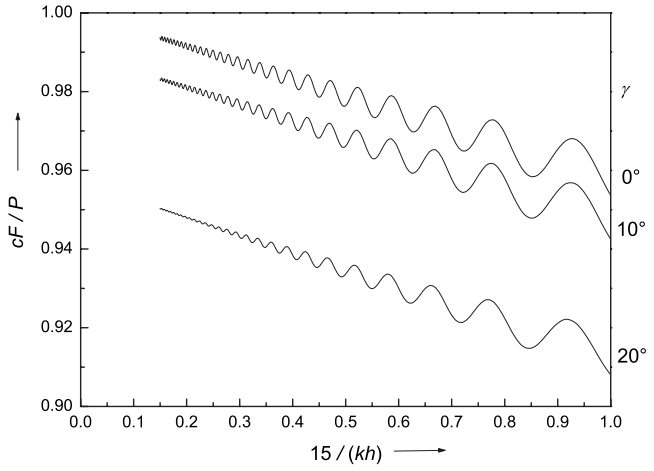


FIG. 7. Radiation force ratio cF/P as a function of $(kh)^{-1}$; AD1, $d_y/d_x=1$, $b/a=1$. Parameter: focus (half-) angle γ.

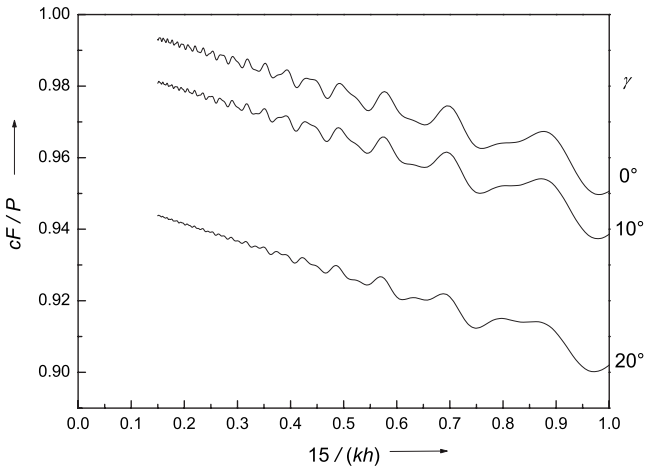


FIG. 8. Same as Fig. 7, but $d_y/d_x=1$, $b/a=1.8$.

The calculations have also been carried out using AD2 and AD3. Just one example is presented in Fig. 9. As already mentioned, the overall behavior of cF/P for apodized amplitude distributions is obviously characterized by a $1/(kh)^2$ dependence (see also Ref. 10). Figure 9, therefore, presents numerical cF/P results as a function of $225/(kh)^2$ in the range $60 \geq kh \geq 15$. Figure 9 is for AD2 and for $b/a=1.8$. The upper curve is the continuation of the dash line in Fig. 3; the lowest curve is the continuation of the dash line in Fig. 6. Resulting curves for AD3 are not shown here for the sake of brevity. They are very similar to those for AD2, but no diffraction oscillations are discernible.

As also found for circular transducers in Ref. 10, focusing leads to an extra decrease in the measured radiation force, even in the high-frequency limit of $kh \to \infty$ or $\lambda \to 0$. The high-frequency values of cF/P obtained from the numerical results by extrapolation to $kh \to \infty$ (i.e., to $15/(kh) \to 0$ or $225/(kh)^2 \to 0$, respectively) are listed in Table I. Column 2 refers to Fig. 7, column 5 refers to Fig. 8, and column 6 refers to Fig. 9. Figures for the cases covered by columns 3, 4, and 7 have been omitted here for the sake of brevity. Columns 8–10 refer to Sec. VI.

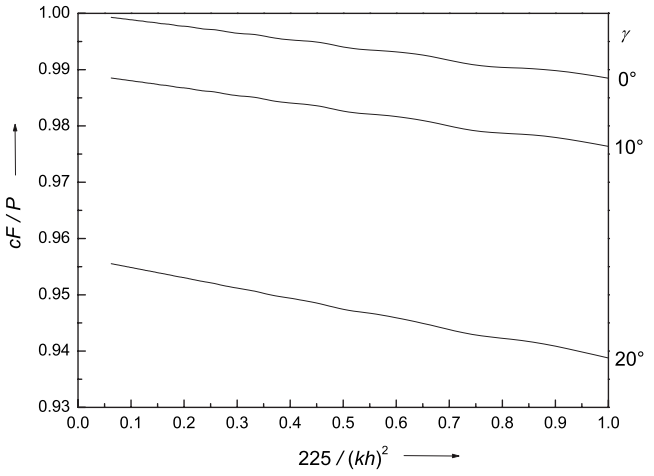


FIG. 9. Radiation force ratio cF/P as a function of $(kh)^{-2}$; AD2, $d_y/d_x=1$, $b/a=1.8$. Parameter: focus (half-) angle γ.

TABLE I. High-frequency values of cF/P obtained from the numerical results by extrapolation to $kh \to \infty$.

1	2	3	4	5	6	7	8	9	10
d_y/d_x	1	1	1	1	1	1	1.8	1.8	1.8
b/a	1	1	1	1.8	1.8	1.8	1	1	1
AD	1	2	3	1	2	3	1	2	3
$\gamma=0°$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\gamma=10°$	0.990	0.991	0.991	0.988	0.989	0.989	0.988	0.989	0.989
$\gamma=20°$	0.957	0.962	0.961	0.951	0.957	0.956	0.951	0.957	0.956

VI. TWO LINE FOCI

We now deal with the case $d_x \neq d_y$, and it is assumed without loss of generality that $d_y > d_x$. The configuration is now characterized by two line foci instead of one point focus. The focal line that is at an axial distance d_x from the transducer lies in the y-z plane. The focal line that is at an axial distance d_y from the transducer lies in the x-z plane. Figure 5 with d_x remains valid and the corresponding one with d_y as well, but the values are now different.

Given d_x and d_y, the main task is to find the principal surface of constant phase and the required phase function first, if possible, in an exact form analogous to Eq. (14) and then in a suitable approximation. The sections of the principal surface of constant phase with the x-z plane and with the y-z plane still are circles but with different curvature, and the question arises as to how the form changes when turning from one plane to the other.

It is quite natural to start assuming that both focal lines are straight which would be the most simple case. However, the author has found in a lengthy analysis that the problem of finding the phase function under this condition is overdetermined and has no exact solution. So the condition needs to be relaxed by assuming that one focal line is straight and the other one (at least slightly) is curved. Now the problem can be solved exactly but there are two different solutions. In both cases, the principal surface of constant phase is from the generalized torus family. The members of this family are briefly described in Ref. 19, for example.

If the straight focal line is to be the more remote one (which means that it is located at d_y in our terminology), the principal surface of constant phase is (part of) a torus in the more strict sense[20] and the phase function is given by

$$\delta/k = \{[(d_y^2 + y'^2)^{1/2} - d_y + d_x]^2 + x'^2\}^{1/2} - d_x. \quad (18)$$

If the straight focal line is to be the closer one (which means that it is located at d_x in our terminology), the principal surface of constant phase is (part of) a lemon[21] and the phase function is given by[22]

$$\delta/k = \{[(d_x^2 + x'^2)^{1/2} - d_x + d_y]^2 + y'^2\}^{1/2} - d_y. \quad (19)$$

The (exact) expressions (18) and (19) are different from each other[23] and are both inseparable. Developing them for weak focusing, i.e., for $x'^2, y'^2 \ll d_x^2, d_y^2$, leads to the following: The approximations of order 1 are equal to each other and are separable. The approximations of order 1.5 (in the similar sense as in Sec. V) are equal to each other and are separable. The approximations of order 2 are different from each other and are inseparable. The reason for the latter effect is that, as in Eq. (16), a mixed term appears in the approximation of order 2 and that it is just this term which is different in the two cases.

As in Sec. V, the approximation of order 1.5 will be used for the practical calculations and the phase function is now given by

$$\delta/k = x'^2/(2d_x) - x'^4/(8d_x^3) + y'^2/(2d_y) - y'^4/(8d_y^3). \quad (20)$$

It should be emphasized that the restriction to the approximation of order 1.5, which in Sec. V was merely introduced to reduce the number of integration stages, here also has a physical aspect, namely, to avoid the difference between the two cases considered. In other words, the restriction to weak focusing in the sense just shown enables us to consider the two focal lines similar in character and practically straight. This is an approximation, of course, but it will also be important in Sec. VII.

We still have $h=(ab)^{1/2}$, and now accordingly $d=(d_x d_y)^{1/2}$. Again $\gamma=\arctan(h/d)$. Numerical calculations are carried out as described in Sec. V. Figure 10 presents results for AD1 with $d_y/d_x=1.8$ and $b/a=1$. As before, the cF/P results are shown as a function of $15/(kh)$ in the range $100 \geq kh \geq 15$ for various values of γ. The upper curve is identical with that in Fig. 7; in the unfocused case, the parameter "ratio of the focal lengths" is irrelevant. Although there are some minor differences, the overall behavior of the curves (including those for AD2 and AD3, not shown here for the sake of brevity) is very similar to those discussed previously. The form of the diffraction oscillations is again determined by b/a.

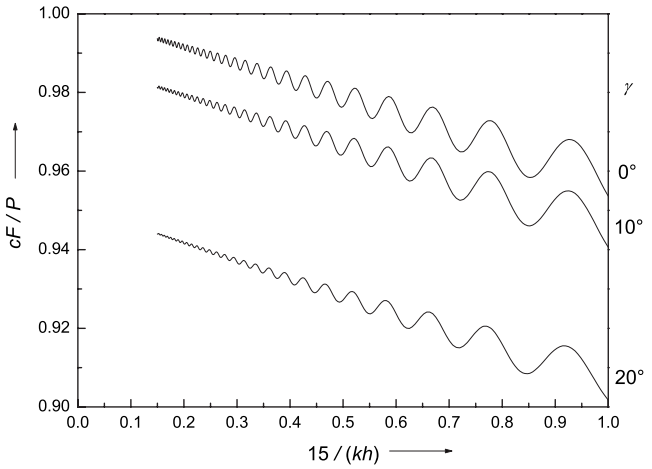


FIG. 10. Same as Figs. 7 and 8, but $d_y/d_x=1.8$, $b/a=1$.

Again, the high-frequency values of cF/P obtained from the numerical results by extrapolation to $kh \to \infty$ are listed in Table I, namely, in columns 8–10. Column 8 refers to Fig. 10.

VII. CLOSED-FORM RESULT NEGLECTING DIFFRACTION

So far, a limited number of sample calculations for certain parameter values have been carried out. What would be needed in practice is a formula for any set of parameter values occurring in reality, but this is not feasible. The values obtained can be understood as resulting from the superposition of two influences, namely, (a) diffraction and (b) the geometric focusing action, both leading to a decrease in cF/P from the plane-wave value. The diffraction effect depends on parameters such as frequency, temporal waveform, apodization, etc., and general predictions for particular cases are difficult, the more so since only continuous-wave behavior has been dealt with here. The reader may draw a general understanding of the diffraction influence from the curves presented and may extend this to other cases of interest.

The geometric focusing action, on the other hand, depends only on the geometric parameters a, b, d_x, and d_y which can be combined to just two crucial quantities, namely, the focus (half-) angles $\alpha = \arctan(a/d_x)$ and $\beta = \arctan(b/d_y)$. An attempt will be made to derive a closed formula for the influence of these parameters on cF/P in a fashion similar to formula (2) for circular beams.[10] This will only be possible by neglecting diffraction (i.e., in the high-frequency limit) and assuming other simplifying conditions as discussed below. The resulting formula is intended for a comparison with the extrapolated values of Table I and should be seen in that context.

As already mentioned, diffraction effects are neglected and a Heaviside step-function behavior of the acoustic quantities at the rim of the beam is assumed, i.e., a sudden step-down from the full values inside the geometric beam[24] to zero outside. Consider first the geometric beam for $d_x = d_y = d$. Starting from the rectangular transducer surface, it is a converging rectangular pyramid until the point focus is reached. At the point focus it starts again as a rectangular pyramid, this time diverging until it hits the spherical inner target surface at r. In conformity with Sec. II, we consider the situation asymptotically in the limit of $r \to \infty$, and it is irrelevant whether we place the starting point of the diverging pyramid at the point focus or at the origin (d being finite). The section of the pyramid with the spherical inner target surface is a spherical quadrangle whose dimensions are determined by the aperture (half-) angles α and β of the pyramid. This spherical quadrangle now is the range of integration in Eqs. (3) and (4). The intensity in Eqs. (3) and (4) is now assumed to be a constant $I_0(r)$ over the surface of integration; otherwise it would be difficult or impossible to obtain a closed-form result.

Consider then the geometric beam for $d_x \neq d_y$. It is assumed that both focal lines are straight.[25] Starting from the transducer surface, the beam cross section changes its rectangular form until it shrinks into the first focal line. Then, a rectangle comes up but shrinks into the second focal line. Then again, the rectangle appears but continues to change its form with increasing z. However, for $z \to \infty$ the width to height ratio tends toward a characteristic value that is determined by α and β alone. Again, for simplicity, the intensity in Eqs. (3) and (4) is taken to be constant.

TABLE II. Selected values of cF/P obtained using Eq. (25).

1	2	3	4
d_y/d_x	1	1	1.8
b/a	1	1.8	1
$\gamma \to 0°$	1.000	1.000	1.000
$\gamma = 10°$	0.990	0.988	0.988
$\gamma = 20°$	0.961	0.956	0.956

To summarize both cases, the integrals (3) and (4) are to be performed over a spherical quadrangle and take on the forms

$$P = r^2 I_0(r)\text{integ1} \quad \text{with} \quad \text{integ1} = \iint \sin\theta d\theta d\varphi \tag{21}$$

and

$$cF = r^2 I_0(r)\text{integ2} \quad \text{with} \quad \text{integ2} = \iint \sin\theta \cos\theta d\theta d\varphi, \tag{22}$$

and the integration is as follows: for φ from 0 to 2π or, for reasons of symmetry, to $\pi/2$, with an extra factor of 4 that cancels out in the ratio; for θ from 0 to a $\theta_{\max}(\varphi)$ which is given by $\arctan(\tan\alpha/\cos\varphi)$ for $0 \leq \varphi \leq \arctan(\tan\beta/\tan\alpha)$ and by $\arctan(\tan\beta/\sin\varphi)$ for $\arctan(\tan\beta/\tan\alpha) \leq \varphi \leq \pi/2$.

The integral (21) can be calculated in the way described but can also be obtained more directly as follows: integ1 is the solid angle subtended by the spherical quadrangle as seen from the origin; this is connected with its area and is equal to its (spherical) excess.[19,26] From a general formula for spherical polygons,[26] the following result can be derived here:

$$\text{integ1} = 4 \arcsin(\sin\alpha \sin\beta). \tag{23}$$

On the other hand, Eq. (22) must be evaluated by explicitly solving the integral that finally leads to

$$\text{integ2} = 2[\sin\alpha \arctan(\cos\alpha \tan\beta) + \sin\beta \arctan(\cos\beta \tan\alpha)], \tag{24}$$

and the characteristic ratio cF/P is given by

$$cF/P = \text{integ2}/\text{integ1}. \tag{25}$$

Table II presents values obtained using Eq. (25) which correspond to the cases included in Table I as a function of d_y/d_x, b/a, and γ. Only α and β appear in Eq. (25) but both values are related to d_y/d_x, b/a, and γ in each case in accordance with $\tan\alpha = [(d_y/d_x)/(b/a)]^{1/2} \tan\gamma$ and $\tan\beta = [(b/a)/(d_y/d_x)]^{1/2} \tan\gamma$. Only one amplitude distribution has been considered in this section, namely, a constant amplitude over the spherical quadrangle, and so the number of

cases is reduced here by a factor of 3 in comparison with Table I.

VIII. CONCLUDING REMARKS

Corresponding values in Tables I and II are almost identical but there are some small differences in detail, a behavior similar to the results for circular fields in Ref. 10. The reasons for these small differences can generally be seen in the approximate assumptions that had to be introduced at several places. Two additional aspects deserve mentioning as follows.

(a) There is no universally valid definition of the effective transducer dimensions. Each definition regards one particular transducer property and brings it into line with the corresponding property of a piston transducer but does not necessarily do the same with other properties. One might consider demanding that in our case, the effective transducer dimensions be such that the results listed in Table I coincide for the various amplitude distributions, but this would not be feasible as it would require a fitting procedure, i.e., repeated runs through the entire algorithm.

(b) In Sec. VII, a constant amplitude over the spherical quadrangle at distance r has been assumed. This is equivalent, at least approximately, with a constant amplitude over the principal surface of constant phase. In Sections III–VI, however, amplitude distributions over the planar transducer surface were considered. The planar transducer surface, on the one hand, and the (curved) principal surface of constant phase, on the other hand, have been related to each other with respect to phase, but this does not necessarily mean a straightforward relation concerning amplitude. The reason for differences in the amplitude distribution is that the waves from the various transducer points have different path lengths until they reach the principal surface of constant phase and, therefore, they arrive there with different amplitudes. These differences increase with increasing focus angle. In summary, Eq. (25), together with Eqs. (23) and (24), is to be regarded as a useful aid for correction and uncertainty estimates in practical radiation force measurements in rectangular fields, but does not exactly take into account all details.[27]

[1] *IEC 61689: Ultrasonics—Physiotherapy Systems—Field Specifications and Methods of Measurement in the Frequency Range 0.5 MHz to 5 MHz*, 2nd ed. (International Electrotechnical Commission, Geneva, 2007).

[2] *IEC 61157: Standard Means for the Reporting of the Acoustic Output of Medical Diagnostic Ultrasonic Equipment*, 2nd ed. (International Electrotechnical Commission, Geneva, 2007).

[3] *IEC 61161: Ultrasonics—Power Measurement—Radiation Force Balances and Performance Requirements*, 2nd ed. (International Electrotechnical Commission, Geneva, 2006).

[4] K. Beissner, "Radiation force and force balances," in *Ultrasonic Exposimetry*, edited by M. C. Ziskin and P. A. Lewin (CRC, Boca Raton, 1993), pp. 127–142; "Primary measurement of ultrasonic power and dissemination of ultrasonic power reference values by means of standard transducers," Metrologia **36**, 313–320 (1999).

[5] C. E. Tschiegg, M. Greenspan, and D. G. Eitzen, "Ultrasonic continuous-wave beam-power measurements; international intercomparison," J. Res. Natl. Bur. Stand. **88**, 91–103 (1983); K. Beissner, W. A. Oosterbaan, R. T. Hekkenberg, and A. Shaw, "European intercomparison of ultrasonic power measurements," Acust. Acta Acust. **82**, 450–458, 671 (1996); K. Beissner, "Report on key comparison CCAUV.U-K1 (ultrasonic power)," Metrologia **39** (2002); Tech. Suppl. 09001, available at www.iop.org/EJ/toc/0026-1394/39/1A

[6] J. A. Rooney and W. L. Nyborg, "Acoustic radiation pressure in a traveling plane wave," Am. J. Phys. **40**, 1825–1830 (1972).

[7] K. Beissner, "The acoustic radiation force in lossless fluids in Eulerian and Lagrangian coordinates," J. Acoust. Soc. Am. **103**, 2321–2332 (1998).

[8] K. Beissner, "Acoustic radiation pressure in the near field," J. Sound Vib. **93**, 537–548 (1984); "Minimum target size in radiation force measurements," J. Acoust. Soc. Am. **76**, 1505–1510 (1984).

[9] L. L. Nider, "Ultrasonic power measurement with a radiation force balance," M.S. thesis, University of Illinois, Urbana, IL, 1979.

[10] K. Beissner, "Radiation force calculations," Acustica **62**, 255–263 (1987).

[11] S. Maruvada, G. R. Harris, B. A. Herman, and R. L. King, "Acoustic power calibration of high-intensity focused ultrasound transducers using a radiation force technique," J. Acoust. Soc. Am. **121**, 1434–1439 (2007).

[12] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out* (Elsevier, Amsterdam, 2004).

[13] In Sec. VI, however, replacing b/a with its reciprocal value would need to be accompanied by a replacement of d_y/d_x with its reciprocal value, but it is assumed there that $d_y > d_x$. This again is independent of "azimuth" and "elevation."

[14] G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).

[15] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, 1968).

[16] T. D. Mast, "Fresnel approximations for acoustic fields of rectangularly symmetric sources," J. Acoust. Soc. Am. **121**, 3311–3322 (2007).

[17] A. H. Stroud and D. Secrest, *Gaussian Quadrature Formulas* (Prentice-Hall, Englewood Cliffs, NJ, 1966).

[18] H. T. O'Neil, "Theory of focusing radiators," J. Acoust. Soc. Am. **21**, 516–526 (1949).

[19] E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics* (Chapman and Hall, London/CRC, Boca Raton, 1999) As the way of presentation there is slightly different from what is needed here, the following is briefly explained: A generalized torus is understood here as a circle (or part of a circle) that is rotated about an axis which lies in the plane of the circle. There are two special lines, namely, (a) the (straight) axis and (b) the (circular) curve formed by the center of the circle during its rotational motion. These two lines can be identified with our focal lines. Members of the family are ring torus, horn torus, apple, sphere, and lemon, depending on the position of the rotation axis with respect to the circle.

[20] That is, a ring torus or a horn torus or an apple, depending on the value of d_y/d_x. For the torus terminology, see Ref. 19.

[21] According to Ref. 19, a lemon is the inside surface of a spindle torus, whereas an apple is the outside surface of a spindle torus.

[22] The sphere can be considered a special form of a torus, and Eqs. (18) and (19) lead to Eq. (14) in the case of $d_x = d_y$.

[23] Equations (18) and (19) look formally similar as they follow from each other by interchanging x' with y' and d_x with d_y. This, however, does not change the fact that, given the two focal line distances from the transducer, there are two different phase functions depending on whether the more distant or the closer line is to be the straight one.

[24] In comparison with the preceding sections, the geometric rim of the beam here is associated with the "effective" parameters of the preceding sections; What is called the geometric beam here has been referred to as the "main beam" in L. W. Schmerr, Jr., A. Sedov, and T. P. Lerch, "A boundary diffraction wave model for a spherically focused ultrasonic transducer," J. Acoust. Soc. Am. **101**, 1269–1277 (1997).

[25] This approximation is in line with Sec. VI and is again important for obtaining a closed-form result.

[26] D. Zwillinger, *Standard Mathematical Tables and Formulae*, 30th ed. (CRC, Boca Raton, 1996), Sec. 4.18.

[27] Note added in resubmission: The present paper deals with symmetric nonscanning (or, in other words, "frozen") beams. Any angular scanning would affect the angular dependencies of the field quantities and would most probably change the results obtained, but this is beyond the scope of the present paper.

Bubble growth by rectified diffusion at high gas supersaturation levels

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

(Received 7 December 2007; revised 24 July 2008; accepted 25 July 2008)

For high gas supersaturation levels in liquids, on the order of 300% as predicted in capillaries of marine mammals following a series of dives [D. S. Houser, R. Howard, and S. Ridgway, J. Theor. Biol. **213**, 183–195 (2001)], standard mathematical models of both static and rectified diffusion are found to underestimate the rate of bubble growth by 10%–20%. The discrepancy is demonstrated by comparing predictions based on existing mathematical models with direct numerical solutions of the differential equations for gas diffusion in the liquid and thermal conditions in the bubble. Underestimation of bubble growth by existing mathematical models is due to the underlying assumption that the gas concentration in the liquid is given by its value for a bubble of constant equilibrium radius. This assumption is violated when high supersaturation causes the bubble to grow too fast in relation to the time scale associated with diffusion. Rapid bubble growth results in an increased gas concentration gradient at the bubble wall and therefore a growth rate in excess of predictions based on constant equilibrium bubble radius. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2973235]

PACS number(s): 43.25.Yw, 43.35.Ei [OAS] Pages: 1950–1955

I. INTRODUCTION

There is continued interest in possible effects of underwater sound on the growth of gas bubbles in capillaries of marine mammals and humans. Numerical simulations motivated by this interest have been reported by Crum and Mao,[1] who considered gas supersaturation levels up to 223%. Houser *et al.*[2] subsequently calculated the evolution of the nitrogen tension and relative supersaturation due to typical dive profiles executed by dolphins, blue whales and beaked whales. Supersaturation levels were predicted to reach 250%–300% upon completion of the dive profiles, depending upon the species. The highest level of 300% was predicted for beaked whales.

One assumption in existing models is that supersaturation in the liquid far from the bubble is constant. However, for constant external pressure P_0 the gas concentration C decays to its equilibrium level C_0 according to the kinetic equation

$$\frac{dC}{dt} = -\frac{C - C_0}{t_w}, \tag{1}$$

where $t_w = t_h/\ln 2$ is the exponential time constant for nitrogen washout, and t_h is the corresponding half-time. Ridgway and Howard[3] measured $t_h \sim 6$ min for washout from dolphin muscle tissue. Other assumptions underlying existing models include large Péclet numbers (dimensionless ratio of frequency to diffusion coefficient), coefficients that do not depend on pressure and temperature, linear theory for thermal and radiation damping of the bubble oscillations, and gas concentration in the liquid given by its value for a bubble of constant equilibrium radius. All of these assumptions can be analyzed numerically. The purpose of our paper is to present a relatively efficient numerical scheme for performing such calculations and to compare results with those from previous work. It is found that the last approximation, that of neglecting the effect of the rate at which the equilibrium bubble radius grows on the gas concentration in the surrounding liquid, leads to underestimation of bubble growth rate by 10%–20% at high supersaturation levels.

While recent work on bubble growth in marine mammals at high gas supersaturation levels motivated the present study, it is not the purpose of this paper to speculate on the impact of our study in this context. *In vivo* bubble dynamics is a complicated process affected by many factors, such as constraints imposed by tissue, nonspherical bubble shapes, and surfactants. Also to be considered is the matter of nucleation.[4] Uncertainties associated with such additional factors can easily dominate the contribution due to the percentage change in growth rate predicted here in connection with rectified diffusion in an unbounded liquid. We nevertheless performed our calculations for sound pressure levels that may be of interest in studies of the effects of underwater sound on marine mammals, for which the percentage change in equilibrium radius predicted by the model presented here is at most about 20%.

II. THEORETICAL MODELS

A. Static (nonacoustic) diffusion

The basis for all existing theoretical models is Epstein and Plesset's model[5] for static (nonacoustic) diffusion. The diffusion equation for spherical symmetry is

$$\frac{\partial C}{\partial t} = \frac{D}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial C}{\partial r}\right), \tag{2}$$

where D is the diffusion constant. The initial concentration of the gas at $t=0$ is assumed to be uniform and given by C_∞, and for $t>0$ the gas concentration C_R at the bubble wall is determined by Henry's law,

$$C_R = P_g/H_D, \tag{3}$$

where P_g is gas pressure in the bubble, and H_D is Henry's constant. The equilibrium concentration of gas in a liquid at a planar boundary and for a given pressure P_0 in the liquid is $C_0 = P_0/H_D$. At the surface of a spherical gas bubble we have, when vapor pressure is negligible,

$$\frac{C_R}{C_0} = 1 + \frac{2\sigma}{P_0 R_0}, \tag{4}$$

where R_0 is the bubble radius, and σ is the surface tension, which are related through $P_g = P_0 + 2\sigma/R_0$.

Epstein and Plesset[5] used the above relations to derive the following result for the rate of change in bubble radius:

$$\frac{dR_0}{dt} = \frac{Dd}{1 + 4\sigma/3P_0R_0}\left[\frac{C_\infty}{C_0} - \left(1 + \frac{2\sigma}{P_0R_0}\right)\right] \times \left(\frac{1}{R_0} + \frac{1}{\sqrt{\pi Dt}}\right), \tag{5}$$

where $d = k_B T/H_D$ is a dimensionless parameter, T is temperature, and k_B is Boltzmann's constant. The physical sense of this parameter is that, for a given temperature, it is the ratio of dissolved gas concentration near the bubble wall, C_R, to the density of gas in the bubble. According to Henry's law, Eq. (3), this ratio does not depend on concentration. However, it does depend on temperature.

It follows from Eq. (5) that the bubble loses mass ($dR_0/dt < 0$) for $C_\infty < C_R$. When the liquid far from the bubble is saturated ($C_\infty = C_0$) Eq. (4) yields $C_R > C_\infty$, and the bubble dissolves. Only for $C_\infty = C_R = C_0(1 + 2\sigma/P_0R_0)$, i.e., for which the supersaturation is given by $1 + 2\sigma/P_0R_0$, is the bubble stabilized. For supersaturation greater than this value the bubble will grow.

When effects of surface tension are ignored ($\sigma = 0$), and for bubble radius sufficiently large ($R_0 \gg 2\sigma/P_0$), Eq. (5) can be solved analytically.[5] However, the analytical solution for this case exists only in complicated parametric form and can be analyzed only by plotting graphs for different parameters. In this case it is preferable to solve Eq. (5) numerically.

Equation (5) is derived from the gas mass flow estimated using the diffusion equation with the boundary condition imposed on a bubble of constant radius. This approach is valid only for low levels of supersaturation and when the bubble growth rate is sufficiently small. Also, the initial conditions are constant gas concentration everywhere from infinity to the bubble wall. Therefore, at the initial time $t = 0$ there is a jump in the gas concentration across the bubble wall. The gradient across this jump in gas concentration is thus singular at $t = 0$, and it is the source of the singular term $1/\sqrt{\pi Dt}$ that appears in Eq. (5). The nonsingular term that contains $1/R_0$ corresponds to the gradient of the concentration at the bubble wall for the stationary concentration

$$C(r) = C_\infty + (C_R - C_\infty)\frac{R_0}{r}. \tag{6}$$

For initial conditions that correspond to this stationary concentration there is no singularity. Therefore, results depend on the initial concentration, and a stationary initial concentration is more realistic than one that is singular.

B. Gas diffusion outside bubble with changing radius

Fick's law, Eq. (2) with a convection term included, is used to account for motion of the bubble wall. Using a transformed version of Fick's law in Lagrangian coordinates, Eller and Flynn[6,7] have shown that for the case of a small dimensionless parameter $1/\mathrm{Pe} = D/\omega R_0^2$, or large Péclet number Pe, where ω is a characteristic angular frequency of the bubble motion, the equation for the equilibrium bubble radius becomes

$$\frac{dR_0}{dt} = \frac{Dd}{1 + 4\sigma/3P_0R_0}\left(\frac{C_\infty}{C_0} - \frac{\langle P_g R^4\rangle}{P_0\langle R^4\rangle}\right) \times \left[\frac{\langle R\rangle}{R_0^2} + \left(\frac{\langle R^4\rangle}{R_0^4 \pi Dt}\right)^{1/2}\right], \tag{7}$$

where $R(t)$ is the instantaneous bubble radius, and the angular brackets indicate time averaging.

Fyrillas and Szeri[8] subsequently improved Eq. (7) by splitting the transformed Fick's equation into an equation for the rapidly oscillating part of the solution and another equation for the slowly varying part, for the same asymptotic case of large Péclet number. They showed that the solution of the oscillatory problem yields zero mean flux of gas particles through the bubble wall. To solve the problem for the slowly varying part they used a standard averaging method to obtain

$$\frac{dR_0}{dt} = \frac{Dd}{1 + 4\sigma/3P_0R_0}\left(\frac{C_\infty}{C_0} - \frac{\langle P_g R^4\rangle}{P_0\langle R^4\rangle}\right) \times \frac{1}{R_0}\int_0^\infty \frac{dx}{\langle[3x + (R/R_0)^3]^{4/3}\rangle}. \tag{8}$$

The singularity $1/\sqrt{\pi Dt}$ was not retained in this formulation.

Both Eqs. (7) and (8) were used by Crum and Mao[1] in their numerical simulations. In general the time averages in these equations must be calculated using numerical solutions of the Rayleigh–Plesset equation. Crum and Mao also discuss an analytical approximation of these time averages, valid only for small bubble oscillations, based on a perturbation solution of the Rayleigh–Plesset equation in powers of the small parameter P_a/P_0, where P_a is the acoustic pressure.

III. NUMERICAL MODEL

The starting point for our numerical model is the form of the diffusion equation, Fick's equation in Lagrangian spherical coordinates, obtained by Eller and Flynn[7] and used by Fyrillas and Szeri:[8]

$$\frac{\partial C}{\partial t} = D\frac{\partial}{\partial s}\left[(3s + R^3)^{4/3}\frac{\partial C}{\partial s}\right], \tag{9}$$

where $3s = r^3 - R^3 = r_L^3 - R_0^3$, and where r_L is the Lagrangian radial coordinate, and r is the Eulerian radial coordinate. The bubble radius R obeys an ordinary differential equation. We will use a standard form of the Rayleigh–Plesset equation,

$$R\ddot{R} + \frac{3}{2}\dot{R}^2 = \frac{1}{\rho}\left(P_g - P_0 - P_a \cos \omega t - \frac{2\sigma}{R} - 4\mu\frac{\dot{R}}{R}\right), \quad (10)$$

where ρ is the density of the liquid, σ is the surface tension, μ is viscosity, P_a and ω are the amplitude and frequency of the acoustic field in the vicinity of the bubble, respectively, P_0 is the ambient pressure, and P_g is the pressure of the gas in the bubble. In absence of an acoustic field $P_g = P_{g0} = P_0 + 2\sigma/R_0$. The number N of gas molecules in the bubble obeys the equation

$$\frac{dN}{dt} = 4\pi R^2 D \frac{\partial C}{\partial r}, \quad (11)$$

where the derivative $\partial C/\partial r$ is calculated at the bubble wall, $r=R$.

The boundary conditions on the bubble wall and at infinity are, respectively,

$$C(s=0) = C(r_L = R_0) = C_R = \frac{P_g}{H_D}, \quad (12)$$

$$C(s=\infty) = C(r_L = \infty) = C_\infty. \quad (13)$$

To avoid the boundary condition at infinity, or for very large r_L, we transform Eq. (9) using the dimensionless, reciprocal distance $\xi = R_0/r_L$,

$$\frac{\partial C}{\partial t} = \frac{D}{R_0^2}\xi^4 \frac{\partial}{\partial \xi}\left\{\left[1 + \xi^3\left(\frac{R^3}{R_0^3} - 1\right)\right]^{4/3}\frac{\partial C}{\partial \xi}\right\}, \quad (14)$$

where $\xi \in (0,1)$. The boundary conditions for Eq. (14) are

$$C(\xi = 1) = C_R, \quad C(\xi = 0) = C_\infty, \quad (15)$$

and Eq. (11) becomes

$$\frac{dN}{dt} = -\frac{4\pi D R^4}{R_0^3}\frac{\partial C}{\partial \xi}, \quad (16)$$

where the derivative $\partial C/\partial \xi$ is calculated at $\xi = 1$.

An implicit finite difference scheme with a homogeneous grid on the interval $\xi \in (0,1)$ was developed to solve Eq. (14). The coordinate ξ is particularly convenient for these calculations because it provides a dense grid in real space near the bubble wall, where the gradient of the concentration is greatest.

A. Radiation losses

The influence of liquid compressibility on the dynamics of a single bubble has been discussed in many papers.[9–11] In Ref. 12 it was shown that the influence of compressibility can be taken into account by adding the term $\dddot{V}/4\pi c$ to the right-hand side of Eq. (10), where V is the bubble volume, and c is the sound speed in the liquid. When expanded in terms of the bubble radius this correction term becomes $(R^2\dddot{R} + 6R\dot{R}\ddot{R} + 2\dot{R}^3)/c$, which coincides with the terms obtained previously by Prosperetti[10] to account for compressibility. In particular, it accounts for additional energy loss during bubble oscillations through acoustic radiation.

We now rewrite $\dddot{V}/4\pi c$ as follows:

$$\frac{\dddot{V}}{4\pi c} = \frac{1}{3c}\frac{d^3R^3}{dt^3} = \frac{1}{c}\frac{d}{dt}[R(R\ddot{R} + 2\dot{R}^2)]. \quad (17)$$

Making use of the fact that the correction to the Rayleigh–Plesset equation that accounts for compressibility is of order $1/c$ we may rewrite Eq. (10) in the form

$$R\ddot{R} = -\frac{3}{2}\dot{R}^2 + \frac{1}{\rho}\left(P_g - P_0 - P_a \cos \omega t - \frac{2\sigma}{R} - 4\mu\frac{\dot{R}}{R}\right) + O(1/c), \quad (18)$$

substitution of which in Eq. (17) yields, up to order $1/c$,

$$\frac{\dddot{V}}{4\pi c} = \frac{1}{2c}\dot{R}^3 + \frac{1}{c}R\dot{R}\ddot{R} + \frac{1}{\rho c}\frac{d}{dt}\left[R\left(P_g - P_0 - P_a \cos \omega t - \frac{2\sigma}{R} - 4\mu\frac{\dot{R}}{R}\right)\right]. \quad (19)$$

Adding this approximation of $\dddot{V}/4\pi c$ to the right-hand side of Eq. (10) and rearranging yields

$$\left(1 - \frac{\dot{R}}{c}\right)R\ddot{R} + \frac{3}{2}\left(1 - \frac{\dot{R}}{3c}\right)\dot{R}^2 = \frac{1}{\rho}\left(1 + \frac{\dot{R}}{c} + \frac{R}{c}\frac{d}{dt}\right) \times \left(P_g - P_0 - P_a \cos \omega t - \frac{2\sigma}{R} - 4\mu\frac{\dot{R}}{R}\right). \quad (20)$$

This formulation is equivalent to Keller's equation[13] at order $1/c$ [like Eq. (20), Keller's equation is only accurate to order $1/c$]. One reason for our use of Eq. (20) is its derivation from the compressibility term $\dddot{V}/4\pi c$, the physical origin of which is easy to understand.[12] Another reason is that there is a slight computational advantage to coding Eq. (20) for acoustic pressures $P_a(t)$ that are more complicated than the simple sinusoidal excitation $P_a \cos \omega t$ considered in the present work, e.g., sonar signals or lithotripsy shock waves. The Keller formulation requires evaluation of a time shifted pressure $P_a[t + R(t)/c]$ rather than a pressure and its derivative at the instantaneous time t.

B. Temperature dependence of Henry's law

Bubble oscillation generates a finite temperature gradient in the surrounding liquid that is proportional to the excitation level and the temperature of the surrounding water.[14] The effect of this gradient on the bubble dynamics was shown to be negligible for the case studied in Ref. 14, but gas solubility is also temperature dependent, and has not previously been accounted for in rectified diffusion studies.

Henry's law for our case is given by Eq. (3), where H_D depends on temperature. In Ref. 15 there are tables for Henry's constants and their temperature dependencies. Our definition of H_D corresponds to $(\mathcal{N}k_H)^{-1}$ in Ref. 15, where $\mathcal{N}$ is the Avogadro constant and k_H is the concentration of a species in the aqueous phase divided by the partial pressure of that species in the gas phase. Therefore, $d = k_B T_0/H_D$ corre-

sponds to k_H^{cc} for $T=T_0$ in Ref. 15, where k_H^{cc} is the dimensionless ratio of the aqueous-phase concentration of a species to its gas-phase concentration.

According to Ref. 15, the measurements fall in the ranges $0.0149<k_H^{cc}<0.0159$ for nitrogen and $0.0294<k_H^{cc}<0.0318$ for oxygen. Using mean values of the reported ranges and approximating air as 80% nitrogen and 20% oxygen we take $k_H^{cc}=0.019$, which yields the value $d(T_0)=0.02$ that was used in our calculations. The temperature dependence according to Ref. 15 is

$$d(T) = d(T_0)\exp\left[\alpha_H\left(\frac{1}{T}-\frac{1}{T_0}\right)\right], \tag{21}$$

where $\alpha_H=1300$ K for nitrogen.

C. Heat conduction in the gas

A polytropic equation of state for the gas inside the bubble was used in Ref. 1, which is common practice because of its ease of implementation. This approach has been shown to be inadequate for large amplitude motion.[14,16] Further, the model in Ref. 14 describes heat conduction in the absence of mass flux. In this section, we develop a model for heat conduction in the gas that accurately describes bubble growth by rectified diffusion under high amplitude acoustic excitation, and which explicitly accounts for mass flux across the bubble wall.

The thermometric conductivity χ_g of the gas is

$$\chi_g = \nu_g/\mathrm{Pr}. \tag{22}$$

For air, $\nu_g=0.15$ cm^2/s is the dynamic viscosity coefficient and Pr=0.733 is the Prandtl number, and therefore $\chi_g \simeq 2.04\times10^{-5}$ m^2/s. The characteristic length scale for the thermal conductivity is

$$L_g = \sqrt{\chi_g/2\omega}. \tag{23}$$

For $\omega/2\pi=100$ Hz we have $L_g=127$ μm, which means that for $R_0\leqslant 100$ μm we can ignore the fact that the gas temperature is not uniform inside the bubble. For $\omega/2\pi=10$ kHz we have $L_g=12.7$ μm and $R_0\geqslant 10$ μm, and it is necessary to take into account the finite value of L_g. The effect of this inhomogeneity is that the equation of state for the gas is close to adiabatic and oscillations of the bubble radius are reduced. In this case rectified diffusion of the gas dissolved in the liquid is also less.

To estimate the effect of a nonuniform temperature distribution we must solve the equation for heat conduction in the gas. An equation such as Eq. (14) cannot be used because it applies only to incompressible liquids, and compressibility of the gas must be taken into account. However, for our problem the full equations of gas dynamics are unnecessary because bubble size is small compared to a wavelength, permitting the simplification whereby the gas pressure P_g may be assumed uniform inside the bubble.

To construct the finite difference scheme for this problem it is convenient to transform the radial coordinate r inside the bubble, which occupies the variable domain $0\leqslant r\leqslant R$, to a normalized spatial coordinate ζ in the fixed domain $0\leqslant\zeta\leqslant 1$. First the heat diffusion equation is written in coordinates that are similar to Lagrangian coordinates but permit bubble growth and variation in the mass of the gas. We thus introduce the variable $m(r,t)$, which is the mass of gas inside a sphere of radius r. Introduction of this variable permits us to ignore the convection term in the heat conduction equation. The equation for the gas temperature T_g thus becomes

$$c_p\rho_g\frac{\partial T_g(m,t)}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(\kappa_g r^2\frac{\partial T_g}{\partial r}\right), \tag{24}$$

where c_p is the specific heat, and ρ_g is the density of the gas. The thermal conductivity κ_g of the gas depends on temperature but not on pressure, and according to the kinetic theory of gases it can be taken proportional to the square root of the gas temperature,

$$\kappa_g = \kappa_{g0}\sqrt{T_g/T_{g0}}, \tag{25}$$

where κ_{g0} is the thermal conductivity at the reference temperature T_{g0}.

Introducing the reference thermometric conductivity for a gas with reference density ρ_{g0},

$$\chi_{g0} = \frac{\kappa_{g0}}{c_v\rho_{g0}}, \tag{26}$$

we may rewrite Eq. (24) as follows:

$$\frac{\partial T_g(m,t)}{\partial t} = \chi_{g0}\frac{\rho_{g0}}{\rho_g}\frac{1}{r^2}\frac{\partial}{\partial r}\left(\sqrt{\frac{T_g}{T_{g0}}}r^2\frac{\partial T_g}{\partial r}\right) \tag{27}$$

$$= \chi_{g0}\frac{\rho_{g0}}{\rho_g}\frac{\partial}{\partial(r^3/3)}\left[\sqrt{\frac{T_g}{T_{g0}}}r^4\frac{\partial T_g}{\partial(r^3/3)}\right] \tag{28}$$

$$= (4\pi)^2\chi_{g0}\rho_{g0}^2\frac{P_g}{P_{g0}}\frac{\partial}{\partial m}\left[\sqrt{\frac{T_{g0}}{T_g}}r^4\frac{\partial T_g}{\partial m}\right], \tag{29}$$

where the final form results from noting that

$$d[(4/3)\pi r^3] = dm/\rho_g. \tag{30}$$

The maximum value of m is $m_0(t)=MN(t)$, which is the total mass of gas in the bubble, and where M is mass per molecule and N is the number of molecules. Since this quantity changes with time it is convenient to introduce the relative mass $\zeta=m/m_0\in(0,1)$. In terms of ζ Eq. (29) takes the form

$$\frac{\partial T_g(\zeta,t)}{\partial t} - \frac{\zeta}{N}\frac{\partial T_g}{\partial\zeta}\frac{\partial N}{\partial t} = \left(\frac{4\pi}{NM}\right)^2\chi_{g0}\rho_{g0}^2\frac{P_g}{P_{g0}} \times\frac{\partial}{\partial\zeta}\left[\sqrt{\frac{T_{g0}}{T_g}}r^4\frac{\partial T_g}{\partial\zeta}\right]. \tag{31}$$

We can calculate r from Eq. (30):

$$r^3 = \frac{3NM}{4\pi}\int_0^\zeta\frac{d\zeta'}{\rho_g} = \frac{3NM}{4\pi\rho_{g0}}\frac{P_{g0}}{P_g}\int_0^\zeta\frac{T_g}{T_{g0}}d\zeta'. \tag{32}$$

After substituting $r^4=(r^3)^{4/3}$ in Eq. (31) we have

$$\frac{\partial T_g(\zeta,t)}{\partial t} - \frac{\zeta}{N}\frac{\partial T_g}{\partial \zeta}\frac{\partial N}{\partial t} = \left(\frac{36\pi}{NM}\right)^{2/3} \chi_{g0}\rho_{g0}^{2/3}\left(\frac{P_{g0}}{P_g}\right)^{1/3} \times \frac{\partial}{\partial \zeta}\left[\sqrt{\frac{T_{g0}}{T_g}} s_T^{4/3}\frac{\partial T_g}{\partial \zeta}\right], \tag{33}$$

where

$$s_T(\zeta,t) = \int_0^{\zeta} \frac{T_g}{T_{g0}} d\zeta'. \tag{34}$$

A semi-implicit finite-difference scheme with a homogeneous grid on the interval $\zeta \in (0,1)$ was developed to solve Eq. (33) numerically. Equation (32) permits calculation of the gas pressure P_g for any given bubble radius $R = r|_{\zeta=1}$ and gas temperature T_g.

D. Model implementation

A computer code was written to calculate bubble growth in supersaturated liquid in the presence of a sound field via solution of a subset of equations presented in the previous sections. Gas diffusion in the liquid is described by Eq. (14), bubble dynamics by Eq. (20), and thermal conductivity of the gas in the bubble by Eq. (33). To account for heat conduction in the liquid surrounding the growing bubble an equation of the same type as Eq. (9) is solved. The time step Δt for the calculations was taken to be of order 10^{-7} s. Every time step consists of four substeps: (i) a substep for the ordinary differential equation for bubble radius and gas mass, Eqs. (20) and (11); (ii) a substep for the adiabatic expansion of the gas to achieve a new bubble radius; (iii) a substep for gas diffusion, Eq. (14); and (iv) a substep for gas thermal conductivity, Eq. (31). At the end of the fourth substep the gas pressure was calculated using Eq. (32) as follows:

$$\frac{P_g}{P_{g0}} = \frac{3NM}{4\pi\rho_{g0}R^3}\int_0^1 \frac{T_g}{T_{g0}} d\zeta = \frac{3NM}{4\pi\rho_{g0}R^3} s_T|_{\zeta=1}. \tag{35}$$

Substeps (ii) and (iv) combined describe the gas thermodynamics in exactly the same way as Prosperetti *et al.*[14] It may be noted that their Eq. (12) depends on c_p whereas our Eq. (31), with χ_{g0} defined in Eq. (26), depends on c_v. However, when our Eq. (31) [substep (iv)] is combined with the adiabatic gas law [substep (ii)], which depends on the ratio c_p/c_v, the net result is a calculation that depends only on c_p, as in Ref. 14.

IV. RESULTS

The numerical calculations described in Sec. III D were performed for an air bubble in water at 300% supersaturation, which corresponds to $C_\infty/C_0=3$. Results are presented in Fig. 1 for (a) static diffusion in the absence of acoustic excitation, and for sound pressure levels of (b) 205 and (c) 215 dB (re 1 μPa) at 100 Hz. In each case the initial bubble radius was 10 μm. The physical parameters used in the calculations were those for air bubbles in water at 20 °C: $d=0.02$, $\rho=10^3$ kg/m^3, $D=2.4\times10^{-9}$ m^2/s, $\chi_g=2\times10^{-5}$ m^2/s, $\sigma=0.072$ N/m, $\mu=0.001$ kg/ms, and $c_p/c_v=1.4$. The number of samples in the dimensionless coordinate ξ was $n_d=1000$, the number of samples for calculations

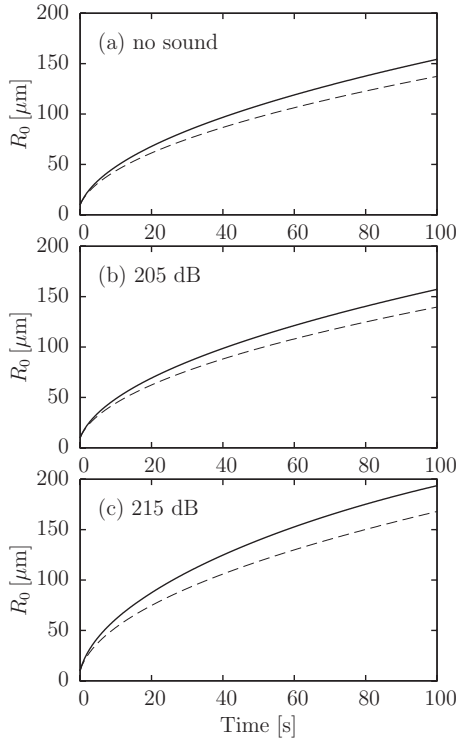


FIG. 1. Dependence of the equilibrium bubble radius R_0 on time for $C_\infty/C_0=3$, (a) in the absence of acoustic excitation and for sound pressure levels of (b) 205 dB and (c) 215 dB (re 1 μPa). Solid lines are results obtained from the present numerical model. Dashed lines are the solutions of Eq. (8), which is Fyrillas and Szeri's model (Ref. 8).

of heat conduction in the gas was $n_g=300$, and the number of time steps per acoustic period was $n_m=2\times10^4$. In each case, results of these numerical calculations, shown with solid lines, are compared with predictions based on Eq. (8) and shown with dashed lines. In the last case, Fig. 1(c), thermal conductivity of the gas in the bubble was taken into account in the same manner as by Crum and Mao.[1] Periodic oscillations of the bubble were calculated using a standard shooting method, with the initial conditions for the bubble radius and its derivative repeated for one period of the external acoustic pressure.

It is clear from Fig. 1 that, at the high level of supersaturation used here, the contribution to bubble growth from rectified diffusion at sound pressure levels of 205 dB and below is negligible in comparison with the contribution from static diffusion. It is also seen that accounting for the finite rate of bubble growth in the present model yields a moderate amount of diffusion enhancement. For static diffusion alone [Fig. 1(a)] the present model predicts about 8% more growth than Eq. (8), while at the highest excitation level [Fig. 1(c)] the present model predicts about 18% more growth than Eq. (8).

The same calculations were repeated for different parameters, including different frequencies, bubble sizes, and the treatment of the gas concentration at infinity. Calculations for 1000 Hz yield approximately the same results but take much longer because the time step must be reduced by one order of magnitude. Likewise, when the initial bubble size is varied, there is little effect on the growth, as long as the excitation is well below the bubble resonance frequency.

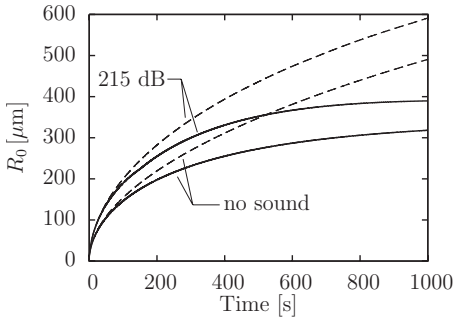


FIG. 2. Dependence of the equilibrium bubble radius R_0 on time with and without sound for washout half times of t_h=360 s (solid lines) and t_h=∞ (dashed lines).

When the supersaturation of the water far away from the bubble is held constant, the bubble radius grows monotonically. In reality the gas concentration in the surrounding liquid approaches exponentially the equilibrium concentration $C_0=P_0/H_D$ corresponding to the external pressure P_0. We can take into account the variation of the concentration due to this relaxation process by, according to Eq. (1), adding the term $-(C-C_0)/t_w$ to the right-hand sides of Eqs. (9) and (15). In this case the gas concentration far away from the bubble will decrease exponentially with time constant t_w, and for durations of order several times t_w the concentration far away from the bubble is nearly at its equilibrium value, and bubble growth is practically stopped. For illustration, shown by solids lines in Fig. 2 are the results of a calculation with t_h=360 s for static diffusion and for a sound pressure level of 215 dB, but otherwise for the same parameters as were used for Fig. 1. Shown for comparison on the same graph are dashed lines for the same parameters but with t_w=∞, i.e., for constant C_∞/C_0=3.

To investigate the temperature dependence of Henry's constant and thermal conductivity of the gas we included Eqs. (21) and (25) in the numerical calculations. It was found that the influence of the temperature dependence of Henry's constant is negligible when the surrounding liquid is at 20 °C. At this equilibrium temperature, changes in the temperature of the liquid near the bubble are very small because of the high thermal conductivity of the liquid in comparison with the thermal conductivity of the gas. There was less than a 0.1% change in bubble growth when Eq. (21) was used. This effect could become more important if the surrounding liquid is at an elevated temperature. For example, the case studied in Ref. 14, which was for an air bubble of radius 0.1 cm at resonance, in water with a temperature of 100 °C, the temperature gradient is more than 30 times larger than the gradient at 20 °C.[14] The effect of radiation losses was studied by substituting Eq. (10) for Eq. (20) in the numerical calculations. For the relatively low acoustic frequencies studied here, 100–1000 Hz, the influence of radiation losses on bubble growth is also negligible.

For all the cases discussed above, the results for bubble growth are higher than what are obtained using models such as Eqs. (7) and (8). Kamath and Prosperetti[17] also observed an increased bubble growth rate using direct numerical solution of the diffusion equation, but they considered only low supersaturation levels near the threshold of bubble growth and did not consider the high supersaturation level used here. A difference between our results and those obtained with models of the type given by Eqs. (7) and (8) is that the latter do not take into account the influence of systematic bubble growth on gas diffusion. The gradient of the dissolved gas concentration $\partial C/\partial r$, and the mass flow of gas into the bubble dm_g/dt according to either Eq. (5) or (7), are estimated for a given equilibrium bubble radius. In reality this gradient and mass flow depend on the growth rate of the equilibrium bubble radius. It can be shown that the correction introduced by the bubble growth rate becomes significant when the characteristic time for bubble growth $t_R=R^2/[Dd(C_\infty/C_0)-1]$ approaches the characteristic time of gas diffusion near the bubble, $t_D=R^2/D$. The ratio $t_D/t_R=d(C_\infty/C_0-1)$ does not depend on D and is of order 0.1 for high levels of the supersaturation C_∞/C_0-1. This in turn means that the growth rate correction may be of order 10% or more. This approximation is indeed borne out, as nearly the same difference in bubble growth appears in Fig. 1 between the predictions of the present model and Eq. (8).

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research.

[1]L. A. Crum and Y. Mao, "Acoustically enhanced bubble growth at low frequencies and its implications for human diver and marine mammal safety," J. Acoust. Soc. Am. **99**, 2898–2907 (1996).
[2]D. S. Houser, R. Howard, and S. Ridgway, "Can diving-induced tissue nitrogen supersaturation increase the chance of acoustically driven bubble growth in marine mammals?," J. Theor. Biol. **213**, 183–195 (2001).
[3]S. H. Ridgway and R. Howard, "Dolphin lung collapse and intramuscular circulation during free diving: evidence from nitrogen washout," Science **206**, 1182–1183 (1979).
[4]L. A. Crum, M. R. Bailey, J. Guan, P. R. Hilmo, S. G. Kargl, T. J. Matula, and O. A. Sapozhnikov, "Monitoring bubble growth in supersaturated blood and tissue ex vivo and the relevance to marine mammal bioeffects," ARLO **6**, 214–220 (2005).
[5]P. S. Epstein and M. S. Plesset, "On the stability of gas bubbles in liquid-gas solutions," J. Chem. Phys. **18**, 1505–1513 (1950).
[6]A. Eller, "Bubble growth by diffusion in an 11-khz sound field," J. Acoust. Soc. Am. **52**, 1447–1449 (1972).
[7]A. Eller and H. G. Flynn, "Rectified diffusion during nonlinear pulsations of cavitation bubbles," J. Acoust. Soc. Am. **37**, 493–503 (1965).
[8]M. M. Fyrillas and A. J. Szeri, "Dissolution or growth of soluble spherical oscillating bubbles," J. Fluid Mech. **277**, 381–407 (1994).
[9]A. Prosperetti and A. Lezzi, "Bubble dynamics in a compressible liquid," J. Fluid Mech. **168**, 457–478 (1986).
[10]A. Prosperetti, "The equation of bubble dynamics in a compressible liquid," Phys. Fluids **30**, 3626–3628 (1987).
[11]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[12]Yu. A. Ilinskii and E. A. Zabolotskaya, "Cooperative radiation and scattering of acoustic waves," J. Acoust. Soc. Am. **92**, 2837–2841 (1992).
[13]A. Lezzi and A. Prosperetti, "Bubble dynamics in a compressible liquid. Part 2. Second-order theory," J. Fluid Mech. **185**, 289–321 (1987).
[14]A. Prosperetti, L. A. Crum, and K. W. Commander, "Nonlinear bubble dynamics," J. Acoust. Soc. Am. **83**, 502–514 (1988).
[15]E. Wilhelm, R. Battino, and R. J. Wilcock, "Low-pressure solubility of gases in liquid water," Chem. Rev. (Washington, D.C.) **77**, 219–262 (1977).
[16]V. Kamath and A. Prosperetti, "Numerical integration methods in gas-bubble dynamics," J. Acoust. Soc. Am. **85**, 1538–1548 (1989).
[17]V. Kamath and A. Prosperetti, "Mass transfer during bubble oscillations," in *Frontiers of Nonlinear Acoustics: Proceedings of the 12th Intl. Symp. on Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier, New York, 1990), pp. 503–508.

Observation of low-frequency acoustic surface waves in the nocturnal boundary layer

Carrick L. Talmadge, Roger Waxler, Xiao Di, and Kenneth E. Gilbert[a)]
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

Sergey Kulichkov
Obukhov Institute of Atmospheric Physics, 3 Pyzhevskii Pereulok, 109017 Moscow, Russia

(Received 18 January 2008; revised 19 June 2008; accepted 22 June 2008)

A natural terrain surface, because of its porosity, can support an acoustic surface wave that is a mechanical analog of the familiar vertically polarized surface wave in AM radio transmission. At frequencies of several hundred hertz, the acoustic surface wave is attenuated over distances of a few hundred meters. At lower frequencies (e.g., below approximately 200 Hz) the attenuation is much less, allowing surface waves to propagate thousands of meters. At night, a low-frequency surface wave is generally present at long ranges even when downward refraction is weak. Thus, surface waves represent a ubiquitous nighttime transmission mode that exists even when other transmission modes are weak or absent. Data from recent nighttime field experiments and theoretical calculations are presented, demonstrating the persistence of the surface wave under different meteorological conditions. The low-frequency surface wave described here is the "quasiharmonical" tail observed previously in nighttime measurements but not identified by S. Kulichkov and his colleagues (Chunchuzov, I. P. *et al.* 1990. "On acoustical impulse propagation in a moving inhomogeneous atmospheric layer," J. Acoust. Soc. Am. **88**, 455–461). © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2967474]

PACS number(s): 43.28.Fp, 43.28.Gq, 43.28.En, 43.28.Js [VEO] Pages: 1956–1962

I. INTRODUCTION

Over the past two decades, one of the authors of this article (Kulichkov) and his collaborators have made a large number of nighttime measurements in various atmospheric and terrain conditions using a propane cannon source (Chunchuzov *et al.*, 1990, 2005). Many of the pressure-versus-time records show a low-frequency "tail" at the end of the received pulse. The tail has been described as having a "quasiharmonical" shape (Chunchuzov *et al.*, 1990) and has been explained by them as the "ground wave" discussed by Cramond and Don (Cramond and Don, 1985; Don and Cramond, 1987). Kulichkov, however, pointed out to the coauthors of the present article that a ground wave does not accurately represent the low-frequency tail, and several years ago, put forth a challenge to explain the origin of the observed tails. The present article is the response to that challenge.

The observed phenomenon has three primary characteristics: it is the last arrival; the higher frequencies are attenuated; and, most importantly, the arrival is present under most nocturnal atmospheric conditions. The last characteristic suggested some fairly universal cause that does not depend on a special or exotic mechanism. After some discussion over a period of time, it was hypothesized that the observed low-frequency tail is an acoustic surface wave due to the porosity of the ground. In Sec. II below, we present experimental evidence to support the hypothesis.

It is useful to make a connection between the low-frequency tail observed by Chunchuzov *et al.* (1990) and the classic acoustic surface wave that exists in a constant sound speed atmosphere. In a downward-refracting atmosphere, when the effective sound speed gradient approaches zero, the lowest order mode in a modal expansion becomes the classic acoustic surface wave that results from the porosity of the ground (Raspet *et al.*, 1992; Waxler, 2002). Hence, throughout this article, we shall refer to the lowest order mode as the "surface wave mode" or simply "the surface wave." The rest of the arrival structure is composed of higher order modes and is referred to as the "refracted waves." A theoretical analysis of the complete arrival structure (surface wave plus refracted waves) shall be reported in a later publication (Waxler *et al.*, 2008). Here, we focus on the surface wave.

The investigation of acoustic surface waves in outdoor sound propagation has a considerable history (Rudnick, 1947; Wenzel, 1974; Donato, 1976; Raspet and Baird, 1989; Attenborough and Chen, 1990; Daigle *et al.*, 1996; Zhu *et al.*, 2002; Albert, 2003, Attenborough *et al.*, 2007). However, prior to the measurements reported here, the only direct observation and identification of a surface wave arrival in outdoor sound experiments has been for snow covered ground (Albert, 2003).

The apparent reason for the lack of experimental evidence for a surface wave arrival over soil is the strong attenuation of the surface wave at frequencies above about 200 Hz (Waxler *et al.*, 2006). At these higher frequencies, the surface wave is strongly attenuated and usually makes no significant contribution beyond a few hundred meters. At distances less than a few hundred meters, where the surface

[a)]Author to whom correspondence should be addressed. Electronic mail: kgilbert@olemiss.edu

 0001-4966/2008/124(4)/1956/7/$23.00

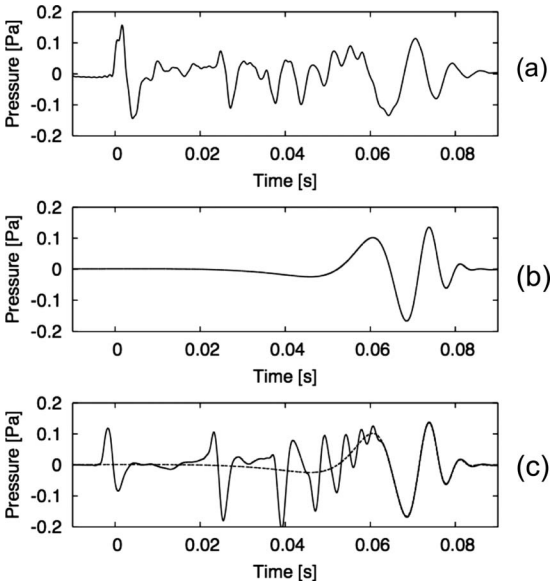


FIG. 1. (a) A sample nighttime pressure-vs-time record showing an example of the "quasiharmonical" (surface wave) tail described in Chunchuzov *et al.*, 1990. The source height, propagation range, and receiver height were 0.3 m, 1.7 km, and 0.625 m, respectively. (b) A theoretical prediction for just the surface wave arrival. (c) A theoretical prediction for the total waveform (solid line) and the surface wave (dashed line). It is clear that the theoretical prediction closely represents the surface wave in the experimental waveform. The computational details of the theoretical prediction are discussed in Sec. IV (Locke Station site).

wave amplitude is substantial, the surface wave arrival is generally not well separated in time from other arrivals and hence cannot be easily identified. In contrast, at frequencies below about 200 Hz, the attenuation of the surface wave can be an order of magnitude less than at higher frequencies, and a propagation distance of thousands of meters is possible (Waxler *et al.*, 2006). Over these distances, the surface wave is often clearly separated from other arrivals and can be easily identified. In this article we present experimental and theoretical evidence to show that a low-frequency surface wave is a common feature of sound propagation at night and can be present even when other modes of propagation are weak or absent.

This paper is organized as follows. Section II presents the surface wave hypothesis and gives evidence supporting its validity. Experimental and theoretical methods are discussed in Secs. III and IV, respectively. Section V compares experimental results with theoretical predictions. A summary and conclusions are given in Sec. VI.

II. SURFACE WAVE HYPOTHESIS

The plausibility of the surface wave hypothesis can be shown by comparing an experimental pressure-versus-time record taken at night in clear weather to a theoretical prediction for sound propagation in a downward-refracting atmosphere. The data and calculation are shown in Fig. 1. Figure 1(a) shows an experimental pressure-versus-time record taken at night. Figure 1(b) shows the theoretical prediction for just the surface wave (first mode) arrival. Figure 1(c) shows the full predicted waveform with the lowest-mode component as a dashed line.

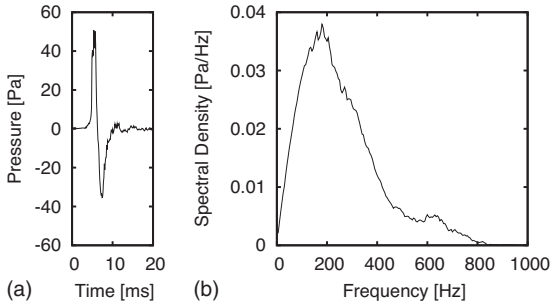


FIG. 2. (a) A pulse shape from the propane cannon used as the sound source. (b) Magnitude of the Fourier transform of the pulse shape. The source and receiver heights were 0.3 m, and the propagation distance was 10 m. Every cannon firing was monitored on a reference microphone.

The experimental source pulse shape and the magnitude of its Fourier transform are shown in Fig. 2. Note that the source has significant Fourier components below 200 Hz. The computed waveform in Fig. 1 uses the modal (eigenfunction) expansion method described in Waxler, 2002. The source waveform in the calculation was the experimental source waveform in Fig. 2. The details of the calculation are discussed in Sec. IV below (Locke Station calculations).

Figure 1 is compelling evidence that the low-frequency tails observed by Kulichkov and his colleagues are, in fact, surface waves. Although many of their measurements show arrival structures similar to that in Fig. 1, the environments of the experiments are often too complex (e.g., due to forests and terrain features) to allow a confident interpretation of the details of the received waveform. In addition, the measurements are for a single microphone. Therefore, it was decided to use a vertical microphone array to pursue a systematic experimental investigation of the surface wave in the nighttime boundary layer under simpler more controlled conditions. In the following we describe some initial findings of that investigation.

III. EXPERIMENTAL METHODS FOR NIGHTTIME MEASUREMENTS

Nighttime measurements are reported for experiments in 2005 and 2006 at farm sites in the Mississippi Delta. The first two experiments were done in 2005, on May 4 and June 23, respectively, at a farm approximately 7 miles north of Marks, MS. The third experiment was done on November 9, 2006 at a farm at Locke Station, MS. Both sites are flat agricultural fields that are free of obstructions and quiet at night. Hereafter, we shall denote the Marks site as "site M" and the Locke Station site as "site L."

The basic geometry of the experiments was the same at the two sites and is shown in Fig. 3. At both sites, acoustic measurements were made on a vertical array of microphones using a propane cannon as the source. As shall be discussed later, the measurements at site M were made under weak downward-refracting conditions and the measurements at site L were made under strong downward-refracting conditions. The primary differences in the experimental setups were the number, location, and types of sensors on the towers (e.g., microphones and meteorological sensors). In addi-

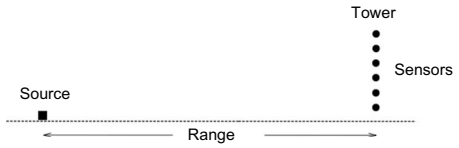


FIG. 3. Experimental setup for nighttime field measurements at sites M and L. The tower was instrumented with microphones and meteorological sensors. The sound source was a propane cannon. The tower instrumentation at site L was more complete than at site M.

tion, the distances of the sources from the tower were different in the three experiments. The details of these differences are discussed below.

The experiments made in 2005 at site M were preliminary measurements used for planning the more complete experiment that was conducted later at site L. Consequently, only limited meteorological measurements were made at site M. Nevertheless, we include the site M measurements here because the gross atmospheric conditions for site M on both dates were very different from those reported at site L and resulted in significantly different acoustic propagation. Since the measurements made in 2006 at site L were more complete, we shall discuss the site L experiment first.

A. Locke Station (site L) experiment, November 9, 2006

The ground surface at Locke Station was an agricultural field that had been smoothed by weathering. There were no crops and little vegetation in the field. On the night of November 9, 2006, the sky was clear with a wind coming out of the southwest. Because of the clear sky, the radiation of ground heat into space (blackbody radiation) was strong. Thus, the ground cooled significantly and a stable nocturnal boundary layer formed.

The 10 m tower at site L (Fig. 3) was instrumented with 16 calibrated Brüel and Kjaer type 4191 microphones spaced uniformly, starting at 0.625 m above the ground. In addition to the microphones, the tower was instrumented with two Campbell Scientific CSAT3 sonic anemometers/thermometers and three Hobo meteorological stations for measuring the temperature and vector wind as function of height. Above 10 m, the temperature and vector wind were measured to a height of 150 m with a Vaisala DigiCORA tethersonde. Above 150 m, the temperature was extrapolated linearly upward and the vector wind was measured with a Remtech PA1 sodar. With these meteorological sensors, the mean temperature and mean vector wind were measured from the ground to a height of 450 m. The strong ground cooling due to radiation of thermal energy into space is evident in the temperature profile in Fig. 4.

The sound source was a propane cannon (C. Frensch Ltd., Beamsville, ON, Canada) at a height of 0.3 m and 1.7 km directly south of the tower. The pressure-versus-time source signature of the cannon was approximately the same as that shown in Fig. 2.

The mean temperature profile and the component of the mean horizontal wind in the propagation direction for run 10 (7:37 pm) are shown in Figs. 4(a) and 4(b), respectively. Mean values for the temperature and vector wind were computed from 15 min averages of the sensor outputs. Continuous mean profiles were then constructed by passing smooth curves through the 15 min averages. The acoustic measurements were made near the middle of the 15 min interval used for the mean temperature and wind profiles. Near the ground, there was a stable layer in which the wind speed was low. At higher altitudes, the wind increased significantly, and there were local maxima in the wind speed. We shall refer to the lowest maxima, near the top of the stable layer, as a "low-level nocturnal jet" (Stull, 1993, pp. 520–526).

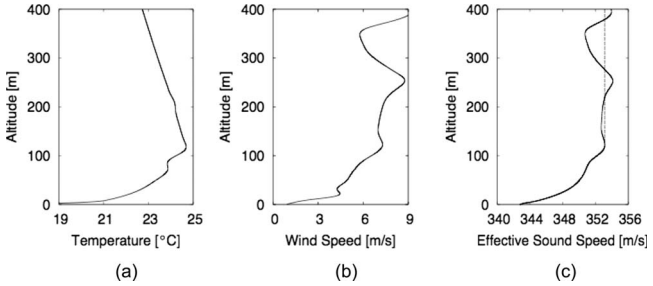


FIG. 4. Mean profiles at site L on run 10 (7:37 pm): (a) temperature profile, (b) horizontal wind component in the propagation direction, and (c) effective sound speed profile determined from (a) and (b). In theoretical calculations, the measured profile was used up to 125 m. Above 125 m, the effective sound speed was taken to be constant, as indicated in (c) by the dashed vertical line starting at 125 m.

The temperature and wind profiles were combined to determine the effective sound speed versus height shown in Fig. 4(c). It is clear from Fig. 4(c) that the effective sound speed profile is strongly downward refracting. Such conditions are typical for downwind sound propagation on clear nights, where there is substantial radiation of heat from the ground into space (Stull, 1993, pp. 506–508).

B. Marks (site M) experiments

The ground surface at site M at the time of the first experiment (May 4, 2005) was a flat plowed agricultural field that had been planted. Weathering had considerably reduced the furrows so that the surface relief was less than about 5 cm. On the night of the second experiment (June 23, 2005), approximately seven weeks later, the surface was smoother, and the planted crop had emerged. The plants were about 10 cm high.

At site M there was a heavy cloud cover on the night of the first experiment. On the night of the second experiment, there were no clouds, but there was a ground-level haze layer. On both nights there was little wind near the ground. Because of the cloud cover on the first night and the haze layer on the second night, the ground did not readily radiate energy into space and hence did not cool as rapidly as it does on a clear night (e.g., as at site L). As a result, the strong temperature inversion (increase in temperature with height) observed during the site L experiment did not occur at site M on either of the two nights. Under such conditions one expects an approximately "neutral" atmosphere near the ground where the temperature gradient is negative and weak, e.g., a decrease of 3–10 °C/km (Arya, 2001, pp. 64–68). Because of the decrease in temperature with height and light winds,

one would expect little or no downward refraction of sound to be detected at ranges of less than a few kilometers.

1. May 4, 2005

The tower used for the first experiment at site M had eight calibrated Brüel and Kjaer type 4191 microphones at heights of 0.20, 0.41, 1.25, 1.9, 2.7, 4.6, and 10.0 m. There were no temperature or wind speed sensors on the tower. Temperature measurements were made with a handheld thermometer on the ground and on a 2 m ladder. The measurements indicated a nearly constant temperature near the ground. The measurements were too crude to detect the weak temperature gradient (0.003–0.01 °C/m) that is typical of a nearly neutral atmosphere. No wind speed measurements were made. However, it was evident that there was very little wind near the ground. The source was a propane cannon of the type used at site L. The propane cannon was fired at a distance of 0.9 km from the tower and a height of 0.3 m.

2. June 23, 2005

The tower used at site M on June 23, 2005 had ten calibrated Brüel and Kjaer type 4191 microphones at heights of 0.0, 0.7, 1.4, 2.1, 2.8, 3.5, 4.2, 4.9, 5.6, and 7.7 m. Sonic anemometers/thermometers (Campbell Scientific CSAT3) were at heights of 1.5 and 7.7 m. The lower anemometer/thermometer malfunctioned, unfortunately, so that temperature and wind speed gradients could not be measured. However, the upper anemometer/thermometer indicated winds of less than 1 m/s and a temperature that decreased by less than 1 °C during the experiment. Thus, one would expect that little ground cooling took place. The propane cannon was fired at a distance of 2 km from the tower and a height of 0.3 m.

IV. THEORETICAL AND COMPUTATONAL METHODS

In Sec. V below, the waveforms measured on the tower at sites L and M are compared to theoretical predictions. For site L, the theoretical waveforms were computed using the eigenfunction expansion method described in Waxler, 2002. For site M, a parabolic equation method (Gilbert and White, 1989; West *et al.*, 1992) was used. At both sites, the measured source signature and spectrum were used in the calculations. In the site L calculations, the measured effective sound speed was approximated with a sound speed profile that matched the measured profile up to 125 m and was constant thereafter, as shown in Fig. 4. At site M, where the temperature and wind speed were not measured, meteorological profiles typical of a neutral atmosphere were assumed in the theoretical predictions.

With strong downward refraction, as at site L, only the discrete part of the horizontal wavenumber spectrum (the "ducted modes") must be included in the calculation. For such conditions, an eigenfunction expansion is efficient and accurate. In contrast, when downward refraction is weak (as at site M), the continuous part of the horizontal wavenumber spectrum can contribute significantly. For these conditions, an eigenfunction expansion is less efficient because the eigenfunctions of the continuous spectrum must be computed and included for an accurate eigenfunction expansion. Since the parabolic equation method automatically includes the continuous spectrum, it is more convenient for predicting propagation in an atmosphere with weak downward refraction. Thus, the theoretical waveforms for site M were computed using a parabolic equation.

In the initial calculations for site L, a locally reacting ground impedance model was used that had nearly equal real and imaginary parts and a magnitude that varied inversely with the square root of the frequency. Such a model is consistent with most standard ground impedance models (see, e.g., Attenborough *et al.*, 2007). With this model, the predicted surface wave arrival came in much too early relative to the refracted arrivals. In order to increase the time separation of the surface wave and refracted waves, it was necessary for the imaginary part of the impedance to be four to five times greater than the real part. That is, the phase of the impedance had to be significantly greater than 45°.

To facilitate varying the phase of the impedance independently from the magnitude of the impedance, the normalized ground impedance was assumed to be of the form $\hat{Z} = |\hat{Z}_0| \exp(i\phi)\sqrt{f_0/f}$, where $|\hat{Z}_0|$ is the magnitude of the normalized impedance at the frequency f_0 and ϕ is the phase of the impedance. Although this impedance parametrization is consistent with standard ground impedance models, it is purely empirical, and no claim is made for its theoretical validity. The parametrization used is convenient for the present investigation, however, because the magnitude of the impedance controls the overall pressure levels and the phase controls the arrival time of the surface wave relative to the refracted waves. For $f_0 = 100$ Hz, the fitted value for the magnitude of the normalized impedance was $|\hat{Z}_0| = 26$, which is large but not unphysical. The phase which yielded the correct relative arrival time of the surface wave was found to be $\phi = 77.8°$ so that the imaginary component of the impedance was 4.6 times the real component. This value of ϕ is consistent with the large phase angle predicted by a ground model with exponentially decreasing porosity (Raspet and Sabatier, 1996). The exponential porosity model approximates a porous layer overlying an acoustically hard layer, as one might expect for an agricultural field. At frequencies below a few hundred hertz, impedance models for a homogenous ground surface (i.e., not layered) generally do not have an imaginary component that is four to five times larger than the real component (Attenborough *et al.*, 2007). Layering is apparently needed to produce such large phase angles in the complex impedance.

The theoretical time domain waveforms at the tower locations were created by computing the complex pressure as a function of frequency and using Fourier synthesis. It was found that a frequency spacing of 5 Hz in the frequency domain was adequate for an accurate numerical Fourier synthesis of the time domain signal. The source height was taken to be 0.3 m, and the received pressure waveforms were computed at the heights of the tower microphones.

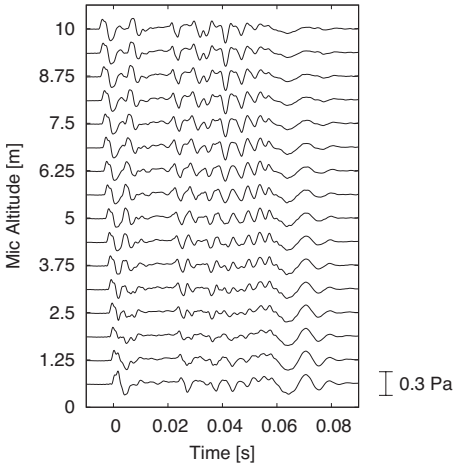


FIG. 5. Measurements made during run 10 (7:47 pm) on a 10 m tower at site L. The source was located 1.7 km south of the tower and 0.3 m above the ground. There were 16 microphones on the tower, starting at 0.625 m and spaced uniformly up to 10 m. The surface wave arrival is evident as the last arrival. It is strongest on the bottom microphone and decreases monotonically with increasing height. Note also that the surface wave arrival is dispersive with the higher frequencies arriving last and with lower amplitude. The arrival structure in the figure was stable over a 2 min run (60 cannon shots) but often was stretched or compressed on a run made 15–20 min later.

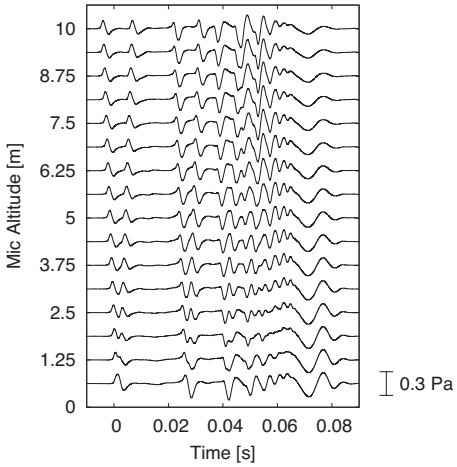


FIG. 6. Theoretical prediction to be compared with the measurements shown in Fig. 5. Although the theoretical prediction does not match all the details of the measurements, it clearly captures the main features. For example, the variation with height of the surface wave and refracted wave amplitudes is accurately represented. Also, as in the measurements, the predicted surface wave arrival is dispersive with the higher frequencies arriving last and with lower amplitude.

V. COMPARISON OF NIGHTTIME MEASUREMENTS AND THEORETICAL PREDICTIONS

A. Locke Station (site L) experiment, November 9, 2006

Figure 5 shows pressure-versus-time records collected on the 16-channel vertical microphone array at site L on November 9, 2006. The surface wave tail is evident as the last arrival in each record, especially near the ground. The front part of the surface wave arrival (0.05–0.06 s) overlaps with the refracted arrivals but the last part does not. The surface wave is clearly dispersive with the lower-frequency components arriving before the higher-frequency components and with less attenuation. While the refracted arrivals are relatively constant or increase with increasing height, the surface wave is strongest on the bottom microphone and decreases monotonically with increasing height.

Although the surface wave is the focus here, it is instructive to briefly discuss the refracted arrival structure as well. The first arrival in Fig. 5 is the direct arrival and it is immediately followed by its ground bounce. This interpretation is clear due to the increasing separation of these two arrivals with increasing height on the tower. While the surface wave is a single mode, the direct arrival and its ground bounce are composed of a large number of modes. Between the first arrival (direct wave) and last arrival (surface wave), the number of modes in the arrivals systematically decreases from a maximum of 36 modes to 1 mode. A complete theoretical analysis of both the refracted arrivals and the surface wave has been carried out and shall be reported in a subsequent article (Waxler *et al.*, 2008).

Over a 2 min run (60 cannon shots) the arrival structure shown in Fig. 5 was surprisingly stable. On a subsequent run made 15–20 min later, the arrival structure was often a "stretched" or "compressed" version of the structure shown in Fig. 5. That is, the basic arrival structure did not change greatly, just the overall temporal width of the pulse. The stretching and compressing of the total arrival are believed to be due to the stability of the near-ground boundary layer and the variability of the low-level nocturnal jet near the top of the stable boundary layer. The surface wave and lower order modes are late arrivals and propagate mainly in the stable boundary layer, which has a lower average effective sound speed than exists at higher altitudes. Due to the near-ground stability of the boundary layer, the late arrivals are hypothesized to have fairly constant absolute travel times. The higher order modes, which make up the early arrivals, extend vertically beyond the stable layer into the low-level nocturnal jet and are hypothesized to have more variable absolute travel times. The sensing of different parts of the atmosphere, which have different stability and different average effective sound speeds, creates a total arrival structure whose early parts are variable and whose later parts are stable. Thus the entire pulse structure stretches and compresses as the low-level nocturnal jet varies. At this time, this explanation is only a hypothesis because the Global Positioning System (GPS) clock at the source malfunctioned so that absolute travel times could not be measured. Future investigations will measure the absolute travel times and attempt to correlate the stretching and compressing of the total arrival structure with direct measurements of the variability of the low-level nocturnal jet.

The computed pressure-versus-time records corresponding to the measurements in Fig. 5 are shown in Fig. 6. The effective sound speed profile in Fig. 4 and the ground impedance discussed in Sec. IV were used in the theoretical calculations. As shown by the dashed line in Fig. 4, the effective sound speed was taken to be constant above 125 m. The computed pressure-versus-time records do not precisely match all of the features in the measurements, but they nevertheless closely represent the main features in the measurements, especially those of the surface wave.

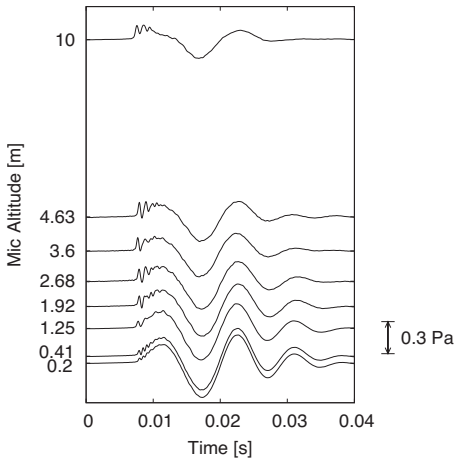


FIG. 7. Measurements during the first nighttime experiment at site M (May 4, 2005) on a 10 m tower under cloudy conditions. The propagation distance was 0.9 km and the source height was 0.3 m. Because of the cloud cover, ground cooling was weak. As a result of the weak ground cooling and calm winds, downward refraction was weak, and the refracted wave arrivals were weak. The refracted arrivals are discernible but are considerably weaker than the surface wave. Under these atmospheric conditions, the surface wave arrival carries most of the acoustic energy.

B. Marks (site M) experiments

Figures 7 and 8 show tower data collected at night at site M on May 4, 2005 and June 23, 2005, respectively. As discussed earlier, because of the cloud cover on the first night and the haze layer on the second night, ground cooling was weak on both nights. Also, it was calm near the ground. As a result, downward refraction was weak and the refracted wave arrivals were consequently weak.

In Fig. 7, the refracted arrivals from a propane cannon 0.9 km away, although weak, are easily discernible. The surface wave arrival clearly carries most of the acoustic energy.

The measurements shown in Fig. 8 were made on the second date, June 23, 2005, at site M on a 7.7 m tower. The refracted arrivals are stronger than in Fig. 7, even though the propane cannon is 2 km away. At the top of the array, the

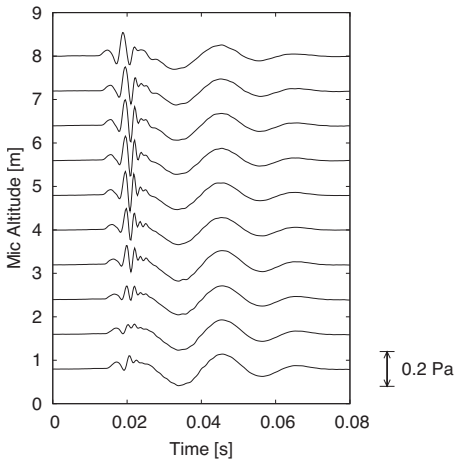


FIG. 8. Measurements made during the second nighttime experiment at site M (June 23, 2005) on a 7.7 m tower in a ground-level haze layer. The propagation distance was 2 km and the source height was 0.3 m. Although the propagation distance was over twice that in Fig. 7, the refracted arrivals are considerably stronger. Nevertheless, near the ground, the surface wave is the dominant arrival.

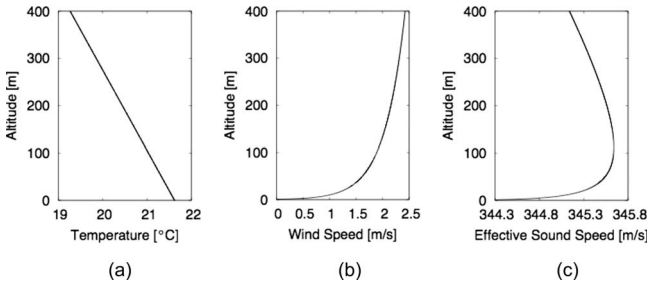


FIG. 9. (a) Assumed temperature profile for a neutral atmosphere. The temperature gradient used to compute the adiabatic sound speed was taken to be a moist adiabatic lapse rate of 6 °C/km. (b) Assumed logarithmic wind speed for a nearly calm night. (c) Effective sound speed for downwind propagation using (a) and (b). Note that the sound duct is very weak, less than 1.5 m/s.

refracted arrivals have considerable amplitude, but near the ground, the surface wave is still the dominant arrival.

As discussed in Sec. III, adequate meteorological measurements were not made at site M. Hence, to make theoretical predictions, it was necessary to use assumed values for the meteorological profiles. For both site M experiments, the atmosphere was taken to be neutral and to have weak winds. A moist lapse rate of 6 °C/km was assumed for the temperature profile. The horizontal wind speed was assumed to be logarithmic and scaled so that the magnitude was 2 m/s at 150 m. The profiles for these conditions are shown in Fig. 9. The propagation was assumed to be directly downwind so that the effective sound speed is given by the sum of the adiabatic sound speed and the horizontal wind speed. The ground impedance was the same as that discussed in Sec. IV. The source is shown in Fig. 2.

Figure 10 shows the predicted arrivals on a 10 m tower with 16 microphones for an atmosphere with the profiles in Fig. 9. The calculation is qualitatively the same as the measurements at site M on both dates. For example, the refracted arrivals are weak near the ground and grow stronger with

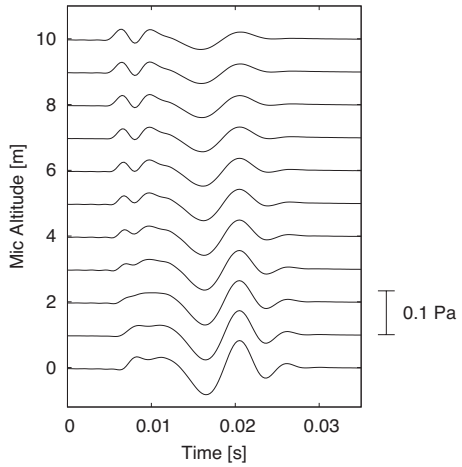


FIG. 10. Theoretical prediction for the effective sound speed in Fig. 9. The source is taken to be 1 km from the tower and 0.3 m above the ground. The source signature is given in Fig. 2, and the ground impedance is given in Sec. IV. The predicted weak refracted arrivals and strong surface wave arrivals are similar to the measurements at site M on both nights. As in the measurements, the predicted refracted arrivals are weak near the ground and increase with height. The reverse is true for the surface wave arrival.

altitude, while the surface wave decays with increasing altitude. As in the data, the calculation shows that, near the ground, most of the energy is carried by the surface wave.

The measurements in Figs. 7 and 8 and the calculations in Fig. 10 show that for downwind propagation, with a nearly neutral atmosphere and weak winds, one expects the surface wave to be the dominant arrival near the ground.

VI. SUMMARY AND CONCLUSIONS

The "quasiharmonical" tail observed by Kulichkov and his collaborators in nighttime measurements has been identified as a low-frequency acoustic surface wave. Data from recent nighttime experiments in agricultural fields and results from theoretical predictions show the existence and persistence of the acoustic surface wave under different nighttime meteorological conditions. The experimental measurements and theoretical calculations both strongly suggest that, at night, a low-frequency source will usually excite a surface wave that propagates to long range. Thus, a surface wave is a persistent nighttime transmission mode for long-range, low-frequency sound and can exist even when refracted arrivals are weak or absent.

The data and analyses presented here are for propagation over agricultural fields at night. It is worthwhile to ask what might happen for other ground surfaces and other meteorological conditions. First, the nighttime data in Chunchuzov *et al.* (1990, 2005) show surface wave arrivals for a variety of ground surfaces. Hence, one would expect that a nighttime surface wave is not limited to agricultural fields but can be present for many ground surfaces. Second, if the ground surface is porous, a strong surface wave arrival is possible in any atmosphere that is downward refracting. For example, under cloudy daytime conditions, a near-surface acoustic duct can exist for downwind propagation. For such atmospheric conditions and a porous ground surface, one would expect a significant low-frequency surface wave component in the long-range sound field. Thus, although the surface wave observations reported here are for specific conditions, the results suggest that a low-frequency surface wave is a common feature for sound propagation in any downward-refracting atmosphere over porous ground.

ACKNOWLEDGMENT

The authors gratefully acknowledge funding from the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.

Albert, D. G. (**2003**). "Observations of acoustic surface waves in outdoor sound propagation," J. Acoust. Soc. Am. **113**, 2495–2500.

Arya, S. P. (**2001**). *Introduction to Micrometeorology*, 2nd Ed. (Academic, New York).

Attenborough, K., and Chen, Y. (**1990**). "Surface waves at an interface between air and an air-filled poroelastic ground," J. Acoust. Soc. Am. **87**, 1011–1016.

Attenborough, K., Li, K. M., and Kirill, H. (**2007**). *Predicting Outdoor Sound* (Taylor & Francis, New York).

Chunchuzov, I. P., Bush, G. A., and Kulichkov, S. N. (**1990**). "On acoustical impulse propagation in a moving inhomogeneous atmospheric layer," J. Acoust. Soc. Am. **88**, 455–461.

Chunchuzov, I. P., Kulichkov, S. N., Otrezov, A., and Perepelkin, V. (**2005**). "Acoustic pulse propagation through a fluctuating stably stratified atmospheric boundary layer," J. Acoust. Soc. Am. **117**, 1868–1879.

Cramond, A. J., and Don, C. G. (**1985**). "Impulse propagation over grassland in the presence of temperature and wind gradients," J. Acoust. Soc. Am. **78**, 1029–1037.

Daigle, G. A., Stinson, M. R., and Havelock, D. I. (**1996**). "Experiments on surface waves over a model impedance plane using acoustical pulses," J. Acoust. Soc. Am. **99**, 1993–2005.

Don, C. G., and Cramond, A. J. (**1987**). "Impulse propagation in a neutral atmosphere," J. Acoust. Soc. Am. **81**, 1341–1349.

Donato, R. J. (**1976**). "Propagation of a spherical wave near a plane boundary with a complex impedance," J. Acoust. Soc. Am. **60**, 34–39.

Gilbert, K. E., and White, M. J. (**1989**). "Application of the parabolic equation to sound propagation in a refracting atmosphere," J. Acoust. Soc. Am. **85**, 630–637.

Raspet, R., and Baird, G. E. (**1989**). "The acoustic surface wave above a complex impedance ground surface," J. Acoust. Soc. Am. **85**, 638–640.

Raspet, R., Baird, G. E., and Wu, W. (**1992**). "Normal mode solution for low-frequency sound propagation in a downward refracting atmosphere above a complex impedance plane," J. Acoust. Soc. Am. **91**, 1341–1352.

Raspet, R., and Sabatier, J. M. (**1996**). "The surface impedance of grounds with exponential porosity," J. Acoust. Soc. Am. **99**, 147–152.

Rudnick, I. (**1947**). "Propagation of an acoustic wave along a boundary," J. Acoust. Soc. Am. **19**, 348–356.

Stull, R. B. (**1993**). *An Introduction to Boundary Layer Meteorology* (Kluwer Academic, Boston).

Waxler, R. (**2002**). "A vertical eigenfunction expansion for the propagation of sound in a downward-refracting atmosphere over a complex impedance plane," J. Acoust. Soc. Am. **112**, 2540–2552.

Waxler, R., Gilbert, K. E., and Talmadge, C. L. (**2008**). "A theoretical treatment of the long range propagation of impulsive signals under strongly ducted nocturnal conditions," J. Acoust. Soc. Am., to be published.

Waxler, R., Talmadge, C. L., Dravida, S., and Gilbert, K. E. (**2006**). "The near-ground structure of the nocturnal sound field," J. Acoust. Soc. Am. **119**, 86–95.

Wenzel, A. R. (**1974**). "Propagation of waves along an impedance boundary," J. Acoust. Soc. Am. **55**, 956–963.

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

Zhu, W., Stinson, M. R., and Daigle, G. A. (**2002**). "Scattering from impedance gratings and surface wave formation," J. Acoust. Soc. Am. **111**, 1996–2012.

Horizontal structure of acoustic intensity fluctuations in the ocean

B. J. Uscinski
Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom

J. R. S. Nicholson
Platform Systems, QinetiQ Ltd., Building A22, Winfrith Technology Centre, Dorchester, Dorset DT2 8XJ, United Kingdom

(Received 16 October 2007; revised 17 June 2008; accepted 24 June 2008)

In April 1989 an acoustic experiment was carried out over the abyssal plain south of Madeira in which transmissions were made, for about an hour, at 482, 680, and 740 Hz from a ship steaming at 5 Kn to a receiving array towed by another ship 65 km away traveling on a parallel course at the same speed. The signals arrived by two paths, an upper path trapped in the surface duct and a lower path via the main sound channel. This paper describes the experiment and analyses of the intensity fluctuations in the signal received by the lower path. This allows the authors to investigate the horizontal structure of acoustic intensity fluctuations in the ocean when these are due principally to internal waves. This aspect of such acoustic intensity fluctuations has received little attention until now. The experimental results are compared with theoretical predictions based on the parabolic moment equations for propagation and scattering in randomly irregular media, and on the standard Garrett–Munk model for ocean internal waves. The experimental results and theoretical predictions agree quite well but the comparison also raises some new questions, in particular, about the correlation of intensity fluctuations as the acoustic transmission frequency is varied.
 [DOI: 10.1121/1.2967485]

PACS number(s): 43.30.Ft, 43.20.Bi, 43.20.Fn, 43.30.Re [JJM] Pages: 1963–1973

I. INTRODUCTION

The effect of ocean internal wave fields on acoustic intensity fluctuations has received wide attention. Apart from theoretical studies[1–3] there have been several propagation sea trials providing experimental information about these intensity fluctuations. In addition to making measurements of their time variation at a single fixed receiver[4–6] their spatial structure in the vertical has been recorded as a function of time[7,8] and also of distance in the direction of propagation.[9] One aspect of their behavior that has not been studied experimentally until now is their horizontal spatial structure transverse to the direction of propagation. This paper describes an experiment that was designed to provide such information, presents the theory necessary to analyze the results, and compares theory and experiment.

The propagation experiments carried out at COBB Seamount in 1971 (Ref. 4) and 1977 (Ref. 5) involved transmissions from a source at a depth of 1000 m on a seamount off the coast of Washington, USA, to a receiver on another seamount about 18.1 km distant. The carrier frequencies used were 2, 4, 8, and 13 kHz and the resulting intensity time series allowed detailed comparisons with the predictions of current acoustic scattering theory. The inadequacies of the then popular Rytov approximation in the case of multiple scattering became clear and an approach based on the fourth moment parabolic equation[10] was adopted. In addition, it was shown that a standard internal wave model with an admixture of "fossil turbulence," i.e., the results of very slowly decaying turbulence, gave satisfactory agreement between theory and experiment.[11,12]

Experiments to measure the vertical structure of intensity fluctuations met with mixed success. NAPOLI 85 (Ref. 8) involved impulse transmissions from a fixed source to a vertical hydrophone array situated at a range of 5 km. The trial was carried out in the Tyrrhenian Sea in the summer of 1985 over a period of 5 days, but there was only a very weak internal wave field. However, the accompanying oceanographic measurements[11] allowed the structure of ocean irregularities to be characterized and so a link could be made to the observed intensity patterns. Similarly the AIWEX trial,[7] which measured the vertical structure of a sound field transmitted under ice in the Beaufort Sea, revealed very weak intensity fluctuations. This was due to the fact that internal wave activity is much suppressed under the permanent Arctic ice cover.

BODRUM 86 (Ref. 9) attempted to study the structure of intensity fluctuations in the vertical but as a function of distance from the source in the direction of acoustic propagation. A 3.5 kHz pulsed source was moored at a depth of 500 m in a region of the Eastern Mediterranean Sea south of the Turkish port of Antalya. A vertical hydrophone chain was lowered and raised in a yo-yo pattern as the research vessel drifted away from the source. The resulting experimental data allowed the two-dimensional acoustic intensity pattern to be reconstructed. A striking result of this experiment was the emergence of ribbons of high intensity extending a long

way in the direction of acoustic propagation and also of regions of very low, almost zero intensity, extending for considerable distances in the same direction.

The above trials confirmed many predictions of the theory of acoustic scattering by ocean irregularities, particularly internal waves. The ribbonlike spatial structure observed in numerical simulation studies[12–14] was detected and the dimensions of the ribbons and their separation were estimated. All these results were in reasonable agreement with theory and simulation. However, one aspect of the ribbons had still not been studied experimentally. This was the structure of intensity fluctuations in the horizontal direction transverse to the direction of acoustic propagation. In 1989 a transmission trial was carried out with the aim of providing such experimental data and so completing the picture of the spatial structure of acoustic intensity fluctuations in the ocean. This paper describes the trial, presents the appropriate acoustic propagation theory, and makes some comparisons between theory and experiment.

Finally, we should consider the significance of these experiments. The spatial structure of acoustic intensity fluctuations is not just of academic interest. Ocean acoustic transmissions provided the first experimental data on intensity fluctuations that could be compared with a carefully measured scattering medium thus allowing multiple scattering theory to be verified. In the case of optical transmissions the scattering medium could not be reliably characterized and so such comparisons were impossible. In the event it was the COBB and MATE results that showed the inadequacies of the then popular Rytov approximation and set the scene for the adoption of the fourth moment parabolic equation. Of equal importance is the fact that a good understanding of the spatial ribbonlike structure of acoustic intensity fluctuations enables us to make better predictions of the performance of acoustic arrays in both their active and passive roles. An understanding of this structure also allows us to optimize strategies for detecting underwater objects using active sonar, and, on the other hand, may help a target avoid detection.

II. THE EXPERIMENT

A. Experimental arrangement

The experiment reported here was carried out in April of 1989 in the sea area south of Madeira, over the Madeira abyssal plain.

During the week leading up to the measurements, heavy seas of between Sea States 5 and 6 had dominated the weather pattern. Upper layer mixing combined with the increase in sound speed due to hydrostatic pressure resulted in an acoustic surface duct extending to a depth of about 200 m. During the experiment, good weather prevailed with Sea State varying between 1 and 2. Two vessels made parallel courses at a separation of 65 km, the empirically determined separation of the first convergence zone. One vessel towed an omnidirectional sound projector while the other towed a linear array of hydrophones to receive the transmitted signal (Fig. 1). The sound projector was maintained in a broadside direction to the receiving array. From depth sensor

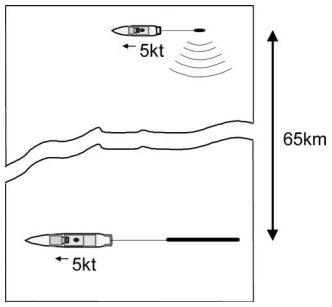


FIG. 1. Disposition of the "transmit" and "receive" vessels.

measurements, the towed array was known to be horizontal. The transmitter depth was approximately 80 m and that of the receiving array was about 100 m. With the vessels maintaining a steady broadside parallel course at 5 Kn under autopilot in light seas, the measured depth variation of the transmitter and receiver was typically 1 m over a period of 1 h.

Figure 2 shows a typical sound speed profile for the experiment.

For the purpose of ray tracing, the extended sound speed profile of Fig. 2 has been used. A selection of ray traces from this profile is shown in Fig. 3.

A recognized feature of deep oceanic sound transmission is the formation of regularly spaced acoustic convergence zones, which are regions of crudely focused sound formed some distance from a transmitter. They are a consequence of deterministic focusing brought about by the bulk sound speed profile, which is dominated by hydrostatic pressure at greater depths. Near the surface, the center of the first convergence zone occurs at a range of approximately 65 km from the transmitter for the extended sound speed profile of Fig. 2. Using this profile, it has been determined from ray tracing that rays with launch angles between −3.5° and −9.5° in elevation, with respect to the horizontal, emerge at the convergence zone centered some 65 km from the source, without surface or bottom interaction. In this manner, the

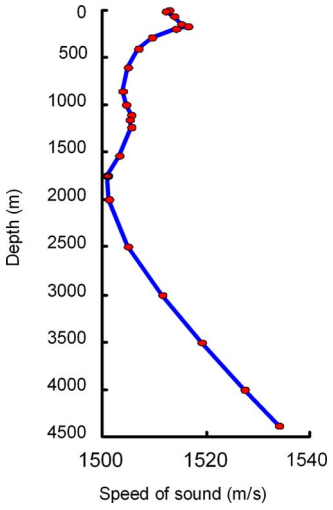


FIG. 2. (Color online) Sound speed profile derived from the mean XBT data to 1200 m and extended to 4400 m with climatological data for the location and time of year.

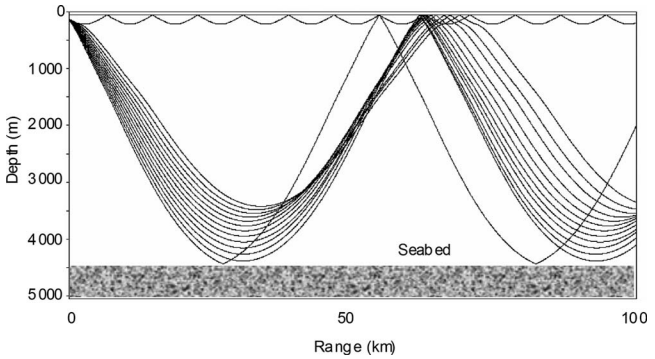


FIG. 3. Ray traces calculated using the extended sound speed profile of Fig. 2.

wholly refracted path from a continuous transmission can be isolated at a distant receiver. Figure 3 depicts a ray trace for launch angles between −3° and 10° at 0.5° spacing. In this case, the limiting rays at −3° and −10° originating from a source at 80 m depth either enter the surface duct or become bottom reflected, respectively. From ray tracing, it was also determined that the typical flight time for a ray launched at −4°, for example, was 43.34 s.

Continuous transmissions were made at the following frequencies: 482, 680, and 740 Hz. These frequencies were chosen to lie in relatively quiet regions of own-ship's noise spectrum to improve signal/noise at the receiver.

B. Acoustic measurements

The receiving array consisted of 25 hydrophones sparsely distributed over a distance of 684 m. One hydrophone was not functioning and was omitted from the analysis. Variations in sensitivity between hydrophone channels were removed by scaling the signal using a long term average of the received signal magnitude. To give an overall picture of the results we now show some of the acoustic measurements for the convergence zone propagation path. Results for rays trapped in the upper layer are not discussed here. Constraints of space prevent us from showing all the acoustic measurements. Some typical values of acoustic intensity down the receiving array as a function of time are shown in Fig. 4 for different transmission frequencies. The intensity variations at a specific frequency have been normalized to the mean intensity at that frequency.

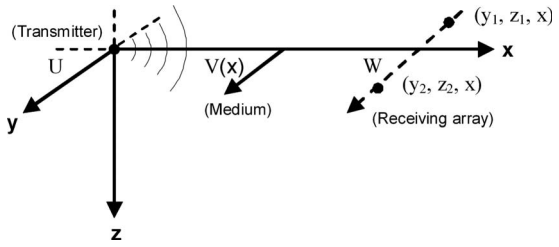


FIG. 5. Coordinate system with moving source, medium, and receivers.

The first and most obvious feature of these records is their banded structure, which is related to the horizontal high and low intensity ribbons caused by acoustic scattering in the internal wave field encountered by the convergence zone path. Also striking is the slope of the bands, which is related to the motion of source and receiver, V. In order to discuss and analyze the acoustic results it will be necessary to consider the appropriate propagation theory for acoustic scattering by internal waves, which is presented below.

III. THEORY

A. Fundamentals

The theory for intensity fluctuations arising from a point source situated in a moving medium, when both source and receiver move, has been treated elsewhere.[15] The principal results are summarized here and applied to the particular case of a medium containing a standard ocean internal wave field. The theory has been derived for a source situated at the origin of a Cartesian set of axes (x,y,z), Fig. 5. The source moves with velocity U in the y direction, the medium moves

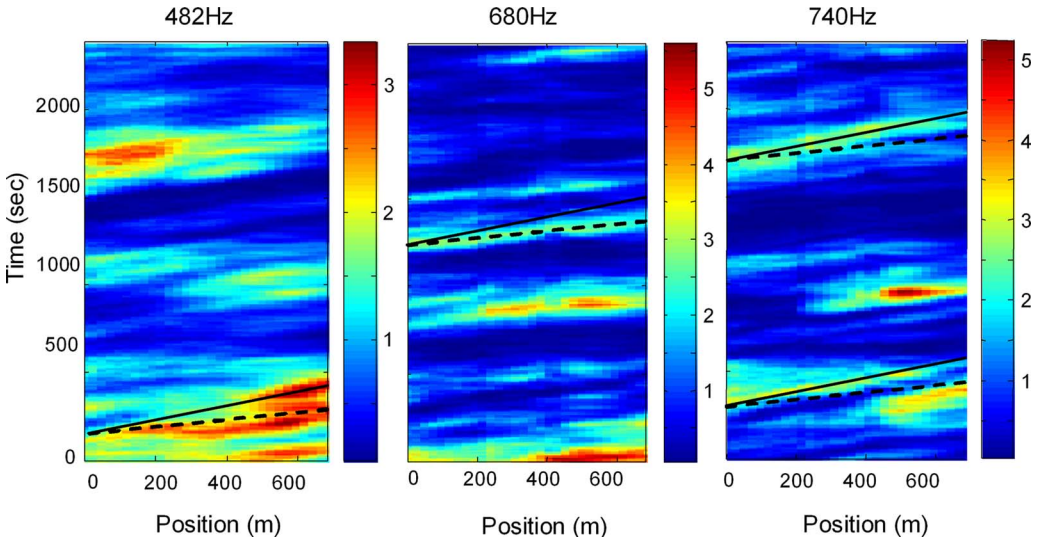


FIG. 4. (Color online) Signal intensity observed at the towed array as a function of position along the array and time for simultaneous transmissions at 482, 680, and 740 Hz. The solid lines indicate trajectory of motion at the towing speed (V=2.5 m/s). The dashed lines indicate trajectory of motion at twice the towing speed.

with a velocity $V(x)$ in the y direction with a magnitude that can depend on its position on the x axis, and the observing points (y_1,z_1,x) and (y_2,z_2,x) move with a velocity W in the y direction.

The acoustic source has wavenumber k. The coordinate space is filled with an ocean medium, the refractive index of which is

$$n(y,z,x,t) = n_0(y,z,x,t) + \mu n'(y,z,x,t), \tag{1}$$

where n_0 is its mean value and $\mu n'(y,z,x,t)$ is the randomly varying part with mean value zero and standard deviation μ, assumed to be very small. The spatial scale sizes of $n'(y,z,x,t)$ are L_y, L_z, and L_x in the y, z, and x directions, respectively, its variance is unity, and its autocorrelation function is

$$\langle n'(y_1,z_1,x_1,t_1)n'(y_2,z_2,x_2,t_2)\rangle = \rho(y_1-y_2,z_1-z_2,x_1 - x_2,t_1-t_2). \tag{2}$$

The field at point (y_i,z_i,x) in the medium at time t_i is $E(y_i,z_i,x,t_i)$ and its fourth moment is

$$m_4 = \langle E(y_1,z_1,x,t_1)E^*(y_2,z_2,x,t_2)E(y_3,z_3,x,t_3) \times E^*(y_4,z_4,x,t_4)\rangle. \tag{3}$$

The equation governing the propagation of this fourth moment is, in the parabolic approximation of the wave equation,

$$\frac{\partial m_4}{\partial x} = \frac{-i}{2k}(\nabla_{T_1}^2 - \nabla_{T_2}^2 + \nabla_{T_3}^2 - \nabla_{T_4}^2)m_4 - \beta(2 + f_{13} + f_{24} - f_{12} - f_{14} - f_{23} - f_{34})m_4, \tag{4}$$

where

$$f_{ij} = \sigma_{ij}(y_i - y_j, z_i - z_j, t_i - t_j)/\sigma(0,0,0),$$

$$\sigma_{ij} = \int_{-\infty}^{\infty} \rho(y_i - y_j, z_i - z_j, x_i - x_j, t_i - t_j)d(x_i - x_j),$$

$$\beta = k^2\mu^2\sigma(0,0,0),$$

$$\nabla_{T_i}^2 = \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{dz_i^2}. \tag{5}$$

It can be seen from Eq. (5) that σ_{ij} is the projection of the autocorrelation function in the x direction. The solution[10,15] of Eq. (4) is the four-point moment m_4, Eq. (3). This is then reduced to a two-point moment of intensity I_i by letting point 4 go to point 1 and point 3 go to point 2. This form of the solution can be represented as[10,15]

$$\langle I_1,I_2\rangle = m_4(\eta_{a'},\zeta_a,\tau) = \iint \Phi_4(q_y,q_z,x,\tau)\exp\{i[q_y\eta_a + q_z\zeta_a]\}dq_y dq_z. \tag{6}$$

Here,

$$\eta = y_1 - y_2, \quad \zeta = z_1 - z_2, \quad \tau = t_1 - t_2$$

are the separations of two points in the y and z directions, respectively, in the plane of observation, q_y and q_z are their corresponding spatial frequencies, and τ is the separation in time. Here, Φ_4 is the spectral function of the intensity fluctuations,

$$\Phi_4(q_y,q_z,x,\tau) = \frac{1}{(2\pi)^2}\int\int \exp\left\{-2\beta x\left[1 - \int_0^1 g\left(s\eta',[1-s]q_y\frac{x}{k};s\zeta',[1-s]q_z\frac{x}{k}\right)ds\right]\right\} \times \exp\{-i(q_y\eta' + q_z\zeta')\}d\eta' d\zeta'. \tag{7}$$

For the three-dimensional case but with $W=U=0$, $\zeta_a=0$, and V in the y direction only,

$$g\left(s\eta',[1-s]q_y\frac{x}{k};s\zeta',[1-s]q_z\frac{x}{k}\right) = f(s\eta' - V\tau;s\zeta') + f\left([1-s]q_y\frac{x}{k};[1-s]q_z\frac{x}{k}\right) - \frac{1}{2}\left[f\left(s\eta' - V\tau + [1-s]q_y\frac{x}{k};s\zeta' + [1-s]q_z\frac{x}{k}\right) + f\left(s\eta' - V\tau - [1-s]q_y\frac{x}{k};s\zeta' - [1-s]q_z\frac{x}{k}\right)\right]. \tag{8}$$

Since this paper deals with acoustic scattering by internal waves solution (6) will need to be determined for this case. However, before proceeding it will be helpful to summarize some important results from the theory of ocean internal waves.

B. Internal waves

Below the mixed layer, the oceans are stable with a density that increases with depth. The surfaces of equal density, isopycnals, can support wavelike disturbances, internal waves. These regions of the world's oceans are filled with a background of fully developed internal waves in a steady state, which has been extensively studied. The waves propagate isotropically in the horizontal and their relative phases are uncorrelated. They thus constitute a randomly inhomogeneous density field with Gaussian statistics. The distribution of energy over the internal wave modes is surprisingly similar throughout the oceans and has been successfully modeled by Garrett and Munk[16] using the spectrum

$$H(j) = (j^2 + j_*^2)^{-1} \Big/ \sum_{j=1}^{\infty}(j^2 + j_*^2)^{-1}, \tag{9}$$

where j is the mode number and j_* is a characteristic mode number that determines the width of the spectrum, which is normalized as follows:

$$\sum_j H(j) = 1. \tag{10}$$

In what follows, we shall use a continuous representation of the spectrum H, replacing the mode number with the vertical wavenumber α_3,[2] related to j by

$$\alpha_3 = j\pi(b^2(z) - f^2)^{1/2}/b_0 B, \tag{11}$$

where

$$b(z) = b_0 \exp(-z/B). \tag{12}$$

Here b_0 is the buoyancy frequency at the surface and $b(z)$ is the buoyancy frequency at depth z. B is the vertical scale and is of order 1 km, while f is the inertial frequency at the latitude in question.

The continuous form of H is then, from Ref. 2,

$$H(\alpha_3) = D_0 \alpha_*^{-1} [1 + (\alpha_3/\alpha_*)^2]^{-1}. \tag{13}$$

The normalization constant D_0 is

$$D_0 = [\pi/2 - \arctan(1/j_*)]^{-1}. \tag{14}$$

Also α_* is the characteristic width of the continuous spectrum of vertical wavenumbers and so it is natural to define the vertical scale sizes of the internal wave field as

$$L_V = \alpha_*^{-1}. \tag{15}$$

The corresponding internal wavenumbers in the horizontal y and x directions are α_2 and α_1 while ω is the internal wave temporal frequency. The random medium constituted by the internal waves has an autocorrelation function[1,2] given by

$$\rho(\eta,\zeta,\xi,\tau) = \frac{r}{\pi^2} \int\int \frac{\sqrt{\alpha_1^2 + \alpha_2^2}}{[\alpha_1^2 + \alpha_2^2 + r^2\alpha_3^2]^2} H(\alpha_3)\alpha_3 \times \exp\{i[\alpha_1\xi + \alpha_2\eta + \alpha_3\zeta + \omega\tau]\} d\alpha_1 d\alpha_2 d\alpha_3, \tag{16}$$

$$r = f/b(z), \tag{17}$$

where ζ, η, and ξ are separations in the z, y, and x directions and τ is a separation in time, while α_3, α_2, and α_1 are the corresponding wavenumbers and ω is the temporal frequency.

The internal wave dispersion relation is

$$\omega^2[\alpha_1,\alpha_2,\alpha_3] = [(\alpha_1^2 + \alpha_2^2)b^2 + \alpha_3^2 f^2]/(\alpha_1^2 + \alpha_2^2 + \alpha_3^2). \tag{18}$$

IV. THE SCATTERING REGIME

A. Evaluation of scattering parameters

Some important quantities appearing in the propagation Eq. (4) can now be evaluated, enabling us to determine the scattering regime that is obtained in the present case. This will in turn allow us to select the particular form of solution for the fourth moment appropriate to the experiment.

The projected spectrum, σ, Eq. (5), is obtained by integrating ρ, Eq. (16), with respect to ξ. This yields

$$\sigma(\eta,\zeta,\tau) = \frac{r}{\pi} \int\int \frac{H(\alpha_3)}{[\alpha_2^2 + r^2\alpha_3^2]^2} \exp\{i[\alpha_2\eta + \alpha_3\zeta + \omega\tau]\} \times |\alpha_2| \alpha_3 d\alpha_2 d\alpha_3. \tag{19}$$

The quantity $\sigma(0,0,0)$ can be regarded as L_H the horizontal scale size of the irregularities in the scattering medium. Set $\eta = \zeta = \tau = 0$ in Eq. (19) to obtain

$$\sigma(0,0,0) = \frac{r}{\pi} \int K(\alpha_3) H(\alpha_3) \alpha_3 d\alpha_3, \tag{20}$$

where

$$K(\alpha_3) = \int_{-\infty}^{\infty} [\alpha_2^2 + (r\alpha_3)^2]^{-2} \alpha_2 d\alpha_2 = \frac{1}{(r\alpha_3)^2}. \tag{21}$$

Then

$$\sigma(0,0,0) = \frac{1}{r\pi} \int_{\alpha_0}^{\infty} H(\alpha_3) \alpha_3^{-1} d\alpha_3, \tag{22}$$

where α_0 is the value of α_3 corresponding to the lowest internal wave mode with $j_0 = 1$. With

$$u = \alpha_3/\alpha_* = j/j_*, \tag{23}$$

$$\sigma(0,0,0) = 2D_0(\pi r \alpha_*)^{-1} \int_{u_0}^{\infty} \frac{du}{[1 + u^2]u} = D_0(\pi r \alpha_*)^{-1} \ln(1 + u_0^{-2}). \tag{24}$$

On setting D_0, α_*, and r as given by Eqs. (14), (27), and (17), respectively, into Eq. (24) we obtain

$$\sigma(0,0,0) = L_H = \frac{Bb_0}{2\pi^2 f j_*} \ln(1 + j_*^2) \Big/ \left[\frac{\pi}{2} - \arctan(1/j_*)\right]. \tag{25}$$

The transverse autocorrelation $\sigma(\eta,0,0)$ is derived and discussed in Appendix B. Since it will be used to examine just the region around the main peak of the experimental intensity autocorrelation, we need to consider $\sigma(\eta,0,0)$ for small values of η only.

1. Horizontal scale size L_H

The horizontal scale size can be evaluated for the present case using the following values of the quantities involved:

$B = 1000$ m,

$b_0 = 5.2 \times 10^{-3}$ s^{-1},

$f = 2\sin(\text{lat})\text{c.p.d.} = 7.3 \times 10^{-5}$ s^{-1}(at Madeira),

$j_* = 3$.

These values are similar to those adopted by Munk and Zachariasen[1] when applying the Garrett–Munk model for internal waves to the deep ocean. Using them in Eq. (25) we obtain

$$\sigma(0,0,0) = L_H \approx 2130 \text{ m}. \tag{26}$$

2. Vertical scale size L_V

We adopt the following expression for the vertical scale size of the irregular structures due to internal waves (Ref. 1, p. 830):

TABLE I. Estimated scattering parameters for each frequency. The speed of sound in water is 1500 m/s.

f	482 Hz	680 Hz	740 Hz
k (m^{-1})	2.02	2.85	3.10
Γ	468	1315	1692
X	0.113	0.080	0.074
ΓX	52.8	105.2	125.2

$$L_V = \alpha_*^{-1} = b_0 B / b(\pi j_* - 1). \tag{27}$$

For a mean ray depth of about 1500 m, Eq. (27) gives

$$L_V = 530 \text{ m}. \tag{28}$$

Another estimate of L_V can be obtained from the experimentally measured sound speed profiles. It turns out to be quite close to that obtained from the standard model. These measurements are discussed in Appendix A.

3. Sound speed variance μ^2

The strength of the random internal wave features varies with depth. Estimates of μ for the Garrett–Munk internal wave model are given in Ref. 1, p. 829, Table I, for a range of depths. Using these we obtain, by interpolation, the following value of μ^2 at 500 m:

$$\mu^2 = 9 \times 10^{-8}. \tag{29}$$

This agrees with the value obtained from the experimental expendable bathythermograph (XBT) casts, as discussed in Appendix A.

4. The scaled scattering strength Γ

The scaled scattering strength of the random medium due to internal waves [Eq. (31)] can now be calculated using Eq. (5) for β with the values of the parameters L_V and $\sigma(0,0,0)$ given above. This is done for the three acoustic frequencies involved and the results are given in Table I. In order to determine the scattering regime it is convenient to use the standard scaled form of Eq. (4) (Ref. 10) in which the scaled distance of propagation is

$$X = x/kL_V^2 \tag{30}$$

and the scaled scattering strength of the medium is

$$\Gamma = k\beta L_V^2. \tag{31}$$

B. Scintillation index

A quantity frequently used when considering fluctuations in intensity I of a wave propagating in a random scattering medium is the normalized mean variance

$$S_I^2 = (\langle I^2 \rangle - \langle I \rangle^2)/\langle I \rangle^2 \tag{32}$$

sometimes referred to as the "scintillation index." The behavior of $S_I^2(X)$ for a point source is shown in Fig. 6 as a function of scaled distance X for one of the acoustic frequencies (740 Hz) used in the trial. The behavior of $S_I^2(X)$ for the other two acoustic frequencies is similar. The curve shown in

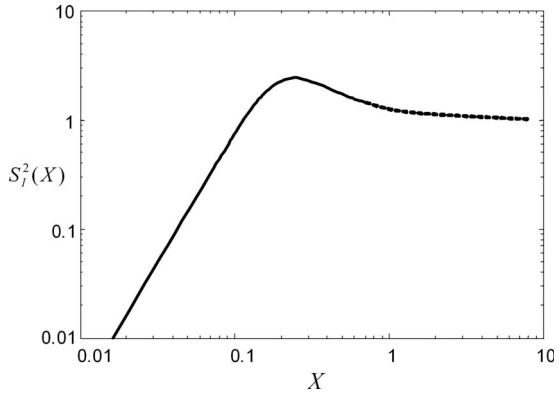


FIG. 6. The scintillation index $S_I^2(X)$ for the case of $f=740$ Hz and $\Gamma = 1692$.

Fig. 6 was calculated by direct numerical solution of the fourth moment equation for a point source using the methods described in Ref. 17 and the scattering parameters of this experiment.

Starting from zero at the source, $S_I^2(X)$ initially behaves like X^n where n is some power usually between 2 and 3. When the scattering parameter is large, the scintillation index rises to a peak greater than unity before falling off to a limiting value of unity for large X. Estimates for the position and height of the peak have been made.[17] For the random medium constituted by ocean internal waves, the value of the power n is approximately 2.5 and S_I^2 behaves initially like

$$S_I^2 = \tfrac{1}{5}\Gamma X^{2.5}. \tag{33}$$

Estimates for X_{f0}, the position of the peak; for X/X_{f0}, the position of the observing hydrophones relative to the peak; and for S_I^2, from Eq. (33), at the point of observation are given in Table II.

C. The overall picture

We have made estimates of the vertical and horizontal scale sizes and scattering strength of the random medium constituted by the deep ocean internal wave field at the site of the acoustic sea trial. These have allowed us to determine the behavior of the intensity fluctuations for each of the three acoustic frequencies used in the experiment. The approximate behavior of the scintillation index curve has been determined as a function of distance of propagation. It turns out that for all three frequencies the value of the scattering strength is large and that there is a peak in the scintillation index curve at some range X_{f0}, an estimate of which has been obtained for each frequency. It also turns out that in each case the range at which the observations were made was about 0.22 of the distance to the peak. Armed with this overall picture of the scattering regime, we are now in a position

TABLE II. Scattering range and scintillation index for each frequency.

f	482 Hz	680 Hz	740 Hz
Xf_0	0.54	0.36	0.33
X/Xf_0	0.21	0.22	0.23
$S_I^2(X)$	0.40	0.48	0.50

TABLE III. Experimental and theoretical scintillation index for each frequency.

f	482 Hz	680 Hz	740 Hz
S_I^2 (theor.)	0.40	0.48	0.50
S_I^2 (obs.)	0.46 ± 0.11	0.84 ± 0.15	0.62 ± 0.10

to look at the experimental results in some detail and compare them with those predicted by the theory of internal waves and propagation in a random medium.

V. COMPARISON OF THEORY WITH EXPERIMENTAL RESULTS

A. Scintillation index

The normalized variance of intensity fluctuations [Eq. (32)] has been evaluated from the observed data for the three transmitted acoustic frequencies. The results are shown in Table III together with their standard error bars. The values calculated from propagation theory are also given in the same table.

At 680 Hz, the experimental value of S_I^2 is much larger than the theoretical value. Agreement with the two other frequencies is reasonably good.

B. Theoretical estimate of transverse scale size

A quantity of particular interest in the context of the present experiment is the horizontal scale size of the intensity fluctuations L_I transverse to the direction of propagation. This is a feature of the acoustic field that has not been investigated previously and which this transmission trial measures. A theoretical estimate has already been made of the horizontal scale of the irregular pancakelike features due to ocean internal waves. The corresponding horizontal spatial autocorrelation function of intensity fluctuations, $m_4(\eta,0,0)$, is derived in Appendix B, Eq. (B7). This gives the horizontal scale size of the intensity fluctuations as

$$L_I = L_H(\Gamma X/3)^{-1/2}. \tag{34}$$

The result of Eq. (34) for the width of the main peak of the intensity autocorrelation is valid for small values of η, in the case when $\beta x = \Gamma X \gg 1$.

Estimates of L_I calculated in this way, and using the theoretical value of $L_H \approx 2130$ m, are given in Table IV, together with the experimental values L_I'.

C. Experimental estimate of transverse scale size

The experiment does not give us the transverse intensity scale directly. What we record is the intensity fluctuation pattern as a function of time as the array is towed through the ocean and the time scale of this pattern is T_I'. If the receivers pass through the intensity pattern with a speed V_I then the experimental estimate of the transverse scale of the intensity fluctuations is

$$L_I' = V_I T_I'. \tag{35}$$

The pattern speed V_I is discussed in the next section. For the towing speed of 5 Kn in the present trial V_I turns out to have a value of 5.14 m/s. The time autocorrelation function of the acoustic intensity pattern gives us an estimate of the time scale T_I'. Examples of this experimental autocorrelation function are shown in Fig. 7 for the three acoustic frequencies and estimates of T_I' obtained from the half-height points of these curves are given in Table IV. Estimates of L_I' obtained from Eq. (35) using the theoretical value of the pattern velocity V_I are also given in Table IV. The relation of V_I to the

TABLE IV. Transverse length and time scales for the high intensity ribbons.

f	482 Hz	680 Hz	740 Hz
$\sqrt{\Gamma X/3}$	4.2	6	6.5
L_I (m)	507	355	328
T_I' (s)	90	75	80
L_I' (m)	463	386	411

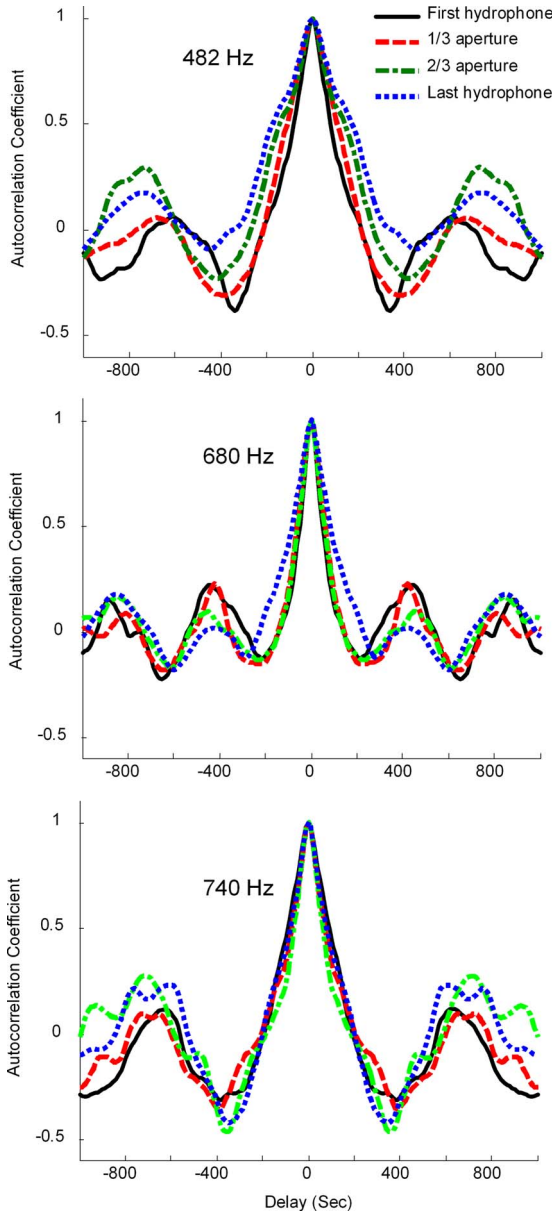


FIG. 7. (Color online) Time autocorrelation functions of the observed intensity patterns for the three acoustic frequencies at different positions down the array: first hydrophone, 1/3 aperture, 2/3 aperture, and last hydrophone.

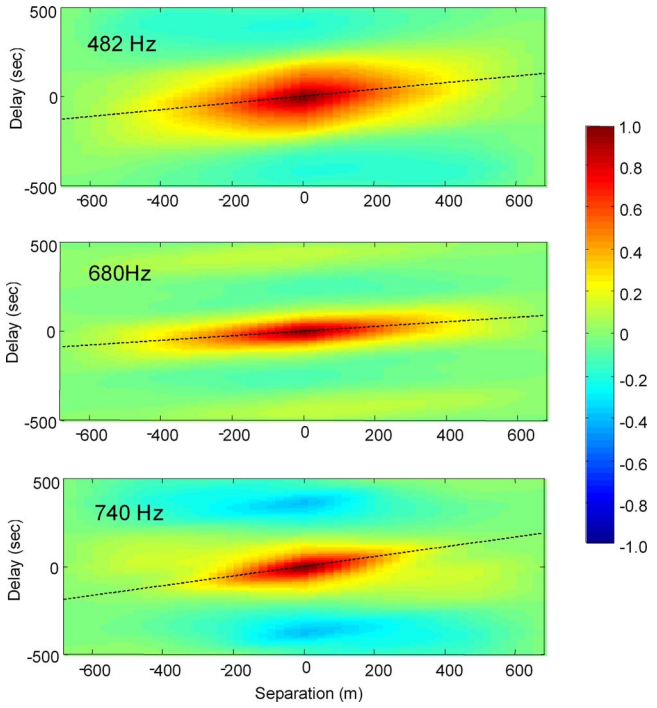


FIG. 8. (Color online) Space-time autocorrelation functions with estimated major axes superimposed.

experimentally observed values V_I' is discussed in the next section.

D. Apparent velocity of the intensity pattern V_I

In the present experiment both source and receiver move uniformly through the ocean on parallel paths with the same speed V. This, in a coordinate frame attached to the source, is the same as keeping both sources and receiver stationary and moving the medium with a uniform speed $-V$. The speed at which the acoustic intensity pattern moves across the receiver can be calculated by using the space-time intensity autocorrelation function obtained from the solution of fourth moment Eq. (6).[18] This is dominated by its large peak at small time separations τ, see Fig. 7. The solution for m_4 in the region of this peak is defined in Appendix B, Eq. (B6). This result is discussed in detail in Ref. 15, Eqs. (46)–(49), where it is shown that the speed of the intensity pattern V_I is twice that of the uniformly moving medium.

$$V_I = 2V. \tag{36}$$

We have made use of this result when calculating the spatial transverse scale of the intensity ribbons L_I.

The speed of the intensity pattern over the receiving array observed experimentally can be judged from the sets of sloping lines superimposed on Fig. 4 in which the dashed line corresponds to a pattern velocity of $2V$ and the full line corresponds to a velocity V. We see that in general the $2V$ line best reflects the slope of the high intensity features, with the occasional feature that appears to have a speed V. These latter could be explained if the high intensity feature was mostly due to scattering by an irregular feature situated close to the receiver.

Finally Fig. 8 presents the space-time intensity correlations obtained from the experimental data for the three

TABLE V. Theoretical and experimental cross-frequency intensity covariances at observation range X.

f_{ij}	f_{21}	f_{31}	f_{32}
f_i/f_j	1.42	1.54	1.09
X	0.096	0.093	0.077
C	0.35	0.2	0.5
C'	0.62 ± 0.19	0.85 ± 0.13	0.64 ± 0.24

acoustic frequencies. The experimental values for the velocity of the intensity patterns over the array can be obtained from the line corresponding to the major axis of the central ellipselike peak of this correlation function. The values for this speed are 11.4 Kn (482 and 680 Hz) and 6.8 Kn (740 Hz), giving an average value of

$$2V \approx 9.9 \text{ Kn},$$

which compares favorably with the predicted value of $V_I = 10$ Kn (5.14 m/s).

E. Cross-frequency correlation

Plots of the received acoustic signal (Fig. 4) exhibit regions of high and low intensities at all three frequencies 482, 680, and 740 Hz. The positions of these features are not quite the same in the different cases. The question arises as to whether we should expect the high intensity ribbons to be in the same place for the different acoustic frequencies.[18]

One approach to this question is to calculate the theoretical cross-correlation between intensity fluctuations at different frequencies. This has been done for a point source and a multiscale medium similar to that of an ocean containing internal waves.[19] Figure 6 shows the cross-frequency variance of intensity at different distances in the scattering medium as a function of the frequency ratio f_i/f_j. If we use the values of scaled distance into the medium calculated as averages for the cross frequencies (see Table I) and the relevant frequency ratios, Fig. 6 of Ref. 19 gives us theoretical values of the covariances C. These are presented in Table V.

The covariances C' obtained from the intensity data are given in Fig. 9 as functions of time lag and spatial separation over the array. Estimates of the covariance maxima C' for the different frequency ratios are also given in Table V. The theoretical and experimental values roughly agree for two of

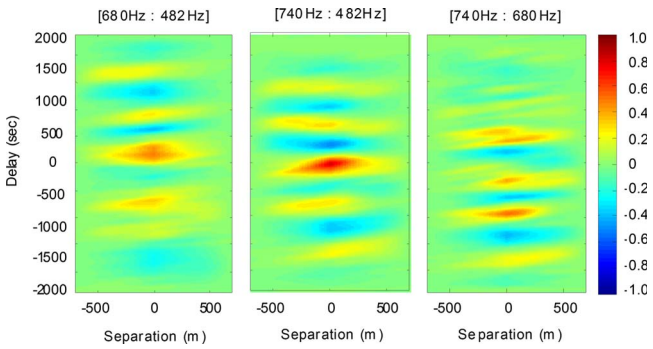


FIG. 9. (Color online) Covariances of intensity for the different frequency pairs as functions of time lag and spatial separation along the receiving array.

the frequency ratios, but correlation for the case 740 Hz:482 Hz is higher than that predicted theoretically.

Finally, we note that the covariances are mostly centered around zero spatial separation along the array implying that the higher intensity regions, when encountered at the two different frequencies, have a similar distribution along the array. The same can be said for the covariances as a function of time lag except for the case 680 Hz:740 Hz when the maxima occur for nonzero time separations, implying that the higher intensity patches are not encountered simultaneously for the two frequencies.

VI. SUMMARY

The data from an experiment investigating the horizontal structure of acoustic intensity fluctuations in the ocean have been compared with theoretical predictions. The theory employed is based on the parabolic moment equations for propagation and scattering in randomly irregular media, and on the standard Garrett–Munk model for ocean internal waves responsible for the random ocean structure.

The quantities investigated include the horizontal scale size of the acoustic intensity fluctuations transverse to the direction of propagation, the variance of the fluctuations or the "scintillation index," the correlation between intensity fluctuations at the different transmitted acoustic frequencies, and the apparent velocity of the intensity pattern over the receiving array. In general, the experimental results and the theoretical predictions agree quite well. This gives us some reassurance that we can use the present theoretical framework to help us deal with acoustic fluctuations in practical situations, in particular, when it is important to predict their behavior and structure in the horizontal, transverse to the direction of propagation. This study leaves some questions unanswered, such as the apparent lesser degree of correlation between intensity fluctuations for the frequency pair at 680 and 740 Hz as compared with the pair at 482 and 740 Hz for which the frequency separation is greater. However, this and other questions must wait until a more detailed study of the data can be carried out.

ACKNOWLEDGMENTS

B.J.U. was supported in this work by the US Office of Naval Research, Award No. N00014-06-1-0608.

The authors wish to acknowledge the support of Dr. G. A. Williams of DSTL Winfrith. The data analyzed in this paper were collected under the UK MoD funded Applied Research Programme.

APPENDIX A: THE XBT RESULTS

1. Results

Direct information about the irregular structures in the ocean medium can be obtained by analyzing the sound speed profiles derived from the XBT casts made during the trial. If the mean measured sound speed profile is subtracted from the individual profiles we are left with residuals giving the fluctuating component of sound speed. A few examples are shown below in Fig. 10. In all cases there is a long wavelength component with superimposed fluctuations of much smaller scale. The large scale features could be due to internal wave activity since both their wavelengths and amplitudes appear to be very similar to those expected from a standard deep ocean internal wave field. The very small scale features could be partly due to instrumental effects, and so the data were smoothed in order to suppress details of the order of the interhydrophone spacing, i.e., about 30 m. Figure 10 shows both the original and the smoothed data series.

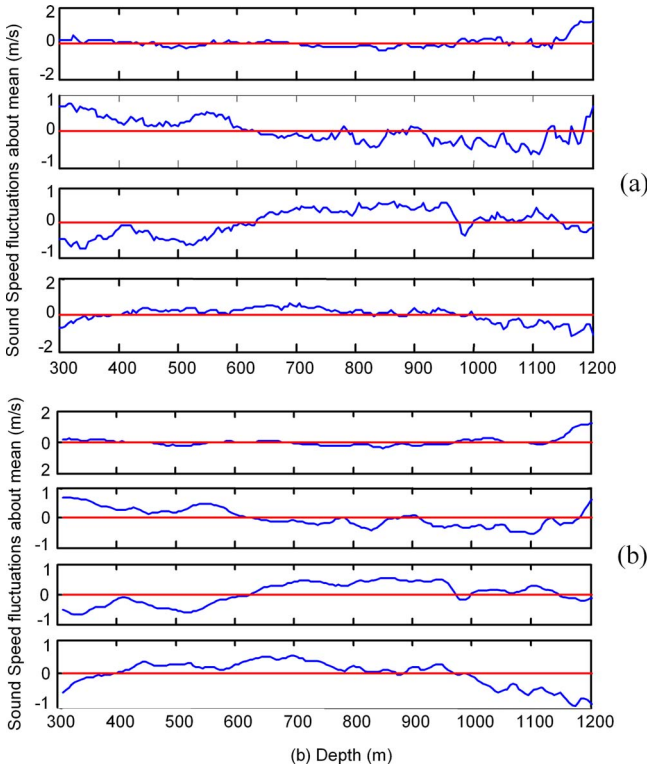


FIG. 10. (Color online) Some examples of the fluctuating part of the sound speed profile obtained from XBT casts: (a) unsmoothed data and (b) smoothed data.

A useful method of investigating the scale sizes and scattering strengths of these components is to obtain and study the mean spatial frequency spectrum of the set of residual sound speeds. This spectrum is given in Fig. 11 below. There are, in fact, two spectra in the figure. The first (a) is obtained from the data before smoothing of the small scale

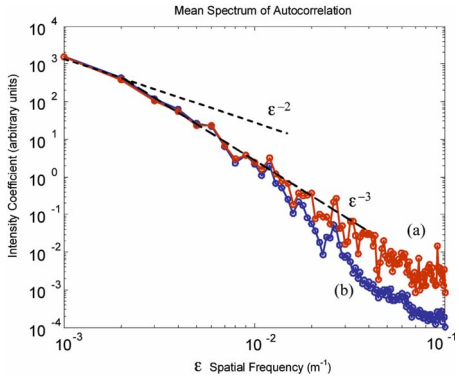


FIG. 11. (Color online) The mean spatial frequency spectrum of sound speed irregularities: (a) before smoothing and (b) after smoothing.

features, and the second (b) comes from the smoothed data. Elimination of fluctuations with lengths comparable to the interhydrophone separation leads to a negligible change in the total energy content of the spectrum. Features from the longest scales down to about 100 m remain unaffected. These could well be due to the ubiquitous presence of deep ocean internal waves.

2. Scale size

The spatial frequency spectrum obtained from the XBT casts should correspond to the spectrum of internal waves in continuous form, Eq. (13). However, we do not have a sufficient depth of record to make a proper comparison. We can identify the vertical scale size that is important in practice by noting that there is a break point in the experimental spectra where the slope changes from a power of −2 to −3. This occurs at a spatial frequency of about 2×10^{-3}/m (see Fig. 11), which implies a characteristic scale size of about 500 m. This is not too different from the value of $L_V \approx 530$ m given in Eq. (28) calculated on the basis of internal wave theory, and fits well with the appearance of the large scale undulations visible in the XBT data shown in Fig. 10.

3. The value of μ^2

The experimental sound speed fluctuation records allow us to obtain a value for the standard deviation of the randomly varying part of the acoustic refractive index. The rms value of sound speed fluctuations obtained from the residuals is 0.440 ± 0.027 m/s. The sound speed at the mean depth of the lower ray path is about 1505 m/s. Thus the experimental value of μ^2 is

$$\mu^2 = (8.6 + 1.0)\times 10^{-8}.$$

This compares favorably with the value

$$\mu^2 = 9.0\times 10^{-8}$$

obtained from the standard internal wave model.[1]

APPENDIX B: TRANSVERSE CORRELATIONS AND MOMENTS

1. The autocorrelation $\sigma(\eta,0,0)$

Set $\zeta=\tau=0$ in Eq. (19) and integrate with respect to α_3 from α_0 to infinity, to obtain

$$\sigma(\eta,0,0) = \int F(\nu)\exp(i\nu\eta_1)d\nu, \tag{B1}$$

where

$$\nu = \alpha_2(r\alpha_*)^{-1},$$

$$\eta_1 = \eta(r\alpha_*),$$

$$F(\nu) = G_0\left[\frac{\nu}{(1-\nu^2)^2}\left|\ln\left(\frac{\nu^2+j_*^{-2}}{1+j_0^{-2}}\right)\right| + \frac{\nu}{(1-\nu^2)(\nu^2+j_*^{-2})}\right], \tag{B2}$$

and

$$G_0 = D_0(\pi r\alpha_*)^{-1}.$$

We require $\sigma(\eta,0,0)$ for small values of η only, which corresponds to large values of the frequency ν.

For $\nu \gg 1$

$$F(\nu) \sim G_0(2\ln(\nu)-1)\nu^{-3}. \tag{B3}$$

The high frequency behavior of $F(\nu)$ is dominated by the ν^{-3} term. We can thus investigate the behavior of $\sigma(\eta,0,0)$ for $\nu\gg1$ by approximating Eq. (B1) as

$$\sigma(\eta_1,0,0) \sim G_0\int\frac{\cos(\nu\eta_1)}{(1+\nu^2)^{3/2}}d\nu = G_0\eta_1K_1(\eta_1). \tag{B4}$$

Then, from Eq. (5), and expanding $K_1(\eta)$ in small powers of η, we have for $\eta_1\ll1$

$$f(\eta_1,0,0) = \frac{\sigma(\eta_1,0,0)}{\sigma(0,0,0)} \approx 1 - 2\left(\frac{\eta_1}{2}\right)^2\left|\ln\left(\frac{\eta_1}{2}\right)\right|. \tag{B5}$$

2. The solution for $m_4(\eta,\zeta,\tau)$

To present a full development of the analytical solution for m_4, Eqs. (6)–(8) is beyond the scope of the present paper. This moment gives us the theoretical autocorrelation function of intensity fluctuations. However, we are interested in comparing it with data only for small time delays, i.e., in the region of the main peak that, as we see from Fig. 7, dominates the function. Now small time delays in an autocorrelation correspond to high frequencies q in its Fourier transform, i.e., its spectral representation [Eq. (7)]. For large q all terms on the right-hand side of Eq. (8), except the first, tend to zero. Using this fact, we set the resulting Eq. (7) into Eq. (6) and carry out the integrals to obtain

$$m_4(\eta,\zeta,\tau) = \exp\left\{-2\beta x\left[1-\int_0^1 f(s\eta - V\tau, s\zeta)ds\right]\right\}. \tag{B6}$$

3. The moment $m_4(\eta,0,0)$ for $\beta x\gg1$

When $\beta x\gg1$, the quantity $m_4(\eta,0,0)$ is significant only for values of $f(\eta,0,0)$ close to unity. We set f as given by Eq. (B5) into Eq. (B6) and integrate with respect to s to obtain, in terms of the scaled variables Γ and X,

$$m_4(\eta,0,0) = \exp\left\{-\Gamma X\left(\frac{2}{3}\left(\frac{\eta_1}{2}\right)^2\left|\ln\left(\frac{\eta_1}{2}\right)-\frac{1}{3}\right|\right)\right\} \approx \exp\{-\eta^2/L_I^2\}, \tag{B7}$$

where

$$L_I = L_H(\Gamma X/3)^{-1/2} \tag{B8}$$

since the square term in the exponent in Eq. (B6) dominates over the log term.

[1]W. H. Munk and F. Zachariasen, "Sound propagation through a fluctuating stratified ocean," J. Acoust. Soc. Am. **59**, 818–838 (1976).
[2]B. J. Uscinski, "Parabolic moment equations and acoustic propagation through internal waves," Proc. R. Soc. London, Ser. A **372**, 117–148 (1980).
[3]B. J. Uscinski and D. E. Reeve, "The effect of ocean inhomogeneities on array output," J. Acoust. Soc. Am. **87**, 2527–2534 (1990).
[4]T. E. Ewart, "Acoustic fluctuations in the open ocean, a measurement using a fixed refracted path," J. Acoust. Soc. Am. **60**, 46–59 (1976).
[5]T. E. Ewart and S. A. Reynolds, "The mid-ocean acoustic transmission experiment, MATE," J. Acoust. Soc. Am. **75**, 785–802 (1984).
[6]S. M. Flatte, S. A. Reynolds, R. Dashen, B. Buehler, and P. Maciejewski, "AFAR measurements of intensity and intensity moments," J. Acoust. Soc. Am. **82**, 973–980 (1987).
[7]T. E. Ewart and S. A. Reynolds, "Instrumentation to measure the depth/time fluctuations in acoustic pulses propagated through Arctic internal waves," J. Atmos. Ocean. Technol. **7**, 129–139 (1990).
[8]B. J. Uscinski, J. R. Potter, and T. Akal, "Broadband acoustic transmission fluctuations during NAPOLI 85," J. Acoust. Soc. Am. **86**, 706–715 (1989).
[9]J. R. Potter, B. J. Uscinski, and T. Akal, "Random focusing of sound into spatially coherent regions," Waves Random Media **10**, 199–216 (2000).
[10]B. J. Uscinski, "Intensity fluctuations in a multiple scattering medium. Solution of the fourth moment equation," Proc. R. Soc. London, Ser. A **380**, 137–169 (1982).
[11]J. R. Potter, "Fine structure in NAPOLI 85, an ocean-acoustic experiment," J. Acoust. Soc. Am. **89**, 1643–1655 (1991).
[12]C. Macaskill and T. E. Ewart, "Numerical solution of the fourth moment equation for acoustic intensity correlations and comparison with the mid-ocean acoustic transmission experiment," J. Acoust. Soc. Am. **99**, 1419–1429 (1996).
[13]B. J. Uscinski, "Numerical simulations and moments of the field from a point source in a random medium," J. Mod. Opt. **36**, 1631–1643 (1989).
[14]B. J. Uscinski and M. Spivack, "High intensity ribbons in multiply scattering media," Waves Random Complex Media **15**, 339–352 (2005).
[15]B. J. Uscinski, "Intensity fluctuations in a moving random medium," Waves Random Complex Media **15**, 437–450 (2005).
[16]C. Garrett and W. Munk, "Space-time scales of ocean internal waves," Geophys. Fluid Dyn. **2**, 225–264 (1972).
[17]D. E. Reeve, S. R. Leonard, and M. Spivack, "Numerical solution of the fourth-moment equation for a point source," J. Mod. Opt. **37**, 965–975 (1990).
[18]B. J. Uscinski, *The Elements of Wave Propagation in Random Media* (McGraw-Hill, London, 1977).
[19]B. J. Uscinski and C. Macaskill, "Frequency cross-correlation of intensity fluctuations," Opt. Acta **32**, 71–89 (1985).

Numerical design of Alberich anechoic coatings with superellipsoidal cavities of mixed sizes

Sven M. Ivansson
Department of Underwater Research, Swedish Defence Research Agency, SE-164 90 Stockholm, Sweden

(Received 30 January 2008; revised 26 June 2008; accepted 9 July 2008)

Thin rubber coatings with cavities in a doubly periodic lattice are able to reduce reflections of underwater sound by redistributing normally incident energy such that absorption in the surrounding rubber is enhanced. For spherical scatterers, the anechoic effect can be studied numerically by the layer-multiple-scattering (LMS) method. In comparison to more flexible but also more computer intensive methods, such as finite-element method modeling, there are two important advantages. An improved physical understanding of the anechoic effect can be achieved by simplified semianalytical analysis, and the high computational speed allows modern global optimization techniques to be applied for coating design. In this paper, the flexibility of the LMS method is improved by combination with an efficient algorithm for numerical computation of transition matrices for superellipsoidal scatterers. (A superellipsoid is a generalization of an ellipsoid, allowing more box-filling shapes, for example.) Extensions to mixtures of nonspherical scatterers of different types are also considered, in order to enhance the broadband performance. Symmetry properties are used to reduce the size of the pertinent equation systems. Examples of numerical coating design for underwater acoustic applications are presented, using differential evolution algorithms for the optimization. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967840]

PACS number(s): 43.30.Ky, 43.20.Bi, 43.30.Gv [KGF] Pages: 1974–1984

I. INTRODUCTION

The phenomenon of optic band gaps for transmission through photonic crystals has recently stimulated much research on wave propagation through phononic crystals, i.e., analogous structures with periodically located scatterers for the acoustic case. Two mechanisms that can lead to gaps in the frequency response have been identified:[1] Lattice resonances[2] and localized resonances.[3,4] As described in recent review papers,[5,6] several computational methods have been adapted and developed, for example the layer-multiple-scattering (LMS) method[7,8] and the finite-difference time domain (FDTD) method. The LMS method has typically been used with identical spherical [or cylindrical, in the "two dimensional (2D) case"] scatterers. One objective with the present paper is to extend the LMS analysis to scatterers of more general shapes. Extensions to lattices including scatterers of mixed sizes are also made.

Phononic crystals have mainly been applied to sound shielding, i.e., reduction of sound transmission, but the new research has also been useful to reconsider reduction of sound reflections. For example, the LMS method has recently been used to study underwater anechoic coatings of Alberich type,[9–11] i.e., rubber coatings with a periodic lattice of cavities or other scatterers. Numerical design of such Alberich coatings is an important application topic of the present paper. A coating should be thin and provide significant reflection reduction over a broad frequency band. It seems reasonable that combinations of different cavity length scales could be useful. In particular, the mentioned extensions of the LMS method are used to investigate to what extent nonspherical cavity shapes and cavities of mixed sizes can improve the coating performance.

The design problem can be formulated as a nonlinear optimization problem. Hence, global optimization techniques from inverse theory can be applied. Illustrative broadband examples using a differential evolution (DE) algorithm[12] are included. Coating optimization has previously been attempted using a simplex procedure with a broadband objective function based on mechanical resonator formulas[13] and using a zero-order method for a single-frequency objective function with finite-element method (FEM) forward modeling.[14] The modeling flexibility of FEM is useful for applications to Alberich anechoic coatings,[15,16] but repeated calculations for optimization purposes can be computationally demanding.

Hence, the semianalytical LMS method and the purely numerical methods, such as FDTD and FEM, are complementary tools. Advantages with the LMS method are its computational speed, which makes it useful for forward modeling in connection with extensive optimization computations and the physical insight it provides. The flexibility to handle various scatterer shapes has limitations, but extensions are possible and desirable. The present paper takes a step in this direction.

The plan of this paper is as follows. Section II gives a brief review of the basic LMS computational method. Transition (T) matrices are an essential ingredient. Moreover, it is described how computational speed can be enhanced by splitting equation systems and by utilizing symmetry. The main extensions are developed in Sec. III. An algorithm by Boström[17] is incorporated for computation of T-matrices for general rotationally symmetrical objects, coupled equation systems are formed to deal with scatterers of mixed types, and translation properties of spherical wave functions are

 0001-4966/2008/124(4)/1974/11/$23.00

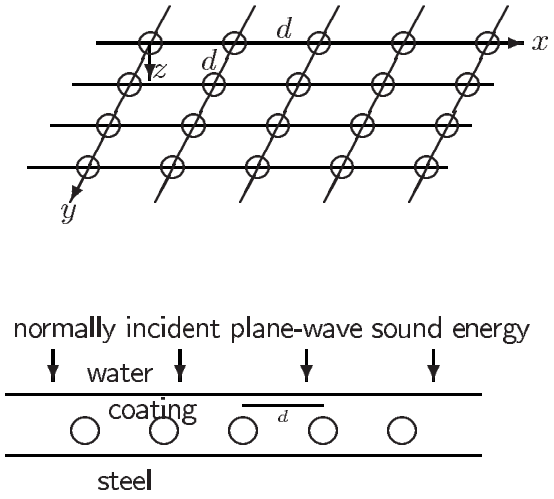


FIG. 1. Top: A scatterer lattice with period d is shown. Horizontal xy coordinates are introduced along with a z depth coordinate axis. The rotation angle between the xy axes and the lattice period directions, denoted χ in the text, vanishes in this illustration. Bottom: A steel plate is covered with an Alberich rubber coating with spherical cavities, and a plane sound wave is incident from the water above. The section of the paper corresponds to the xz plane from the top panel.

used to handle depth differences within the coating for the different scatterer types. Section IV treats the scattered field by an isolated cavity. It is shown by a few relevant examples how scattering and absorption cross sections are changed when the scatterer shape is changed. Some implications are noted for the expected performance of a coating with many cavities in a lattice. Section V provides a few numerical coating design examples. Section VI, finally, summarizes the results.

II. BASIC LAYER-MULTIPLE-SCATTERING COMPUTATIONAL METHOD

As in Fig. 1, a right-hand Cartesian xyz coordinate system is introduced in a fluid-solid medium surrounded by homogeneous half-spaces. The medium is periodic with period d in horizontal directions rotated an angle χ from the x and y directions, respectively. Consideration of nonvanishing χ is convenient later on, although the period directions coincide with the axes directions in Fig. 1. Sound waves with time dependence $\exp(-i\omega t)$, to be suppressed in the formulas, are considered, where ω is the angular frequency. It follows that an incident plane wave with horizontal wavenumber vector $\mathbf{k}_\parallel$ gives rise to a linear combination of reflected and transmitted plane waves with displacement vectors

$$\mathbf{u}(\mathbf{r}) = \exp(i\mathbf{K}^s_{\mathbf{g}j} \cdot \mathbf{r})\mathbf{e}_j. \tag{1}$$

Here, $\mathbf{r}=(x,y,z)$, $j=1,2,3$ for a wave of type P, SV, SH, respectively, $s=+(-)$ for a wave in the positive (negative) depth direction z, and

$$\mathbf{K}^{\pm}_{\mathbf{g}j} = \mathbf{k}_\parallel + \mathbf{g} \pm [\omega^2/c_j^2 - |\mathbf{k}_\parallel + \mathbf{g}|^2]^{1/2}(0,0,1) \tag{2}$$

$$= \frac{\omega}{c_j}(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{3}$$

where $\mathbf{g}$ belongs to the reciprocal lattice

$$\mathbf{g} = (k_x, k_y, 0) = 2\pi m/d(\cos\chi, \sin\chi, 0) + 2\pi n/d(-\sin\chi, \cos\chi, 0) \tag{4}$$

with m and n running over the integers. Furthermore, c_1 is the compressional-wave velocity α, and $c_2=c_3$ is the shear-wave velocity β. The angular variables θ,ϕ of $\mathbf{K}^{\pm}_{\mathbf{g}j}$ are defined by Eq. (3), with a possibly complex $\cos\theta$. The vectors $\mathbf{e}_j=\mathbf{e}_j(\mathbf{K}^{\pm}_{\mathbf{g}j})$ are defined by $\mathbf{e}_1=(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, $\mathbf{e}_2=(\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)$, $\mathbf{e}_3=(-\sin\phi, \cos\phi, 0)$.

As detailed in Refs. 9 and 10, and references therein, reflection and transmission (R/T) matrices $\mathcal{R}_B, \mathcal{T}_B$ and $\mathcal{R}_A, \mathcal{T}_A$ can now be introduced, for the discrete set of waves specified by Eqs. (1)–(4). Including one scatterer interface within the rubber layer, three interfaces are involved in the lower panel of Fig. 1. Individual R/T matrices can be combined recursively,[18,19] and layer thicknesses are conveniently accounted for by phase shifts of the complex amplitudes of the plane-wave components.

A. Interface with periodically distributed scatterers of a common type

Explicit expressions for the R/T matrices are well known for an interface between two homogeneous half-spaces.[19] Spherical vector solutions to the wave equations can be used to handle an interface with periodically distributed scatterers of one and the same type.[7] As in Refs. 9 and 10, $\mathbf{u}^{0L}_{lm}(\mathbf{r}), \mathbf{u}^{0M}_{lm}(\mathbf{r}), \mathbf{u}^{0N}_{lm}(\mathbf{r})$ and $\mathbf{u}^{+L}_{lm}(\mathbf{r}), \mathbf{u}^{+M}_{lm}(\mathbf{r}), \mathbf{u}^{+N}_{lm}(\mathbf{r})$ are here used to denote regular solutions involving the spherical Bessel function j_l and outgoing solutions involving the spherical Hankel function h^+_l, respectively. The L and N solutions are spheroidal, while the M solutions are toroidal. The pertinent wave velocity is α for the L solutions and β for the M and N solutions. The index $l=0,1,2,\ldots$ with $m=-l,\ldots,l$, since the spherical harmonics are involved, but it is understood that $\mathbf{u}^{0M}_{00} \equiv \mathbf{u}^{0N}_{00} \equiv \mathbf{u}^{+M}_{00} \equiv \mathbf{u}^{+N}_{00} \equiv \mathbf{0}$. For identical cavities or other scatterers at $\mathbf{R}=md(\cos\chi, \sin\chi, 0) + nd(-\sin\chi, \cos\chi, 0)$, with integers m,n, and an incident plane wave as in Eq. (1), the total scattered field $\mathbf{u}_{\mathrm{sc}}$ can be written as[7]

$$\mathbf{u}_{\mathrm{sc}}(\mathbf{r}) = \sum_{Plm}\left(b^{+P}_{lm}\sum_{\mathbf{R}} e^{i\mathbf{k}_\parallel\cdot\mathbf{R}}\mathbf{u}^{+P}_{lm}(\mathbf{r}-\mathbf{R})\right), \tag{5}$$

where the index P runs over $P=L,M,N$. The vector $\mathbf{b}^+ = \{b^{+P}_{lm}\}$ is determined by solving

$$(I - T\cdot\Omega)\cdot\mathbf{b}^+ = T\cdot\mathbf{a}^0, \tag{6}$$

where I is the identity matrix, $\mathbf{a}^0=\{a^{0P}_{lm}\}$ gives the coefficients for expansion of the incident plane wave in regular spherical waves $\mathbf{u}^{0P}_{lm}(\mathbf{r})$, $\Omega=\Omega(d,\chi,\mathbf{k}_\parallel,\omega/\alpha,\omega/\beta)$ is the lattice translation matrix $\{\Omega^{PP'}_{lm;l'm'}\}$, and $T=\{T^{PP'}_{lm;l'm'}\}$ is the transition matrix for an individual scatterer. Specifically, $\mathbf{b}'=\Omega\cdot\mathbf{b}^+$ and $\mathbf{b}^+=T\cdot(\mathbf{a}^0+\mathbf{b}')$, where $\mathbf{b}'=\{b'^P_{lm}\}$ gives the coefficients for expansion in regular spherical waves $\mathbf{u}^{0P}_{lm}(\mathbf{r})$ of the scattered field from all scatterers except the one at the origin. A computer program for calculating Ω can be found in Ref. 18.

The R/T matrices are obtained by transforming the expansion (5), for each pertinent incident plane wave, to plane waves of type (1). Specifically,[7] Eq. (5) can be rewritten as

$$\mathbf{u}_{\mathrm{sc}}(\mathbf{r}) = \sum_{\mathbf{g}} \sum_{j=1}^{3} \Delta(\mathbf{g}, j, \mathbf{k}_{\parallel}, \mathbf{b}^{+}) \exp(i\mathbf{K}_{\mathbf{g}j}^{\pm} \cdot \mathbf{r}) \mathbf{e}_j, \tag{7}$$

where the coefficients Δ additionally depend on the lattice and medium wavenumber parameters d, χ, and ω/α, ω/β. The sign choice for $\mathbf{K}_{\mathbf{g}j}^{\pm}$ is given by the sign of z.

The matrices $\mathcal{R}_B, \mathcal{T}_B$ are for incident plane waves from above, whereas $\mathcal{R}_A, \mathcal{T}_A$ concern incidence from below. For scatterers which are symmetric with respect to reflection in the interface, the R/T matrices $\mathcal{R}_B, \mathcal{T}_B$ can be obtained from $\mathcal{R}_A, \mathcal{T}_A$, and vice versa.

B. Equation system split for spherical scatterers

For the case of a spherical scatterer, explicit analytical expressions can be given for the transition matrix $T = \{T_{lm;l'm'}^{P,P'}\}$, e.g., Ref. 7. Furthermore, scattering only appears to the same l,m components ($l'=l$, $m'=m$), and also to the same type ($P'=P$) except that L waves can be scattered to N waves and vice versa.

The matrices Ω have the following properties.[7,20]

(1) Compressional- and shear-wave elements are not mixed, i.e., $\Omega_{lm;l'm'}^{PP'}$ vanishes when precisely one of P,P' is L and the other is M or N.
(2) $\Omega_{lm;l'm'}^{LL}$, $\Omega_{lm;l'm'}^{MM}$, and $\Omega_{lm;l'm'}^{NN}$ vanish unless $(l'+m')-(l+m)$ is even.
(3) $\Omega_{lm;l'm'}^{MN}$ and $\Omega_{lm;l'm'}^{NM} = -\Omega_{lm;l'm'}^{MN}$ vanish unless $(l'+m')-(l+m)$ is odd.

As is well known,[7] it follows by considering $T \cdot \Omega$ that the equation system (6) can be split into two subsystems. One subsystem concerns the b_{lm}^{+L}, b_{lm}^{+N} with odd $l+m$ together with the b_{lm}^{+M} with even $l+m$, while the other subsystem concerns the b_{lm}^{+L}, b_{lm}^{+N} with even $l+m$ together with the b_{lm}^{+M} with odd $l+m$.

C. Utilizing symmetry

The recursive combination of individual R/T matrices[18,19] involves matrix inversions, where the matrix dimension equals the finite number of included plane waves from Eqs. (1)–(4). For a compressional wave at normal incidence, $\mathbf{k}_{\parallel}=\mathbf{0}$, on the square scatterer lattice(s), symmetry arguments show that many of the reflected and transmitted plane waves have equal coefficients. By forming a new wave basis from sums of plane waves with equal scattering coefficients, a matrix dimension reduction is achieved with almost a factor of 8. It is waves with horizontal wavenumber vectors $(\pm k_x, \pm k_y, 0)$ and $(\pm k_y, \pm k_x, 0)$ that can be combined. Normally incident waves still appear separately. Other waves with $k_x=0$ or $k_y=0$, or $k_x=\pm k_y$, appear in groups of four waves, rather than eight. The dramatic reduction of computation time that results is essential for the coating design examples in Sec. V.

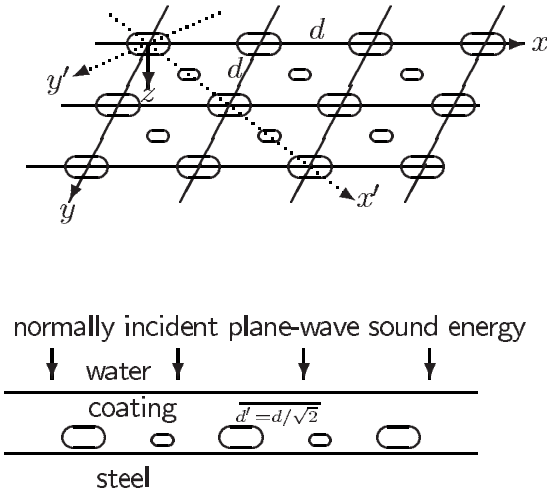


FIG. 2. The configuration from Fig. 1 is extended here. Top: The new scatterer lattice with period d is shown. Horizontal xy coordinates are introduced in the plane of the centers of one of the two scatterer types, along with a z coordinate axis. Rotated $x'y'$ axes are also introduced, tilted an angle $\chi'=\pi/4$ with respect to the xy lattice period directions. Bottom: A steel plate is covered with an Alberich rubber coating with nonspherical cavities of two different types. The section of the paper corresponds to the $x'z$ plane from the top panel. Hence, the distance between two neighboring cavities equals $d'=d/\sqrt{2}$.

Each sum of plane waves is expanded in regular spherical waves to get the $\mathbf{a}^0$ vectors for the equation system (6). As pointed out in Ref. 18, symmetry arguments make it possible to reduce the dimension of this equation system as well.

III. EXTENSIONS

The LMS computational method is commonly applied for spherical scatterers, while other methods such as FDTD have been used for more general scatterer shapes. As described in Sec. III A, however, an effective method for nonspherical scatterers is obtained by combining the LMS method with an efficient method for computing T-matrices of nonspherical objects.

It is also possible to handle scatterers of different shapes, and vertical center positions, within the same layer. This is shown in Secs. III B and III C. An illustration of the extended scatterer configuration is given in Fig. 2. There are two types of scatterers. All scatterers of the same type are centered in the same horizontal plane. For the situation in Fig. 2, there is no horizontal plane separating the two types of scatterers, and recursive combination of individual R/T matrices is not applicable.

With some restrictions, to be indicated, the possibilities for equation system split, as described in Sec. II B, remain. Symmetry arguments, as in Sec. II C, can still be used to reduce equation system dimensions.

A. Nonspherical scatterers

The extension from the configuration in Fig. 1 to the one in Fig. 2 is made in a stepwise fashion. The first step is to allow nonspherical scatterers of one and the same type in Fig. 1.

For such a scatterer, the transition matrix T must in general be computed numerically. Methods for this purpose have

been developed, e.g., the null-field approach, which is attractive and efficient when it is applicable. As described in Ref. 17, for example, general representation formulas are used to express the expansion coefficients of the incident and the scattered fields, respectively, in terms of integrals of displacements and tractions over the scatterer surface. The surface fields are expanded in some suitable basis of vector wave functions, and the desired transition matrix is obtained by matrix inversion followed by matrix multiplication. The scatterer can be a cavity or a penetrable body, either fluid or solid.

The null-field approach works best for scatterers for which the ratio between the largest and smallest radii does not essentially exceed 2. For large ratios, many terms must be included in the field expansions, making the method less attractive. Symmetry is a general property of a transition matrix, i.e., $T^{PP'}_{lm;l'm'}=T^{P'P}_{l'm';lm}$. Unless this is used explicitly in the computational routine, it provides a useful test on the accuracy.

With a subroutine for numerical computation of the transition matrices, it becomes a straightforward matter to extend the plane-wave R/T matrix calculations to layers including lattices with nonspherical scatterers. Restriction is here made to scatterers which are symmetric with respect to rotation around a vertical axis (parallel to the z axis). In this case, a Fourier expansion in the azimuthal coordinate ϕ shows that $T^{PP'}_{lm;l'm'}$ must vanish unless $m'=m$. Furthermore, the azimuthal parts of the surface integrals required in the numerical computation of the transition matrices become trivial. Only line integrals for the polar direction remain, which can be handled conveniently by Gauss quadrature.

The transition matrix is not diagonal in the l index anymore, however, and product matrices like $T\cdot\Omega$ in Eq. (6) become more complicated to compute. In general, the coupling between elements with different l and l' in $T^{PP'}_{lm;l'm}$ destroys the possibility to split the equation system (6) in two subsystems. For the particular, and common, case of scatterers which are additionally symmetric with respect to reflection in the xy plane, the split from Sec. II B is still possible. This follows easily from the specified properties of the Ω matrices, together with the following properties[21] of the T matrices for such objects.

(i) When none or both of P,P' equal $M,T^{PP'}_{lm;l'm}$ vanishes unless $l'-l$ is even.

(ii) When precisely one of P,P' equals $M,T^{PP'}_{lm;l'm}$ vanishes unless $l'-l$ is odd.

Only scatterers of superellipsoidal shape are considered in this paper. In a local centered coordinate system, with horizontal coordinates x,y and a depth coordinate z, the equation of the scatterer surface is

$$\left(\frac{|z|}{a}\right)^p+\left(\frac{r}{b}\right)^p=1, \tag{8}$$

where $r=|(x,y)|$. The number p equals 2 for an ordinary ellipsoid. For reasons of computational efficiency and accuracy, the examples are limited to $\max(a,b)/\min(a,b)\leqslant 2.5$, and $p=2$ or 4 (or, in a test case, 8). In its local coordinate system, each scatterer is apparently symmetric with respect to rotation around the z axis as well as reflection in the xy plane.

B. Different types of scatterers at the same interface

The next step leading to the configuration in Fig. 2 is to allow two types of scatterers at the same horizontal interface. The following notation is introduced, for an xyz coordinate system as in Fig. 2. Scatterers of the first type, with transition matrix T and scattered-field expansion coefficients denoted $\mathbf{b}^+$, appear at, and are rotationally symmetric around vertical axes through, the points $\mathbf{R}=m\cdot(d,0,0)+n\cdot(0,d,0)$, for integers m,n. Scatterers of the second type, with transition matrix U and scattered-field expansion coefficients denoted $\mathbf{c}^+$, appear at, and are rotationally symmetric around vertical axes through, points in between, i.e., $\mathbf{S}=(m+1/2)\cdot(d,0,0)+(n+1/2)\cdot(0,d,0)$. The reciprocal lattice vectors become $\mathbf{g}=(2\pi m/d,2\pi n/d,0)$, where m and n run over the integers.

The generalization of Eq. (5) for the scattered field becomes

$$\mathbf{u}_{\mathrm{sc}}(\mathbf{r})=\sum_{Plm}\left(b^{+P}_{lm}\sum_{\mathbf{R}}e^{i\mathbf{k}_\parallel\cdot\mathbf{R}}\mathbf{u}^{+P}_{lm}(\mathbf{r}-\mathbf{R})\right)+\sum_{Plm}\left(c^{+P}_{lm}\sum_{\mathbf{S}}e^{i\mathbf{k}_\parallel\cdot\mathbf{S}}\mathbf{u}^{+P}_{lm}(\mathbf{r}-\mathbf{S})\right). \tag{9}$$

It follows that

$$\mathbf{b}^+=T\cdot(\mathbf{a}^0+\mathbf{b}'+\mathbf{b}''),\quad \mathbf{c}^+=U\cdot(\mathbf{a}^0+\mathbf{c}'+\mathbf{c}''), \tag{10}$$

where, for a scatterer of the first type at $\mathbf{R}$, $\exp(i\mathbf{k}_\parallel\cdot\mathbf{R})\mathbf{b}'$ and $\exp(i\mathbf{k}_\parallel\cdot\mathbf{R})\mathbf{b}''$ give the coefficients for expansion in regular spherical waves $\mathbf{u}^{0P}_{lm}(\mathbf{r}-\mathbf{R})$ of the scattered field from all other scatterers of the first and the second type, respectively. For a scatterer of the second type at $\mathbf{S}$, $\exp(i\mathbf{k}_\parallel\cdot\mathbf{S})\mathbf{c}'$ and $\exp(i\mathbf{k}_\parallel\cdot\mathbf{S})\mathbf{c}''$ analogously give the coefficients for expansion in regular spherical waves $\mathbf{u}^{0P}_{lm}(\mathbf{r}-\mathbf{S})$ of the scattered field from all other scatterers of the second and the first type, respectively.

With $\Omega^0=\Omega(d,0,\mathbf{k}_\parallel,\omega/\alpha,\omega/\beta)$, it follows that $\mathbf{b}'=\Omega^0\cdot\mathbf{b}^+$ and $\mathbf{c}'=\Omega^0\cdot\mathbf{c}^+$. For a certain matrix Q, to be determined, $\mathbf{b}''=Q\cdot\mathbf{c}^+$ and $\mathbf{c}''=Q\cdot\mathbf{b}^+$. The equation system to be solved for $\mathbf{b}^+$ and $\mathbf{c}^+$ becomes

$$\begin{pmatrix} I-T\cdot\Omega^0 & -T\cdot Q \\ -U\cdot Q & I-U\cdot\Omega^0 \end{pmatrix}\cdot\begin{pmatrix}\mathbf{b}^+\\ \mathbf{c}^+\end{pmatrix}=\begin{pmatrix}T\cdot\mathbf{a}^0\\ U\cdot\mathbf{a}^0\end{pmatrix}. \tag{11}$$

In order to form the R/T matrices, incident plane waves with different horizontal wavenumber vectors $\mathbf{k}_\parallel+\mathbf{g}_{\mathrm{inc}}$ have to be considered, where $\mathbf{g}_{\mathrm{inc}}$ belongs to the reciprocal lattice $\{(2\pi m/d,2\pi n/d,0)\}$. Noting that the union of the scatterer positions is a small square lattice with period $d'=d/\sqrt{2}$ tilted an angle $\chi'=\pi/4$ with respect to the xy axes, see Fig. 2, the following expression for Q as a difference of Ω matrices is directly obtained:

$$Q=\Omega(d/\sqrt{2},\pi/4,\mathbf{k}_\parallel+\mathbf{g}_{\mathrm{inc}},\omega/\alpha,\omega/\beta)-\Omega^0. \tag{12}$$

Only those $\mathbf{g}_{\mathrm{inc}}$ in $\{(2\pi m/d,2\pi n/d,0)\}$ for which $m-n$ is even are reciprocal vectors for the small tilted lattice. Since a lattice translation matrix Ω is periodic in its third argument

with the same periods as the corresponding reciprocal lattice, there are two groups of $\mathbf{g}_{\text{inc}}$ with different Q matrices according to Eq. (12).

The transformation of the expansion (9) to plane waves of the type (1) can be made separately for each of the **R** and **S** sums. For the **S** case, the translation from the origin causes a sign change for some combinations of incident $\mathbf{g}_{\text{inc}}$ and scattered $\mathbf{g}_{\text{sc}}$ reciprocal lattice vectors. Specifically, with $\mathbf{g}_{\text{inc}}=(2\pi m_{\text{inc}}/d, 2\pi n_{\text{inc}}/d, 0)$ and $\mathbf{g}_{\text{sc}}=(2\pi m_{\text{sc}}/d, 2\pi n_{\text{sc}}/d, 0)$, the double sum involving $\mathbf{c}^{+}$, corresponding to the one in Eq. (7), appears as

Provided that *both* types of scatterers are symmetric with respect to reflections in the xy plane, $T^{PP'}_{lm;l'm}$ as well as $U^{PP'}_{lm;l'm}$ fulfill restriction properties, Sec. III A, for coupling between elements with different l and l'. The system (11) can thus be split in the same way as the system (6) in this case. Specifically, one subsystem concerns the b^{+L}_{lm}, b^{+N}_{lm} and c^{+L}_{lm}, c^{+N}_{lm} with odd $l+m$ together with the b^{+M}_{lm} and c^{+M}_{lm} with even $l+m$. The other subsystem concerns the b^{+L}_{lm}, b^{+N}_{lm} and c^{+L}_{lm}, c^{+N}_{lm} with even $l+m$ together with the b^{+M}_{lm} and c^{+M}_{lm} with odd $l+m$.

C. Computation of transition matrices for vertically displaced scatterers

The centers for the two types of scatterers in Fig. 2 are vertically displaced relative to each other. For a displacement large enough for the two kinds of objects to be separable by a horizontal plane, the general technique for recursive computation of R/T matrices is of course preferable, with one scatterer interface for each kind. It is for smaller displacements that the special developments in Sec. III B are necessary.

All points **R** and **S** from Sec. III B are in the chosen xy plane, with $z=0$. The transition matrices T and U are defined for spherical vector wave functions with origins at these points.

On the other hand, the transition matrix for a superellipsoidal scatterer, with surface given by Eq. (8), is most conveniently computed relative to a local coordinate system with the center of the scatterer at the origin. The equation of the surface in polar coordinates can be expressed in closed form, and all coordinate planes are symmetry planes with respect to reflection. Moreover, for direct application of the null-field method, the local origin should be inside the scatterer and the radius to a point on the scatterer surface should be a continuous one-valued function of the spherical angles.[17]

Transition matrices for different coordinate origins can be related, however, by applying explicit formulas for translation of vector wave functions. Specifically,[22] there are matrices $S(\zeta,m)=\{S^{PP'}_{ll'}(\zeta,m)\}$, with elements depending on the azimuthal index m and the translation ζ along the z axis, as indicated, such that

$$\mathbf{u}^{0P}_{lm}(\zeta\mathbf{e}_z+\mathbf{r})=\sum_{P'l'} S^{PP'}_{ll'}(\zeta,m)\cdot\mathbf{u}^{0P'}_{l'm}(\mathbf{r}), \tag{14}$$

$$\mathbf{u}^{+P}_{lm}(\zeta\mathbf{e}_z+\mathbf{r})=\sum_{P'l'} S^{PP'}_{ll'}(\zeta,m)\cdot\mathbf{u}^{+P'}_{l'm}(\mathbf{r}). \tag{15}$$

That $|\mathbf{r}|>|\zeta|$ is a natural restriction for the validity of the relation (15). The explicit expressions for the S matrix elements involve the Wigner 3-j symbols. It is useful to note that the inverse $S(-\zeta,m)$ of $S(\zeta,m)$ is readily obtainable without explicit matrix inversion. In fact, the matrices are "almost Hermitian," and a slightly different definition of the vector wave functions makes the transformation matrices orthogonal.[22] As detailed in Ref. 22, for example, explicit formulas also exist for general translation vectors, but coupling among azimuthal indices m then appears.

Suppose now that a transition matrix U, as for the second type of scatterers in Sec. III B, has been computed relative to the spherical wave functions $\{\mathbf{u}^{0P}_{lm}(\mathbf{r})\},\{\mathbf{u}^{+P}_{lm}(\mathbf{r})\}$. It follows easily from Eqs. (14) and (15) that the corresponding transition matrix relative to the spherical wave functions $\{\mathbf{u}^{0P}_{lm}(\zeta\mathbf{e}_z+\mathbf{r})\},\{\mathbf{u}^{+P}_{lm}(\zeta\mathbf{e}_z+\mathbf{r})\}$ is given by

$$[S(-\zeta,m)]^t\cdot U(m)\cdot[S(\zeta,m)]^t, \tag{16}$$

where a transposed matrix is denoted by the superscript t. As usual, restriction is made to scatterers with rotational symmetry around the z axis, implying that the transition matrix U is split into uncoupled blocks $\{U(m)\}$ for the different azimuthal indices m.

It was noted at the end of Sec. II A that the R/T matrices $\mathcal{R}_B, \mathcal{T}_B$ can sometimes be obtained directly from $\mathcal{R}_A, \mathcal{T}_A$, and vice versa, without further computation. With two types of vertically displaced scatterers, as in Fig. 2, the symmetry with respect to reflection in a horizontal plane is of course lost, and incident waves in the upward as well as downward directions have to be considered explicitly.

D. Geometrical restrictions

The LMS method used puts some mild geometrical restrictions on the positions of the different scatterers in Fig. 2, which must not be too close. The smallest spheres centered at the chosen lattice points **R** and **S**, such that each sphere contains the corresponding scatterer, should not overlap. Indeed, the field is expanded in spherical wave functions centered at these points, in Eq. (9), for example. Each such expansion requires a homogeneous spherical shell region in which to be formed. Apparently, an appropriate horizontal plane should be chosen for the points **R** and **S**, close to the vertical center positions for the two types of scatterers.

It might also be argued that these scatterer enclosing spheres must not intersect horizontal layer boundaries, such as the water-rubber and the rubber-steel interfaces in Fig. 2. It appears that such intersections can be allowed, however, since the expansion in spherical waves is only an intermediate step in forming the R/T matrices, see Eq. (7). Layer thicknesses on each side of the scatterers only appear in phase factors for the elements of the R/T matrices, see Eqs. (5.1) in Ref. 7. It follows that the R/T matrices for a layer are analytic as functions of the layer thickness, and a sufficiently thick layer can be envisaged initially.

Cases with scatterers that themselves intersect a horizontal interface between media with different properties are not

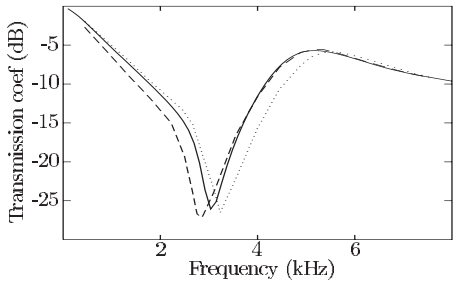


FIG. 3. Transmission coefficients in decibel as functions of frequency for a certain 20 mm thick panel, with cylindrical inclusions as described in the text, immersed in water. The dotted- and dashed-line curves are FEM results from Fig. 9 in Ref. 15, obtained with mesh sizes $\lambda/4$ and $\lambda/12$, respectively, where λ is the wavelength. The solid-line curve is obtained with the amended LMS method of the present paper, with a superellipsoid approximation of the cylindrical inclusions.

considered in the present paper. The expansions in plane waves, needed for the R/T matrices to be formed, require homogeneous horizontal sublayers above and below, as parts of a layer containing scatterers.

E. Validation

A number of tests have been made to validate the described computational procedures and the program code implementation. Concerning wave propagation through phononic crystal slabs with spherical scatterers, comparisons to several examples in the literature have been made successfully, for example, to those in Ref. 23. The procedure from Sec. II C, to speed up computations by utilizing symmetry, has been validated in the obvious way by checking that the results remain the same when it is introduced.

The implementation of the T-matrix routine for nonspherical scatterers has been tested by comparison to a number of relevant examples in Ref. 17. Concerning its integration in the LMS package, Fig. 3 shows the results for an example from Ref. 15, Fig. 9 there, that has subsequently been used by other authors as well.[16,14] The transmission coefficient is computed, as a function of frequency, for plane waves normally incident on a 20 mm thick panel immersed in water with a sound velocity of 1489 m/s. The density of the panel material is 1100 kg/m^3, and its compressional- and shear-wave velocities are 1504.6 and 210.69 m/s, respectively. (There appears to be a misprint in Ref. 15 concerning Young's modulus,[14] which should be 1.4×10^8 Pa rather than 1.4×10^9 Pa.) The absorption is 6.195 dB/wavelength for compressional as well as shear waves. In the middle of the panel, there is a square lattice of period $d=30$ mm with cylindrical air-filled (density of 1.2 kg/m^3 and sound velocity of 340 m/s) inclusions with a radius of 10 mm and a height of 15 mm.

There are three curves in Fig. 3. The dotted- and dashed-line ones are the FEM results from Fig. 9 in Ref. 15, obtained with different mesh sizes. The solid-line curve is obtained with the amended LMS method of the present paper, with the cylindrical inclusions approximated by superellipsoids with parameters $a=7.5$ mm, $b=10$ mm, and $p=8$, as defined in Eq. (8). A satisfactory agreement is obtained, see Fig. 9 of Ref. 16 and Fig. 8 of Ref. 14.

Concerning the mixture of different scatterer types from Sec. III B, an obvious test is to make computations in two different ways for the case with identical parameters for the two scatterer types. An alternative expression to Eq. (12) for the matrix Q has also been used, involving four Ω matrices for axes parallel but rectangular lattices.

A sphere, vertically displaced from the coordinate system origin, is very useful for various validation purposes. The transition matrix can be computed in two different ways: by the numerical T-matrix routine for nonspherical scatterers and by the translation formulas of Sec. III C (starting from analytical formulas for a centered sphere). R/T matrices have also been computed and compared for different vertical locations of a hypothetical interface for spherical scatterers.

IV. ABSORPTION AND SCATTERING CROSS SECTIONS FOR NONSPHERICAL SCATTERERS

According to classical papers by Meyer *et al.*[24] and Gaunaurd *et al.*,[25] a spherical cavity in an elastic solid with a reasonably small shear-wave velocity has a monopole scattering resonance, for the radial displacement amplitude at the sphere surface, when the circumference equals two shear-wave wavelengths in the solid. This resonance is important in connection with Alberich anechoic coatings, as discussed in Ref. 9, for example.

A natural question is how this resonance is changed when the shape of the cavity is changed from spherical to (super)ellipsoidal. Some observations were given already in Ref. 24, based on measurements for rotationally symmetric cavities with ratios between the largest and smallest principal radii not exceeding 2. It was claimed that the resonance frequency is roughly proportional to the inverse of the cube root of the cavity volume.

The T-matrix method has been used to study the scattering and resonance frequencies of nonspherical objects in various circumstances.[26,27] To the author's knowledge, however, it has not been used to theoretically verify the mentioned experimental findings of Meyer *et al.* A few examples are now given, relevant for the design cases in Sec. V.

A rubber material is chosen with a density of 1000 kg/m^3, and compressional- and shear-wave velocities of 1500 and 300 m/s, respectively. Three different kinds of cavities are considered, all with the same height of 8 mm: (a) a sphere with a radius of 4 mm, (b) an ellipsoid with parameters $a=4$ mm, $b=8$ mm, and $p=2$, as defined in Eq. (8), and (c) a superellipsoid with parameters $a=4$ mm, $b=8$ mm, and $p=4$. They are illustrated in Fig. 4. The cavity volumes are (a) 268.1 mm^3, (b) 1072.3 mm^3, and (c) 1405.9 mm^3. It can be noted that the superellipsoid with $p=4$ almost fills an enclosing cylindrical section with a radius of 8 mm and a height of 8 mm, which has a volume of 1608.5 mm^3.

The referred results of Meyer *et al.* predict resonances at 23.9 kHz for the sphere, approximately 15.0 kHz for the ellipsoid, and approximately 13.7 kHz for the superellipsoid. Interestingly, this agrees reasonably well with the solid-line scattering cross-section curves (a), (b), and (c) in Fig. 5, computed with the T-matrix method. A plane compressional wave is incident from above, propagating downward in the direction of increasing z, see Eq. (8). The scattering cross

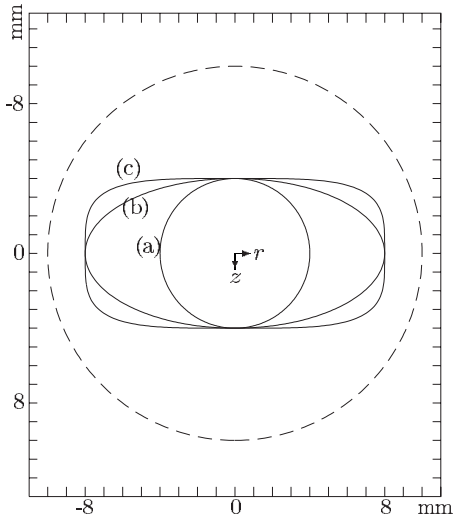


FIG. 4. The shapes of the scatterers considered in Sec. IV are shown: (a) a sphere with radius 4 mm, (b) an ellipsoid with parameters $a=4$ mm, $b=8$ mm, $p=2$, as defined in Eq. (8), and (c) a superellipsoid with parameters $a=4$ mm, $b=8$ mm, $p=4$. The dashed curve shows an enclosing sphere with radius 10 mm.

sections in the figure are all given relative to the area 64π mm^2, which agrees with the geometrical cross-sectional area for the ellipsoid and superellipsoid cavities.

No absorption is involved for the solid-line curves of Fig. 5. Including a rubber shear-wave absorption of 20 dB/wavelength, the resonance effect is not clearly seen at long ranges anymore, and the dashed-line scattering cross-section curves are obtained.

With application to anechoic coatings in mind, the anelastic absorption in the vicinity of the scatterers is of great interest. Absorption cross sections are shown in Fig. 6, for the three chosen cavity types, for a rubber shear-wave absorption of 20 dB/wavelength. They were computed using the T-matrix method in connection with formulas such as Eq. (24) of Ref. 9. As before, a plane compressional wave is incident from above, and the cross sections are given in units of 64π mm^2. Interaction with the incident field is included, in contrast to the scattering cross sections of Fig. 5, which only involve the scattered field.

In comparison to the resonance frequencies seen in the solid-line curves in Fig. 5, the frequencies of maximum ab-

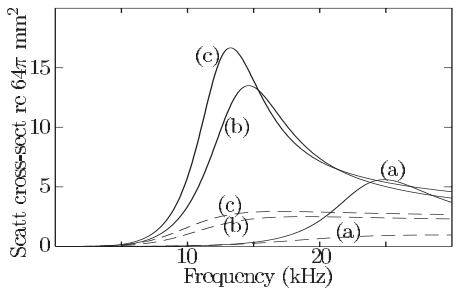


FIG. 5. Total scattering cross-sections for the three cavity types (a), (b), and (c) from Fig. 4. They are given in units of the geometrical cross-sectional area 64π mm^2 for the ellipsoid and superellipsoid cavities. A compressional plane wave is incident from above. There is no absorption for the solid-line curves, whereas the dashed-line ones concern cases with a shear-wave absorption of 20 dB/wavelength.

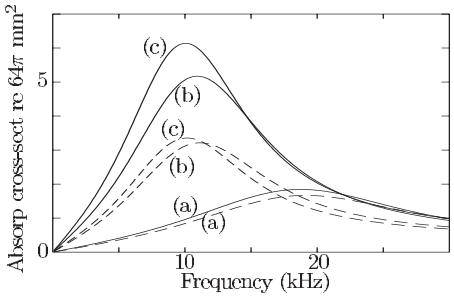


FIG. 6. Absorption cross sections for the three cavity types (a), (b), and (c) from Fig. 4. The total absorption cross sections are shown by the solid-line curves, whereas the dashed-line ones only concern the losses suffered within a surrounding sphere with radius 10 mm.

sorbance from Fig. 6 are shifted slightly downward. The peak frequency is still decreased as the volume of the scatterer is increased.

The dashed-line curves of Fig. 6 show how much of the loss that is suffered within the enclosing sphere of radius 10 mm that is depicted in Fig. 4. As expected from Ref. 9, most of the lost energy is absorbed close to the scatterer.

The underlying resonance concerns the monopole part of the field.[24,25] By redoing the calculations for Fig. 6, this time including only the $l=0$ term [in Eq. (24) of Ref. 9, for example], very similar results are obtained. It is important to note the significance of the shear-wave properties of the medium for generating losses in a monopole field, even when a plane compressional wave propagates without losses. As noted in Ref. 9, in connection with Eq. (41) there, the local time-averaged loss density, or strain power density, includes the imaginary part of the shear modulus as a factor.

A. Implications for coating design

Some observations can now be made concerning the application to Alberich anechoic coatings. The resonance frequency of an individual cavity can be lowered by increasing its volume or by decreasing the rubber shear-wave velocity. Hence, for a fixed coating layer thickness and a fixed rubber shear-wave velocity, oblate ellipsoidal or superellipsoidal shapes seem appropriate to get a thin coating for low frequencies. For the example with a cavity height limited to 8 mm and a rubber shear-wave velocity of 300 m/s, Figs. 5 and 6 show that a coating with spherical cavities would have much higher single-cavity resonance frequencies.

The dashed-line curves (b) and (c) in Fig. 6, for the ellipsoidal and superellipsoidal cavities, show normalized "close-range" absorption cross sections around 3 at 10 kHz. In principle, each cavity could thus take care of the energy of an incident plane wave within an area of $3\times64\pi$ mm^2, indicating that a cavity lattice with period about 25 mm could be appropriate to reduce reflections at 10 kHz. This is only a gross estimate, however, since a simple spherical close-range surrounding has been chosen and important effects have been neglected, such as reflections from the backing, lattice resonances, and other effects of multiple scattering in the lattice. In practice, full-wave computations are needed for coating design, as illustrated in Sec. V.

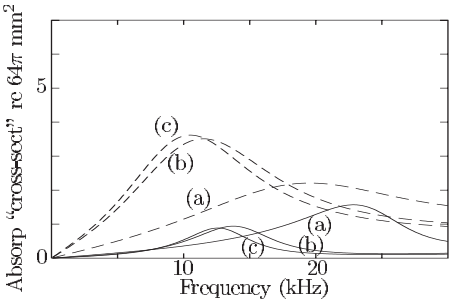


FIG. 7. Close-range absorption cross sections for the three cavity types (a), (b), and (c) from Fig. 4 when compressional-wave absorption of 10 dB/wavelength is included. The absorbed power within a surrounding sphere with radius 10 mm is divided by a reference intensity, taken as the incident wave intensity at the horizontal center level $z=0$. There is only compressional-wave absorption for the solid-line curves, whereas a rubber shear-wave absorption of 20 dB/wavelength is included for the dashed-line ones.

B. Including absorption for compressional waves

Strictly speaking, shear-wave absorption must be accompanied by some absorption for the compressional waves in order to meet a physical constraint for the bulk modulus.[9] In the case discussed, 0.83 dB/wavelength would have been needed. Even at the highest frequency considered (30 kHz), the wavelength of the compressional waves (50 mm) is reasonably large in comparison to the dimensions in Fig. 4, and the omission of this loss can be justified. It has facilitated the discussion of cross sections, which would have been less clear with absorption loss already in an incident compressional wave.

Nevertheless, compressional-wave absorption can be useful to reinforce the effect of the shear-wave losses and to widen the frequency band for the losses. Except when the absorption is weak, and perturbation techniques can be used, it appears difficult to analytically separate the effects of compressional- and shear-wave absorptions, even for simple monopole fields. Hence, an example is here considered, Fig. 7, for which a rubber compressional-wave absorption of 10 dB/wavelength is included. The spherical, ellipsoidal, and superellipsoidal cavities from Fig. 4 are still used. With losses already in the incident compressional wave, the incident plane-wave intensity now decreases with increasing z and the definition of an absorption cross section becomes less obvious. The "close-range cross sections" in Fig. 7 concern the absorbed power within the 10 mm radius sphere surrounding a cavity, divided by the incident wave intensity at the horizontal center level $z=0$.

Because of the losses for the incident wave alone, the absolute "cross-section" levels in Fig. 7 should not be compared directly to those in Fig. 6. The six curves can be compared among themselves, however, since the same compressional-wave absorption of 10 dB/wavelength is used for all of them. No shear-wave absorption is introduced for the solid-line curves, whereas the shear-wave absorption is 20 dB/wavelength for the dashed-line ones. It appears from these examples that the compressional-wave absorption provides favorable basic loss conditions, for subsequent inclusion of the crucial shear-wave absorption.

For simplicity, only vacuum cavities are considered in the examples. In reality, the cavities may be filled by air. As shown by complementary computations, the corresponding results are virtually identical to those presented, except that some narrowband peak may appear due to the resonance scattering from the air inclusion.

V. NUMERICAL COATING DESIGN USING DIFFERENTIAL EVOLUTION

Global optimization methods can be used to design anechoic coatings. Simulated annealing, genetic algorithms, and DE are three kinds of such methods that have become popular during the past 15 years. DE, which is applied here, is related to genetic algorithms, but the parameters are not encoded in bit strings, and genetic operators such as crossover and mutation are replaced by algebraic operators. For applications to underwater acoustics, DE has been claimed to be much more efficient than genetic algorithms[12] and comparable in efficiency to a modern adaptive simplex simulated annealing algorithm.[28]

Realistic rubber materials, for the coatings, can be complicated as concerns the dependence of the acoustical material parameters on temperature and frequency. The purpose here is only to provide illustrative computational design examples, with simple frequency-independent search intervals for some important rubber parameters. It is well known that the shear-wave absorption losses in rubber can be significant. For the Buna-N compound B-5, Nolle[29] reported loss factors (quotients between imaginary and real modulus parts) exceeding 1 for Young's modulus at temperatures around 20 °C and frequencies higher than 5 kHz. Loss factors of this size have later been used in connection with finite-element modeling of resonant sound absorbers.[16] It follows that shear-wave losses up to about 25 dB/wavelength, as considered below, are not unrealistically high.

The objective function in the examples, for DE minimization, is specified as the maximum reflectance in the frequency band 8–22 kHz. Reflectance is here time- (and space-) averaged reflected energy flux relative to the time-averaged energy flux of a normally incident monofrequency plane compressional wave. Only the rubber, modeled as a viscoelastic solid, is anelastic. Parameters for steel are fixed at 5850 and 3230 m/s for the compressional- and shear-wave velocities, respectively, and 7700 kg/m^3 for the density. The sound velocity of water is 1480 m/s.

All the time, all cavities are placed at the bottom of the coating, in direct contact with the steel plate, since this is preferable from a construction point of view. Except for some complementary computations with spherical cavities, the shape parameter p, from Eq. (8), is fixed at 4 for all cavities. Such cavities almost fill enclosing cylindrical sections, see Fig. 4. With two types of cavities present, they differ in size but not in shape. Specifically, the quotient a/b, where a and b are parameters from Eq. (8), is the same for both types.

A. Minimal rubber compressional-wave absorption

The solid-line curves of Fig. 8 concern two cases, (a) and (b), of optimized coatings with superellipsoidal cavities.

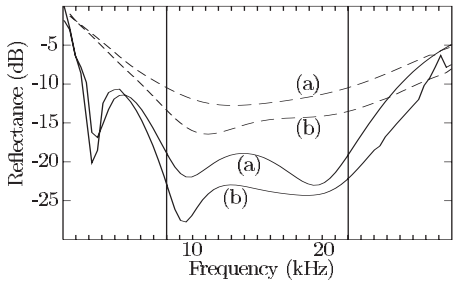


FIG. 8. Reflectances in decibel as functions of frequency obtained by DE coating optimization. Case (a) involves one cavity type as in Fig. 1, whereas case (b) involves two cavity types as in Fig. 2. Superellipsoidal and spherical cavities are used for the solid-line and dashed-line results, respectively. The coatings are specified in Table I and in the text. The largest reflectance within the band 8–22 kHz, marked with vertical lines, is minimized. Some 25 000 coating models were tested during the optimization for each case.

TABLE I. Specification of optimized coatings with minimal rubber compressional-wave absorption, for the solid-line curves of Fig. 8. The parameters are lattice period $p_1=d$, coating thickness p_2, fraction of coating thickness between water and closest cavity surface p_3, superellipsoid cavity shape $p_4=a/b$ from Eq. (8), quotient between the length-scale sizes of the smallest and the largest cavities p_5, rubber density p_6, rubber shear-wave velocity $p_7=\beta$, absorption p_8, and rubber compressional-wave velocity $p_9=\alpha$.

	Case (a)	Case (b)
Optimum (dB)	−18.9	−22.2
$p_1=d$ (mm)	46.8	53.4
p_2 (mm)	11.9	12.0
p_3	0.20	0.20
$p_4=a/b$	0.41	0.40
p_5	⋯	0.40
p_6 (kg/m^3)	1214.3	1257.2
$p_7=\beta$ (m/s)	332.3	280.3
p_8 (dB/wavelength)	25.0	25.0
$p_9=\alpha$ (m/s)	1427.1	1400.7

As in Fig. 1, but with the superellipsoidal shape, only one cavity type is considered for case (a). Case (b) involves a mixture of two types as in Fig. 2.

For the solid-line (b) case of Fig. 8, nine parameters, denoted $p_1, p_2, \ldots, p_9$, are varied within the following search space: lattice period [$p_1=d$, 11.3–113.2 mm], coating thickness [p_2, 8–12 mm], fraction of coating thickness between water and closest cavity surface [p_3, 0.2–0.8], cavity shape parameter [$p_4=a/b$ from Eq. (8), 0.4–1.0], quotient between the length-scale sizes of the smallest and the largest cavities [p_5, 0.3–1.0], rubber density [p_6, 900–1300 kg/m^3], rubber shear-wave velocity [$p_7=\beta$, 60–400 m/s], absorption [p_8, 1–25 dB/wavelength], and rubber compressional-wave velocity [$p_9=\alpha$, 1400–1600 m/s]. The compressional-wave absorption for the rubber is in each case set to the least value consistent with the physical constraint for the bulk modulus. Special constraints are additionally included to avoid too large cavities in relation to the lattice period.

The small tilted lattice used in Sec. III B, which includes all scatterers regardless of type, has the period $d'=d/\sqrt{2}$, with variation within 8–80 mm for case (b). Hence, the parameter p_1 is varied within 8–80 mm for case (a), in order to allow the same scatterer density. Moreover, the search interval for the geometrical parameter p_3 is extended to 0.2–0.94 for case (a), in order to allow as small scatterers as in case (b). The third and final parameter, by which the two cases differ, is p_5, which is irrelevant, of course, for case (a).

Table I specifies the optimal coatings for the solid-line cases of Fig. 8. The inclusion of a second type of cavities in case (b) brings about a reduction of the maximum reflectance within the chosen 8–22 kHz band with more than 3 dB, from −18.9 to −22.2 dB.

For several parameters (p_2, p_3, p_4, p_8, and perhaps p_9), the optimal value appears at or close to the border of the corresponding search interval. It is not surprising that a thick coating (large p_2) with high absorption (large p_8) is preferred. Within the given search bounds, it also appears that big cavities (small p_3 and p_4) are useful. In order to avoid cavity collapse, the fraction of the coating thickness between the water and the closest cavity surface, parameter p_3, should not be too small. With $p_2=12$ mm and $p_3=0.2$, its lower limit, there is everywhere 2.4 mm of rubber covering the cavities toward the water. The height of the (big) cavities is about 12−2.4=9.6 mm and, with $p_4=a/b=0.4$, the horizontal diameter of these cavities is about 24 mm. With $p_5=0.4$, the height of the small cavities for the optimal case (b) coatings is about 3.8 mm and their horizontal diameter is about 9.6 mm. Table I shows that a small rubber compressional-wave velocity, parameter p_9, is preferable. Nevertheless, the acoustic impedance of the rubber is larger than that of the water, since the rubber density, parameter p_6, is rather large.

The lattice period p_1, the cavity length-scale size quotient p_5 [relevant for case (b) only], and the rubber shear-wave velocity p_7 have optimal values in the interior of their corresponding search intervals. With lattice period $d=53.4$ mm for case (b), the corresponding value of d' is about 37.8 mm. Hence, the distance between adjacent cavities is smaller in case (b) than in case (a). The resonance frequency of an isolated cavity of a given size is related to the rubber shear-wave velocity p_7, as discussed in Sec. IV. Insertion of the numerical values verifies that the single-cavity resonance frequencies for cases (a) and (b) fall within or close to the chosen band 8–22 kHz.

The rubber compressional-wave absorption for the optimal coatings are 1.23 dB/wavelength for case (a) and 0.91 dB/wavelength for case (b). Indeed, these are the smallest values that are consistent with the physical constraint on the bulk modulus.

The dashed-line cases (a) and (b) of Fig. 8 concern additional design computations, for which the cavities are restricted to be spherical in shape. A conspicuous difference in the results is that the rubber shear-wave velocities for the optimal coatings, $p_7=\beta$, are only about 55%–60% of the corresponding velocities in Table I, specifically 190.3 and 159.1 m/s for cases (a) and (b), respectively. This is readily understandable from the observations in Sec. IV A. Since the spherical cavities cannot expand in the lateral direction, they get smaller volumes than the superellipsoids, and lower rubber shear-wave velocities are needed to keep the cavity resonance frequencies at the frequency band of interest (8–22 kHz). Moreover, the spherical cavities for these coat-

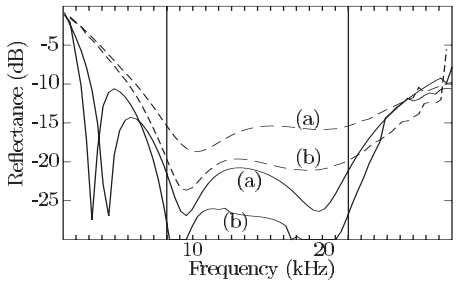


FIG. 9. Reflectances for coating design examples as in Fig. 8, but with variable compressional-wave absorption (parameter p_{10}) included for the rubber. The coatings are specified in Table II and in the text.

ings are more densely spaced: $p_1=d=30.1$ and 35.0 mm for cases (a) and (b), respectively. The remaining parameter values are $p_2=12.0$ mm, $p_3=0.20$, $p_8=25.0$ dB/wavelength, and $p_9=1400$ m/s for (a) as well as (b), $p_6=1143.5$ kg/m^3 for case (a), and $p_6=1171.7$ kg/m^3 and $p_5=0.42$ for case (b). As seen from Fig. 8, the optima are some 8 or 9 dB worse than the corresponding ones for the superellipsoidal cavities.

B. Variable rubber compressional-wave absorption

Including a tenth parameter p_{10} for the rubber compressional-wave absorption, with search interval 0–10 dB/wavelength, the results of Fig. 9 are obtained. As in Fig. 8, there are two solid-line cases, (a) and (b), of optimized coatings with superellipsoidal cavities of one and two types, respectively. The specification details for these optimal coatings are provided by Table II. This time too, the inclusion of a second type of cavities in case (b) brings about a significant reduction of the maximum reflectance within the chosen 8–22 kHz band. The improvement is about 5 dB, from −20.8 to −26.0 dB.

As expected from Sec. IV B, the additional compressional-wave absorption makes possible an enhanced reflection damping within the band 8–22 kHz. Compared to Fig. 8 and Table I, the improvements for case (a) and case (b) are about 1.9 and 3.8 dB, respectively.

TABLE II. Specification of optimized coatings with variable rubber compressional-wave absorption, for the solid-line curves of Fig. 9. In addition to the parameters from Table I, the rubber compressional-wave absorption p_{10} now appears.

	Case (a)	Case (b)
Optimum (dB)	−20.8	−26.0
$p_1=d$ (mm)	49.9	60.5
p_2 (mm)	12.0	12.0
p_3	0.29	0.22
$p_4=a/b$	0.40	0.41
p_5	⋯	0.44
p_6 (kg/m^3)	1225.0	1225.8
$p_7=\beta$ (m/s)	265.2	239.9
p_8 (dB/wavelength)	25.0	23.0
$p_9=\alpha$ (m/s)	1400.9	1413.8
p_{10} (dB/wavelength)	10.0	10.0

Since a large compressional-wave absorption for the rubber is preferable, the optimal value of p_{10} is as high as possible (10 dB/wavelength). Except for p_3, with current optimal values somewhat larger than the lower limit 0.2, the observations from Sec. V A apply concerning the appearance of optimal values at the border or in the interior of the search intervals. The lattice period is slightly larger now, $d=49.9$ mm for case (a) and $d'=d/\sqrt{2}=42.8$ mm for case (b). It seems reasonable that an "effective absorption cross section" for a cavity is increased when more rubber absorption is introduced, implying that a less dense cavity lattice is preferred.

As in Fig. 8, there are two dashed-line cases (a) and (b) in Fig. 9, for additional design computations restricting the cavities to be spherical in shape. Again, the rubber shear-wave velocities for the optimal coatings are small, 160.4 and 139.9 m/s for cases (a) and (b), respectively, or about 60% of the corresponding velocities for the coatings with superellipsoidal cavities. The spherical cavities are more densely spaced: $p_1=d=37.2$ and 42.6 mm for cases (a) and (b), respectively. The remaining parameter values are $p_2=12.0$ mm, $p_3=0.20$, $p_8=25$ dB/wavelength, $p_9=1401$ m/s, and $p_{10}=10.0$ dB/wavelength for (a) as well as (b), $p_6=1179.7$ kg/m^3 for case (a), and $p_6=1160.6$ kg/m^3 and $p_5=0.45$ for case (b). This time, the optima improve by some 6 dB when changing from spherical to superellipsoidal cavities.

The numerical errors in Figs. 8 and 9 appear to be small in relation to the scale of the figures. Error estimates are obtained by redoing the computations with different truncations for the included angular orders l, for Eqs. (6) and (11), and reciprocal lattice vectors, according to Eq. (4).

VI. CONCLUSIONS

The LMS method, a fast semianalytical method for studying wave propagation through layers including scatterer lattices, is extended in Sec. III to handle scatterers of nonspherical shapes and scatterers of different types in the same layer. In the latter case, depth differences within the layer can be allowed for the centers of the scatterers of the different types. Possibilities to split certain equation systems can be lost, however. Symmetry properties are used to reduce the number of unknowns and thereby the equation system sizes (Sec. II).

To make the extensions, null-field type algorithms are incorporated for efficient numerical computation of transition matrices for rotationally symmetric scatterers, and general translation formulas for spherical wave functions are utilized. In principle, any available method could be used to compute the T-matrices. The null-field technique is best adapted to cases where the scatterer shape does not deviate too much from a sphere. Referring to the superellipsoid equation (8), more terms must be included in the field expansions, leading to longer computation times, as the parameter p or the parameter quotient $\max(a,b)/\min(a,b)$ is increased. For the examples, p is set to 2 or 4 and $\max(a,b)/\min(a,b)\leq 2.5$.

The extended LMS computational method is used in Sec. V to model and optimize the performance of Alberich anechoic coatings for reduction of underwater sound reflections. Such a coating consists of a rubber layer with periodically located cavities. Localized resonances by the individual cavities, as modulated by multiple-scattering effects from the lattice, are crucial for the anechoic effect.[9] Early measurements[24] indicate that the single-cavity resonance frequency is inversely proportional to the cube root of the scatterer volume, which is supported by a few computational examples in Sec. IV. For a fixed rubber shear-wave velocity, an oblate superellipsoidal cavity shape can be useful to achieve favorable low-frequency performance of an Alberich anechoic coating, while keeping it reasonably thin.

Global optimization techniques from inverse theory, such as differential evolution, which is used for the optimizations in Sec. V, are useful for numerical coating design. A fast forward model, as provided by the LMS method, is highly desirable. The examples illustrate that the broadband performance can be improved by mixing cavities of different sizes, and by using rubber materials with compressional as well as shear wave absorptions.

ACKNOWLEDGMENTS

Anders Boström kindly provided his computer code for numerical computation of transition matrices. Mikael Odenberg provided advice and encouragement during intellectually inspiring walks around the beautiful Djurgården island.

[1]Z. Liu, C. T. Chan, and P. Sheng, "Three-component elastic wave band-gap material," Phys. Rev. B **65**, 165116 (2002).
[2]K. Maslov, V. K. Kinra, and B. K. Henderson, "Elastodynamic response of a coplanar periodic layer of elastic spherical inclusions," Mech. Mater. **32**, 785–795 (2000).
[3]Z. Liu, X. Zhang, Y. Mao, Y. Y. Zhu, Z. Yang, C. T. Chan, and P. Sheng, "Locally resonant sonic materials," Science **289**, 1734–1736 (2000).
[4]M. Hirsekorn, P. P. Delsanto, A. C. Leung, and P. Matic, "Elastic wave propagation in locally resonant sonic material: Comparison between local interaction simulation approach and modal analysis," J. Appl. Phys. **99**, 124912 (2006).
[5]M. Sigalas, M. S. Kushwaha, E. N. Economou, M. Kafesaki, I. E. Psarobas, and W. Steurer, "Classical vibrational modes in phononic lattices: Theory and experiment," Z. Kristallogr. **220**, 765–809 (2005).
[6]R. Sainidou, N. Stefanou, I. E. Psarobas, and A. Modinos, "The layer multiple-scattering method applied to phononic crystals," Z. Kristallogr. **220**, 848–858 (2005).
[7]I. E. Psarobas, N. Stefanou, and A. Modinos, "Scattering of elastic waves by periodic arrays of spherical bodies," Phys. Rev. B **62**, 278–291 (2000).
[8]Z. Liu, C. T. Chan, P. Sheng, A. L. Goertzen, and J. H. Page, "Elastic wave scattering by periodic structures of spherical objects: Theory and experiment," Phys. Rev. B **62**, 2446–2457 (2000).
[9]S. Ivansson, "Sound absorption by viscoelastic coatings with periodically distributed cavities," J. Acoust. Soc. Am. **119**, 3558–3567 (2006).
[10]S. Ivansson, "Reflections from steel plates with doubly periodic anechoic coatings," in *Theoretical and Computational Acoustics 2005*, edited by A. Tolstoy, E.-C. Shang, and Y.-C. Teng (World Scientific, Hackensack, NJ, 2006), pp. 89–98.
[11]H. Zhao, Y. Liu, J. Wen, D. Yu, and X. Wen, "Tri-component phononic crystals for underwater anechoic coatings," Phys. Lett. A **367**, 224–232 (2007).
[12]C. van Moll and D. G. Simons, "Improved performance of global optimisation methods for inversion problems in underwater acoustics," in Proceedings of the Seventh European Conference on Underwater Acoustics, edited by D. G. Simons, Delft, 2004, pp. 715–720.
[13]V. V. Tyutekin, "Optimization of the characteristics of a sound absorber synthesized from mechanical resonators," Acoust. Phys. **44**, 581–584 (1998).
[14]C. Cai, K. C. Hung, and M. S. Khan, "Simulation-based analysis of acoustic absorbent lining subject to normal plane wave incidence," J. Sound Vib. **291**, 656–680 (2006).
[15]A.-C. Hladky-Hennion and J.-N. Decarpigny, "Analysis of the scattering of a plane wave by a doubly periodic structure using the finite element method: Application to Alberich anechoic coatings," J. Acoust. Soc. Am. **90**, 3356–3367 (1991).
[16]V. Easwaran and M. L. Munjal, "Analysis of reflection characteristics of a normal incidence plane wave on resonant sound absorbers: A finite element approach," J. Acoust. Soc. Am. **93**, 1308–1318 (1993).
[17]A. Boström, "Scattering by a smooth elastic obstacle," J. Acoust. Soc. Am. **67**, 1904–1913 (1980).
[18]J. B. Pendry, *Low Energy Electron Diffraction* (Academic Press, London, 1974).
[19]B. L. N. Kennett, *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, Cambridge, 1983).
[20]R. Sainidou, N. Stefanou, I. E. Psarobas, and A. Modinos, "A layer-multiple-scattering method for phononic crystals and heterostructures of such," Comput. Phys. Commun. **166**, 197–240 (2005).
[21]V. V. Varadan and V. K. Varadan, "Scattering matrix for elastic waves. III Application to spheroids," J. Acoust. Soc. Am. **65**, 896–905 (1979).
[22]A. Boström, "On the systematic use of spherical, cylindrical and plane vector wave functions in elastodynamic scattering problems," in *Acoustic Interactions with Submerged Elastic Structures*, edited by A. Guran, A. Boström, O. Leroy, and G. Maze (World Scientific, Hackensack, NJ, 2002), Chap. 2, pp. 41–79.
[23]R. Sainidou, N. Stefanou, I. E. Psarobas, and A. Modinos, "Scattering of elastic waves by a periodic monolayer of spheres," Phys. Rev. B **66**, 024303 (2002).
[24]E. Meyer, K. Brendel, and K. Tamm, "Pulsation oscillations of cavities in rubber," J. Acoust. Soc. Am. **30**, 1116–1120 (1958).
[25]G. Gaunaurd, K. P. Scharnhorst, and H. Überall, "Giant monopole resonances in the scattering of waves from gas-filled spherical cavities and bubbles," J. Acoust. Soc. Am. **65**, 573–594 (1979).
[26]G. Gaunaurd, "Elastic and acoustic resonance wave scattering," Appl. Mech. Rev. **42**, 143–192 (1989).
[27]C. Feuillade and M. F. Werby, "Resonances of deformed gas bubbles in liquids," J. Acoust. Soc. Am. **96**, 3684–3692 (1994).
[28]H. Gothäll and R. Westin, "Evaluation of four global optimisation techniques (ASSA, DE, NA, Tabu Search) as applied to anechoic coating design and inverse problem uncertainty estimation," M.Sc. thesis, Swedish Defence Research Agency, Stockholm, 2005.
[29]A. W. Nolle, "Dynamic mechanical properties of rubber like materials," J. Polym. Sci. **5**, 1–54 (1950).

A 50 Year comparison of ambient ocean noise near San Clemente Island: A bathymetrically complex coastal region off Southern California

Mark A. McDonald
WhaleAcoustics, 11430 Rist Canyon Road, Bellvue, Colorado 80512

John A. Hildebrand and Sean M. Wiggins
Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093-0205

Donald Ross
8870 Villa La Jolla Drive, La Jolla, California 92037

(Received 31 October 2007; revised 8 July 2008; accepted 8 July 2008)

Repeated ocean ambient noise measurements at a shallow water (110 m) site near San Clemente Island reveal little increase in noise levels in the absence of local ships. Navy reports document ambient noise levels at this site in 1958–1959 and 1963–1964 and a seafloor recorder documents noise during 2005–2006. When noise from local ships was excluded from the 2005–2006 recordings, median sound levels were essentially the same as were observed in 1958 and 1963. Local ship noise, however, was present in 31% of the recordings in 1963 but was present in 89% of the recordings in 2005–2006. Median levels including local ships are 6–9 dB higher than median levels chosen from times when local ship noise was absent. Biological sounds and the sound of wind driven waves controlled ambient noise levels in the absence of local ships. The median noise levels at this site are low for an open water site due to the poor acoustic propagation and low average wind speeds. The quiet nature of this site in the absence of local ships allows correlation of wind speed to wave noise across the 10–220 Hz spectral band of this study.
 [DOI: 10.1121/1.2967889]

PACS number(s): 43.30.Nb, 43.80.Ka, 43.80.Nd [WMC] Pages: 1985–1992

I. INTRODUCTION

A. Background

The noise environment of the ocean is an important aspect of habitat for marine mammals and other organisms, and introduction of human-generated sounds may result in auditory, physiological, or behavioral impacts (National Research Council, 2003 and 2005). Rising levels of ocean noise may negatively impact marine mammals by interfering with their ability to detect sounds, whether these are calls of members of their own species, echoes from prey, or natural sounds that aid in navigation or foraging. Noise may cause generalized stress and affect developmental, reproductive, or immune functions. Noise may exclude animals from areas of critical habitat. Long-term noise measurements in diverse marine environments help to document past changes in noise levels as well as to provide a baseline for future changes. Documenting changes in ocean noise, therefore, is important in understanding the state of the marine environment.

In the deep-water portion of the North Pacific, beyond the continental margins, there has been a steady rise in low frequency ambient noise levels, which is attributed to distant ship traffic corresponding primarily to the increasing propulsion power of the largest commercial ships (Ross, 1993; Andrew *et al.*, 2002; McDonald *et al.*, 2006). It is important to understand whether distant ship traffic also affects ambient noise levels on the continental shelf, an area of critical marine habitat. To study this question changes in ambient noise levels were examined in the relatively shallow coastal waters off Southern California. Shallow water ambient noise studies were conducted at a site west of San Clemente Island in 1958–1960 (Wenz, 1964) and in 1963 (Wenz *et al.*, 1965). The same site was reoccupied in 2005–2006 with a seafloor acoustic recording package (Wiggins, 2003). The earlier study results were compared to the recent results and reexamined in light of current understanding of wind dependent noise, shipping traffic, and biological sound sources.

To understand geographic differences in low frequency shallow water ambient noise levels the contributions of distant and local noise sources must be teased apart as well as differences in seafloor reverberation and acoustic propagation at each measurement site. The primary distant noise source is ship traffic and the primary local noise source is wind driven surface waves. Differences of greater than 10 dB in ambient noise levels under similar wind conditions at different shallow water sites have been attributed to differences in ocean bottom properties, water depths, and sound speed profiles (Ingenito and Wolf, 1989). Vertical hydrophone arrays have been used to separate ship traffic and other distant noise from wind-wave noise (Kuperman and Ferla, 1985; Kewley *et al.*, 1990; Chapman and Cornish, 1993) and to estimate wave source level which can then be used to model omnidirectional noise spectrum levels for given sound speed profiles and seafloor properties. While it should be possible to model the propagation of distant deep-water ship traffic noise onto the continental margins, the approach taken

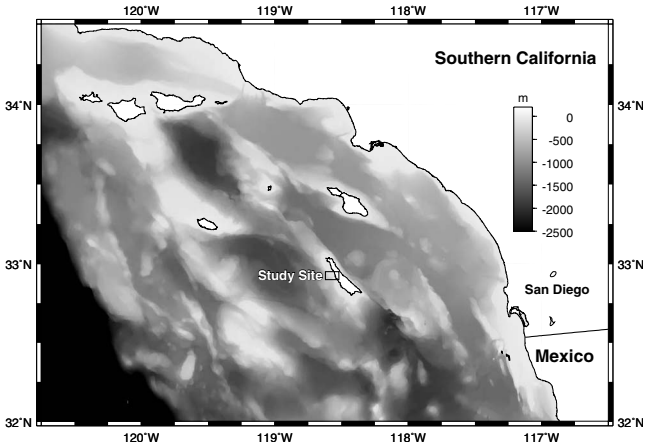


FIG. 1. Bathymetry of the Southern California continental margin showing the ambient noise study site off the west coast of San Clemente Island, CA.

here is to look for any shallow water noise increase at the study site which could be caused by the approximately 15 dB deep-water noise increase over the study time period (McDonald *et al.*, 2006).

B. Study setting

The site for this study is located in the continental borderland region offshore Southern California, along the western flank of San Clemente Island (Fig. 1). The Southern California continental borderland is not a typical continental margin in that the bathymetry is complex and alternates between relatively deep troughs (greater than 1 km) and shallow banks. The distance between the mainland-shore and the oceanic-crustal margin, where full ocean depths (>4000 m) occur, is more than 100 miles for this region. A zone of extended continental crust creates a broad shelf, with alternating banks and troughs created by normal faulting.

The region being studied is important with respect to marine mammals, shipping, and military operations. The study site (Figs. 1 and 2) was chosen because ambient noise was previously characterized at this site. The study site is adjacent to a Navy sonar test range; thus local ship traffic in this region is not necessarily typical of a random sampling off Southern California, as commercial ships may be excluded from the range at times of active naval operations and more Navy ships are expected to operate near this site than at other sites in this region.

II. METHODS

A. Wenz recordings, 1958–1959 and 1963–1964

During 1958–1959 a shore connected cabled hydrophone was deployed near Eel Point, on the western flank of San Clemente Island. The hydrophone was located several meters above the seafloor in 110 m of water over a relatively flat sandy bottom (Wenz, 1964; Wenz *et al.*, 1965). Rockfish and flatfish were observed in seafloor photos taken during the deployment. The location of the hydrophone deployment was determined using bearing angles from the island. The 1958–1959 hydrophone location is estimated to be 32°55.45′ N 118°34.65′ W. A recording anemometer was located onshore approximately 2 miles (3.7 km) from the study site.

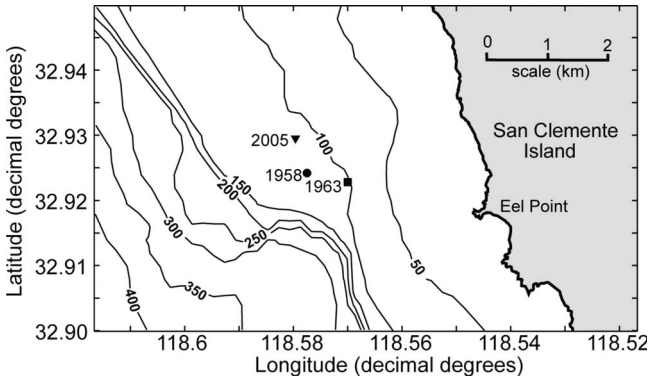


FIG. 2. The locations of the sites occupied in 1958 (circle), 1963 (square), and 2005 (triangle) are shown with bathymetric contours in meters.

During the 1958–1959 time period, four data sets were collected with saved noise spectra being chosen to avoid interference from local ship noise. The data windows were 30 June–31 July (525 spectra), 8 September–19 October (658 spectra), 17 November–22 December (515 spectra), and 3 March–21 May (580 spectra). The first three data groups were analyzed using six one-third octave band filters, where the band level was chosen visually from a strip chart recorder, to avoid transient noises and ship noise. Each spectra sample averaged over 165 s and represented a sample from a different hour. The March–May 1959 data set was analyzed with a more complex system allowing for more one-third octave band levels, the details of which are described by Wenz (1964), but the goal was to have similar data for comparison.

The 1963–1964 data were collected with a different hydrophone system, again deployed near Eel Point, with an estimated location of 32°55.37′ N 118°34.20′ W. The hydrophone was approximately 2 m above the seafloor and an onshore anemometer was again recorded. Several methods, including magnetic tape were used to collect data during this period, but the data used for spectrum comparison were collected with one-third octave filters and strip chart recorders, again discriminating against local ships by eye (Wenz *et al.*, 1965). The magnetic tape records were further analyzed, particularly with regard to blue whale songs (20 Hz longs), although these sounds were not recognized as being from blue whales at that time (Thompson, 1965).

B. Autonomous seafloor recorder, 2005–2006

Recordings were made with an autonomous recording package (ARP) of the same design as that described by Wiggins (2003), and used in previous ambient noise studies (McDonald *et al.*, 2006). Continuous acoustic recordings were collected from 16 August 2005 to 9 February 2006 (179 days) at a sampling rate of 500 Hz. The recorder was deployed at 32°55.768′ N, 118°34.777′ W in a water depth of 110 m, with the hydrophone suspended 10 m above the seafloor. Review of the data revealed no evidence of instrumental problems throughout the recording period. When this same instrument design has been deployed in high current

areas, flow noise and cable strum can be recognized by significant energy below 10 Hz, but during this deployment there was no evidence of current induced flow noise. Wind data relevant to the study site were obtained from the naval air station on San Clemente Island, 10.3 km to the north of the study site.

C. Calibration

Calibration of a similar seafloor acoustic recording package was conducted using reference hydrophones at the U.S. Navy's Transducer Evaluation Center facility in San Diego (TRANSDEC) to verify the theoretical calibration which was based on nominal component specifications. Calibration was conducted from 10 to 250 Hz. Differences between the actual instrument used for measurements at the Eel Point site and the one tested at TRANSDEC are expected to be less than 1 dB, due to slight differences in hydrophone sensitivity and circuitry.

The calibration testing showed the theoretical response of the instrument to be within about 1 dB of the measured response. The seafloor recorder is not expected to have a meaningful response below 2 Hz and absolute calibration is questionable below 10 Hz. The high frequency rolloff of the recorder used begins at 220 Hz and provides −36 dB/octave of protection from aliasing. The noise floor of the instrument is calculated to be near the lowest values observed in this study. Because of possible differences in the antialias filtering between instruments, these data are plotted only to 220 Hz, the start of the antialiasing rolloff.

D. Spectral averaging

The analysis of Wenz (1964) used either 165 or 200 s of data for each spectral average. Wenz (1964) used only three averages per hour, presumably because of data processing limitations, while the 2005–2006 recordings were analyzed using continuous data with no overlap between 200 s spectral averages, processed with a Hann window. The 2005–2006 spectra were calculated as 1 Hz bins, which is a standard procedure for noise spectra, providing more detailed information than one-third octave bands to help identify sound sources. The 2005–2006 spectrum which includes local shipping noise is an average computed continuously over all data from the 179 day deployment using a 200 s average for a total of 62 504 spectra.

To estimate noise levels when local ship noise is not apparent in the data, time intervals of 30 min each were chosen, as near as possible to sunrise, local noon, sunset, and local midnight for each day. When a 30 min time interval free of ship noise was not available a lesser interval of not less than 10 min was used. Sometimes there were no intervals of 10 min or more free of ship noise for as long as several days. The presence of ship noise was determined by examining spectrogram views of the data with 1 Hz spectral bins for either strong tonals associated with ships or for strong cavitation noise. In total, 319 spectra free of local ship noise, averaging near 30 min each in duration, were chosen in this manner.

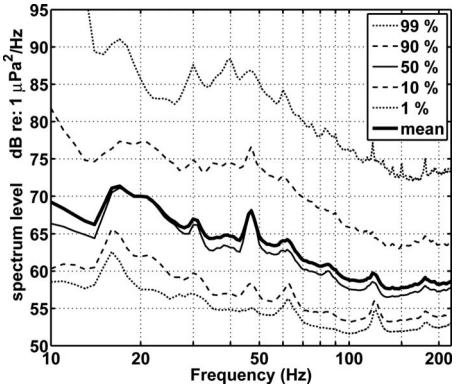


FIG. 3. The 2005–2006 pressure spectrum level distribution at the San Clemente site including local shipping. A total of 62 504 spectra, each 200 s in duration, spanning 179 days was used.

III. RESULTS

The 2005–2006 pressure spectra that include local ship noise reveal average noise levels to be about 70 dB re 1 μPa^2/Hz at 20 Hz, decreasing to about 58 dB re 1 μPa^2/Hz at 200 Hz (Fig. 3). Spectral level cumulative distributions for 2005–2006 are long tailed for higher values, with the ambient noise cumulative 99 percentile about 20 dB above the mean, whereas the 1 percentile is about 10 dB below the mean (Fig. 3). Since Wenz (1964, 1965) did not report 1958 or 1963 spectral levels in the presence of ships, a direct comparison with 2005–2006 spectra including local ship noise is not possible.

Pressure spectrum average levels excluding local ships in 2005–2006 are as much as 10 dB lower than the long-term average including ships (Fig. 4). The differences between

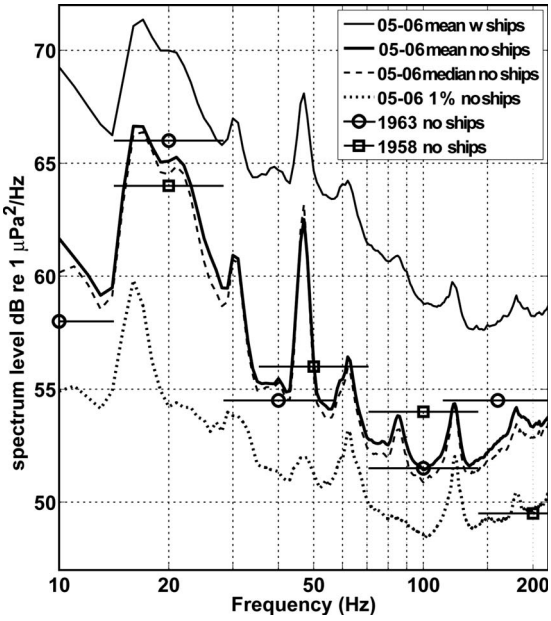


FIG. 4. The 2005–2006 pressure spectrum levels at the San Clemente site excluding local shipping are compared to the spectrum levels from representative one-third octave averages of November-December 1958 (□) and 1963 (○) at 0800 h for the same site with local ship noise excluded and the 2005-2006 mean values with ships included. The horizontal bars on the 1958 and 1963 values indicate the frequency ranges of the one-third octave bands used to compute the values.

noise levels with and without ships are slightly larger at 10–80 Hz (8–10 dB) than for frequencies >100 Hz (5–6 dB). The median and mean spectral levels excluding local ships are essentially the same, as illustrated by the heavy dashed line and heavy solid lines in Fig. 4. The ambient spectrum cumulative 1 percentile distribution without local ship noise (dotted line in Fig. 4) is 3–5 dB below the mean, with levels of 54 dB re 1 μPa^2/Hz at 20 Hz, decreasing to about 48 dB re 1 μPa^2/Hz at 100 Hz.

Ambient pressure spectral levels excluding local ship noise were effectively the same in 1958, 1963, and 2005–2006 (see points in Fig. 4). Data from 1958 and 1963 differ from the 2005–2006 data by less than 5 dB, with equal numbers of the early data points below and above the recent data in noise level. Note that it was necessary to selectively choose 1958 and 1963 data that were not biased upward in noise level due to high winds or high biological noise. The 1958 data were chosen to avoid unusually high winds and the 1963 data were chosen from a time of day that did not contain high levels of fish chorusing.

Both early and recent data sets show mean pressure spectrum levels dominated by the calls of fin and blue whales over the 15–30 Hz band (McDonald *et al.*, 1995). When ships are present, blue and fin whale calls rise about 5 dB above the background noise (see peaks at 15–22 and 48 Hz in the top curve in Fig. 4). When ships are absent, these blue and fin whale call levels are 10 dB above the background noise (bold curve in Fig. 4). Note that in absolute terms, the blue and fin whale spectral peaks are about 5 dB lower when ships are absent, suggesting a responsive change in blue and fin whale source levels in the presence of ship noise.

If one excludes the fin and blue whale calls, the average noise level in 2005–2006, in the absence of ships, slopes from about 60 dB re 1 μPa^2/Hz at 10 Hz to about 53 dB re 1 μPa^2/Hz at 70 Hz. In the 70–220 Hz range the noise level is about 52.5 dB re 1 μPa^2/Hz. Note that noise peaks near 60, 120, and 180 Hz may be related to seafloor electrical power cables and are not thought to be components of the ambient noise field.

IV. ANALYSIS AND DISCUSSION

A. Shipping noise

Comparisons between 1963 and 2005–2006 at the study site suggest that it has experienced increased exposure to local shipping noise. The analysis by Thompson (1965) of eight days of data from July 1963 at this site reports that 31% of the recordings contained ship propulsion sounds. A ship noise analysis of the 179 days of recordings from 2005–2006 was conducted by examining 200 random times, chosen with a random number generator, from throughout the recording interval. Significant local ship noise was present in 89% of the samples to a degree such that these samples could not be used for the local-ship-excluded noise average. The nearby location of the Navy test range to the west of San Clemente Island may contribute local ship noise. This site is not directly on a commercial shipping transit lane and is shielded from ships near Long Beach/Los Angeles and San Diego by the presence of Santa Catalina and San Clemente Islands. However, ships transiting from Long Beach to the southwest (e.g., to Hawaii or Australia) pass nearby this site to the north. Likewise, some local ship traffic may result from fishing activity in this area.

The local-ship-included noise levels are presented in Fig. 3 for comparison to other locations and to this same location in the future. Unfortunately, Thompson (1965) did not present a local-ship-included noise average for comparison. Omnidirectional noise levels at a site 7 nautical miles west of San Diego are reported to be 72 dB re 1 μPa^2/Hz at 62 Hz when local ship noise is included (Heitmeyer *et al.*, 2004). The local shipping included noise level at the site of this study is 64 dB re 1 μPa^2/Hz at 62 Hz (Fig. 3), suggesting greater local ship noise influence near the port of San Diego.

Deep-water ambient noise levels are generally more predictable than coastal ambient noise levels because distant shipping noise typically dominates deep-water noise levels. Shallow water noise levels are highly variable primarily because of differences in local acoustic propagation and seafloor absorption characteristics. The results of this study suggest that distant shipping traffic does not provide a significant contribution to ambient noise at the study site, leaving wind noise, biological noise, and local shipping to determine noise levels.

At many shallow water coastal sites, where acoustic propagation is more efficient over long ranges than at this site, distant shipping provides the major contribution to ambient noise. Many coastal sites show mean spectrum levels at 50 Hz which are 15–20 dB higher than the San Clemente Island site after exclusion of obvious local shipping noise, examples being from the Norwegian Sea (Walkinshaw, 1966), Eastern Canada (Zakarauskas *et al.*, 1990), and the Mediterranean (Kuperman and Ferla, 1985). Even higher levels have been measured in the North Sea and Baltic Sea (Wille and Geyer, 1984). This difference can be attributed to a relatively constant depth flat seafloor allowing for much better propagation of distant shipping noise at these other sites when compared with the irregular bathymetry of the California continental borderlands. A study by Piggott (1964) off Eastern Canada found relatively low noise levels, although still higher than those recorded at the San Clemente Island site. Piggott's (1964) relatively low spectrum levels have been attributed to poor acoustic propagation near his study site (Zakarauskas *et al.*, 1990).

Sites in the Arafura and Timor Sea (Cato, 1976) are the only published results that have ambient noise levels as low as those found in this study. The San Clemente site is also quiet when compared to North Pacific measurements near the axis of the deep sound channel (Wenz, 1969; Andrew *et al.*, 2002; McDonald *et al.*, 2006). Even including the noise from local ships, which are present 89% of the time at the study site, the mean noise levels at this study site are relatively low when compared to most other shallow water sites worldwide.

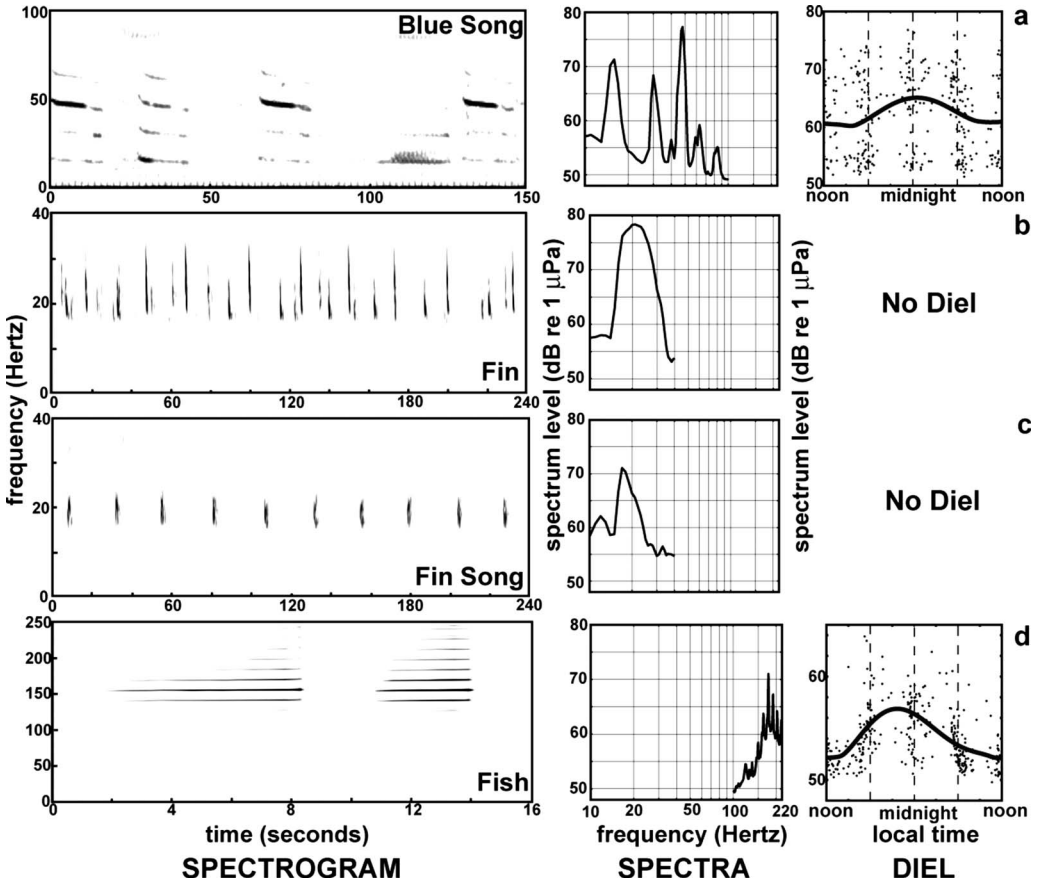


FIG. 5. The four most prominent biological sounds are (a) blue whale song, (b) fin whale countercalling, (c) fin whale song, and (d) a fish chorus referred to as the "motorboat," each spectrogram computed with a 1 s FFT from the 2005–2006 data. Representative spectra are shown for each sound type, computed as a 30 min average over a time period when the given biological sound dominates. The diel variation shown here is the spectrum level of the most distinctive spectral peaks (a) 47.3 and (d) 181 Hz for each sound for the 319 spectral averages.

B. Biological sounds

The strongest seasonal variations in noise at the San Clemente site are due to blue and fin whale calling which appears as a broad spectral peak near 20 Hz (Fig. 5), most prominently in the fall (Thompson, 1965, Burtenshaw *et al.*, 2004). Blue whales are normally present in this region only from June through January, while fin whales are present year-round (Oleson *et al.*, 2007a). The 20 Hz spectral peak results from patterned song calls of blue whales, which increase in occurrence during the fall (Oleson *et al.*, 2007b). The 2005–2006 study period (179 days) coincided with the seasonal occurrence of blue whale song in this region (Burtenshaw *et al.*, 2004), as did the data from 1958 and 1963 (plotted in Fig. 4). The blue whale population off California has been increasing (Calambokidis and Barlow, 2004) though the blue whale song spectrum levels at this site appear more or less the same in 1958, 1963, and 2005–2006. Blue whale singing appears to increase at nighttime [Fig. 5(a)], based on the spectrum level at 48 Hz, the frequency midpoint of the third harmonic of the blue whale song during 2005–2006 (Oleson *et al.*, 2007b). Subtle diel variations, due to peaks in blue whale calling near sunrise and sunset, may be present in these data (Thompson, 1965; Wiggins *et al.*, 2005). Less prominent blue whale sounds include D calls (counter calling), which are occasionally significant in terms of average spectral levels.

The two dominant sounds produced by fin whales in this region include irregularly spaced pulses of about 0.8 s duration [Fig. 5(b)], which are sometimes used as countercalls between traveling fin whales (McDonald *et al.*, 1995; McDonald and Fox, 1999) and similar pulses used in patterned calls [Fig. 5(c)], often with bimodal temporal spacing (Thompson *et al.*, 1992), produced only by males (Croll *et al.*, 2002). These temporally patterned calls are referred to as song and have a stereotypical character. At this site, the fin whale call type changed from dominantly irregular countercalls to dominantly song type calls about the second week of December. No diel pattern was discerned with these data for either the countercall type or the song type calls.

The studies of Wenz (1964), Wenz *et al.* (1965), and Thompson (1965) found at least three different sounds known to be produced by fishes, which at given times and seasons set the ambient noise levels within a frequency band near 150–200 Hz. The 1958 fish sounds differed in frequency and seasonality from the 1963 sounds, suggesting that different species were present. The 2005–2006 data contained only one obvious fish sound [Fig. 5(d)], this showing a strong diel pattern. When a single day of data free of local ship noise is examined the diel pattern is commonly stronger than the mean for all (319) spectral averages as in Fig. 5(d). The diel pattern is asymmetrical within the night, the greatest chorusing occurring shortly after sunset, suggesting a crep-

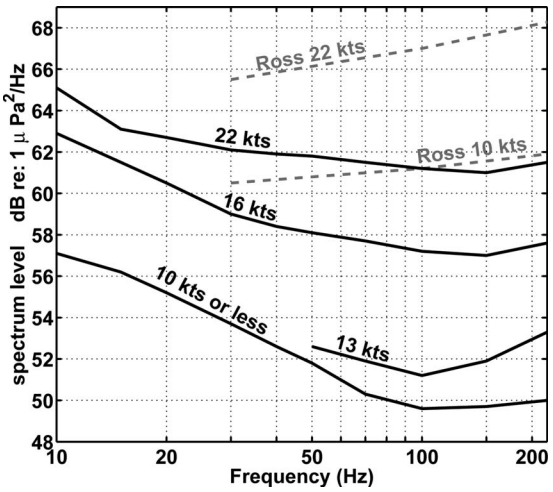


FIG. 6. In the absence of local shipping traffic, surface wind/wave noise is seen across the entire study band when winds exceed about 13 kn. The generalized wind speed/noise curves of Ross (1976) are included for comparison.

uscular pattern. The spectrum presented in Fig. 5(d) shows peaks near 150 Hz, corresponding to what is presumed to be an individual fish near the hydrophone. The precise spectral peaks change from one recording to the next, resulting in an average spectral peak near 180 Hz. The fish sounds in the spectrogram of Fig. 5 are pulsive in nature, a ls duration fast Fourier transform (FFT) resulting in 15 Hz sidebands, corresponding to the 15 pulses per second. The presumed fish sounds described by Thompson (1965) were characterized as a "motorboat" sound because of their similarity to the sound of a single piston small boat motor. A detailed review and comparison of these fish sounds with the current data is beyond the scope of this study, but a 10–20 dB diel pattern with higher intensities during the night was found in 1963.

In January and February gray whale and humpback whale calls and song are present in sufficient quantity to increase the average noise levels over periods of hours, but do not contribute substantially to the ambient pressure spectrum level when averaged over monthly or seasonal time periods. Recordings with greater bandwidth would be better suited for the study of these sounds, as most of the sounds produced by these species occur above the frequencies recorded in this study.

C. Wind driven surface wave noise

The relatively low ambient noise levels that occur at the study site when local shipping noise is absent allow correlation of ambient noise levels to wind speed across the entire 10–220 Hz band considered in this study. The average and median wind speed during the study period was 7 kn, with the highest sustained wind speed being 22 kn. Periods with low biological and ship noise were selected to compute hourly averaged spectra as wind speed changed, particularly selecting for intervals with high wind speeds (Fig. 6). These data suggest a minimum surface wind noise at 100–150 Hz, with about a 6 dB/octave increase in noise level at lower frequencies, and a 1–2 dB/octave increase in noise level at higher frequencies. The correlations in Fig. 6 show notably lower spectral levels than Ross (1976) or Urick (1983). The wind speed noise correlations of Ross (1976) and Urick (1983) peak near 250 Hz and decrease down to where the curves end near 100 Hz (Urick, 1983) or 30 Hz (Ross, 1976). The increase in noise level at the highest frequencies of Fig. 6 corresponds in shape to the plots in Ross (1976) and Urick (1983), but is lower in absolute spectral level.

The 10 km separation between the wind and the ambient noise recording sites for this study may result in some error, although the high winds in this region typically come from the west, with no sheltering effect from the island at this site. While propagation conditions and seafloor absorption characteristics are important to such wind-noise correlations, the data in Fig. 6 are likely better for estimating wind related noise levels in similar water depths in California coastal waters than the deep-water correlations of Ross (1976) or the more generalized correlations of Urick (1983).

Studies of wind noise using vertical hydrophone arrays to separate ship traffic and other distant noise from wind-wave noise (Kewley *et al.*, 1990; Chapman and Cornish, 1993) have estimated source level as opposed to omnidirectional noise spectrum levels versus frequency. The conversion from source level to omnidirectional spectrum level assumes a frequency dependent bottom reflection loss (Burgess and Kewley, 1983). If seafloor reflection loss is large, the conversion is approximately 7 dB, if seafloor reflection loss is 4 dB the conversion is 8.5 dB and for a 2 dB loss the conversion factor is 13.5 dB. The source level at 100 Hz from Kewley *et al.* (1990) and Chapman and Cornish (1993) is 50 dB for a 10 kn wind, predicting a noise level not less than 58 dB at 100 Hz, which is about 8 dB higher than that seen in Fig. 6. This discrepancy might be explained by leakage of near horizontally propagating distant noise into the vertically beamformed noise measurements or it might be that deep-water waves are different than those at this site.

Wenz (1964) plotted wind speed versus spectrum level, but he found as much as 10 dB of unexplained discrepancy in the wind speed to spectrum level correlations from one data set to the next, perhaps due to noise from nonwind related sources, which caused a coincidental correlation in his small data set. Wenz (1964) did not have the technology to readily examine his data in modern spectrogram form to help identify known noise sources.

D. Minimum noise levels

The biological and other-noise lines can be identified and removed from the minimum noise level plot to estimate the background noise level when winds are calm (Fig. 7). The minimum noise level is 55 dB re 1 μPa^2/Hz at 10 Hz, decreasing at about 6 dB/octave at higher frequencies. Above the minimum spectral level are seen three spectral lines at multiples of 60 Hz. These lines vary slightly in frequency and time, but they are not impulsive or diel as may be expected for biological sounds. In addition, the seafloor recorder does not use 60 Hz power, and if these peaks were electronic noise lines associated with the recorder, such broad peaks would not be expected. Their origin remains undetermined, but we speculate that they may be associated

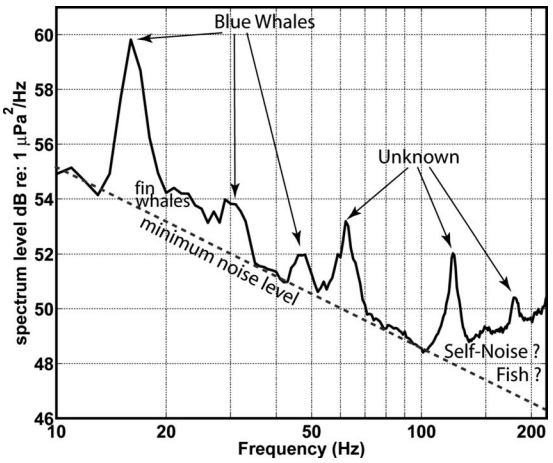


FIG. 7. The lowest one percentile ambient noise curve is shown with noise sources identified. The dashed line is suggested as the minimum noise level in the absence of biological sources. The wind is presumed to be calm at this site for the lowest 1% of noise levels.

with seafloor cables passing near this site, which terminate at San Clemente Island. Above 100 Hz the increased ambient noise seen in Fig. 7 may be due to instrumental self-noise or to the presence of fish chorusing. After removing the biological and other-noise lines, the ambient noise level is 10–15 dB lower than the ambient noise in deep-water offshore of the continental shelf, and as mentioned previously, it has not been increasing over the past 50 years as the deep-water noise levels have.

V. RECOMMENDATIONS

New and better metrics for ship noise should be developed to better examine the question of how local ship noise may be affecting marine mammals. Such metrics would ideally integrate hearing perception curves for each species of interest and percentile plots of noise levels at multiple sites using various averaging intervals appropriate to the hearing integration times of the particular species. For the species which use the frequency band of this study (e.g., blue and fin whales), hearing perception curves have not been measured.

The reoccupations of additional early shallow water ambient noise study sites such as those of Piggott (1964) off Eastern Canada, Walkinshaw (1966) in the Norwegian Sea, and Wenz (1961) in the Bering Sea are needed to better judge long-term changes in ambient sound on continental shelves. These earlier studies also often report the percentage of time local shipping noise was apparent in the data allowing for comparison with more recent recordings.

ACKNOWLEDGMENTS

This work was supported by the US Navy CNO N45 and ONR, and we thank Frank Stone, Ernie Young, Jeff Simmons, and Bob Gisiner.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. (**2002**). "Ocean ambient sound: Comparing the 1960's with the 1990s for a receiver off the California coast," ARLO **3**, 65–70.

Burgess, A. S., and Kewley, D. J. (**1983**). "Wind-generated surface noise source levels in deep water east of Australia," J. Acoust. Soc. Am. **73**, 201–210.

Burtenshaw, J. C., Oleson, E. M., Hildebrand, J. A., McDonald, M. A., Andrew, R. K., Howe, B. M., and Mercer, J. A. (**2004**). "Acoustic and satellite remote sensing of blue whale seasonality and habitat in the Northeast Pacific," Deep-Sea Res., Part II **51**, 967–986.

Calambokidis, J., and Barlow, J. (**2004**). "Abundance of blue and humpback whales in the eastern north Pacific estimated by capture-recapture and line-transect methods," Marine Mammal Sci. **20**, 63–85.

Cato, D. H. (**1976**). "Ambient sea noise in waters near Australia," J. Acoust. Soc. Am. **60**, 320–328.

Chapman, N. R., and Cornish, J. W. (**1993**). "Wind dependence of deep ocean ambient noise at low frequencies," J. Acoust. Soc. Am. **93**, 782–789.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urbán-Ramírez, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809–810.

Heitmeyer, R. M., Wales, S. C., and Pflug, L. A. (**2004**). "Shipping noise predictions: capabilities and limitations," Mar. Technol. Soc. J. **37**, 54–65.

Ingenito, F., and Wolf, S. N. (**1989**). "Site dependence of wind-dominated ambient noise in shallow water," J. Acoust. Soc. Am. **85**, 141–145.

Kewley, D. J., Browning, D. G., and Carey, W. M. (**1990**). "Low-frequency wind-generated ambient noise source levels," J. Acoust. Soc. Am. **88**, 1894–1902.

Kuperman, W. A., and Ferla, M. C. (**1985**). "A shallow water experiment to determine the source spectrum level of wind generated noise," J. Acoust. Soc. Am. **77**, 2067–2073.

McDonald, M. A., and Fox, C. G. (**1999**). "Passive acoustic methods for fin whale population density estimation," J. Acoust. Soc. Am. **105**, 2643–2651.

McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (**2006**). "Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California," J. Acoust. Soc. Am. **120**, 711–718.

McDonald, M. A., Webb, S. C., and Hildebrand, J. A. (**1995**). "Blue and fin whales observed on a seafloor array in the Northeast Pacific," J. Acoust. Soc. Am. **98**, 712–721.

National Research Council (**2003**). *Ocean Noise and Marine Mammals* (National Academies Press, Washington, DC).

National Research Council (**2005**). *Marine Mammal Populations and Ocean Noise: Determining When Noise Causes Biologically Significant Effects* (National Academies Press, Washington, DC).

Oleson, E. M., Calambokidis, J., Barlow, J., and Hildebrand, J. A. (**2007a**). "Blue whale visual and acoustic encounter rates in the Southern California Bight," Marine Mammal Sci. **23**, 574–597.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., and Hildebrand, J. A. (**2007b**). "Behavioral context of call production by Eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.

Piggott, C. L. (**1964**). "Ambient sea noise at low frequencies in shallow water of the Scotian shelf," J. Acoust. Soc. Am. **36**, 2152–2163.

Ross, D. (**1976**). *Mechanics of Underwater Noise* (Pergamon, New York).

Ross, D. (**1993**). "On ocean underwater ambient noise," Institute of Acoustics Bulletin **18**, 5–8.

Thompson, P. O. (**1965**). "Marine biological sound west of San Clemente island," U.S. Navy Electronics Laboratory Report No. 1290, San Diego, CA, p. 42.

Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.

Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York), pp. 202–236.

Walkinshaw, H. M. (**1966**). "Measurements of ambient noise spectra in the south Norwegian Sea," originally published in U. S. Navy J. Underwater Acoust.; reprinted in (**2005**). IEEE J. Ocean. Eng. **30**, 262–266.

Wenz, G. M. (**1961**). "Periodic variations in underwater low-frequency ambient noise levels," U.S. Navy Electronics Laboratory Report No. 1014, declassified 1973.

Wenz, G. M. (**1964**). "Ambient noise measurements west of San Clemente Island," U.S. Navy Electronics Laboratory Report No. 1235, declassified 1970.

Wenz, G. M. (**1969**). "Low-frequency deep-water ambient noise along the Pacific Coast of the United States," U. S. Navy J. Underwater Acoust. **19**, 423–444, declassified 1997.

Wenz, G. M., Calderon, M. A., and Scanlan, T. F. (**1965**). "Underwater acoustic ambient-noise and transmission tests west of San Clemente Island, July 1963," U.S. Navy Electronics Laboratory Report No. 1260, declassified 1994.

Wiggins, S. M. (**2003**). "Autonomous Acoustic Recording Packages (ARPs) for long-term monitoring of whale sounds," Mar. Technol. Soc. J. **37**, 13–22.

Wiggins, S. M., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. (**2005**). "Blue whale (*Balaenoptera musculus*) diel call patterns offshore of Southern California," Aquat. Mamm. **31**, 161–168.

Wille, P. C., and Geyer, D. (**1984**). "Measurements on the origin of the wind-dependent ambient noise variability in shallow water," J. Acoust. Soc. Am. **75**, 173–195.

Zakarauskas, P., Chapman, D. M. F., and Staal, P. R. (**1990**). "Underwater acoustic ambient noise levels on the eastern Canadian continental shelf," J. Acoust. Soc. Am. **87**, 2064–2071.

Numerical modeling of the sound propagation through a rarefied gas in a semi-infinite space on the basis of linearized kinetic equation

Felix Sharipov[a] and Denize Kalempa
Departamento de Física, Universidade Federal do Paraná, Caixa Postal 19044, Curitiba 81531-990, Brazil

(Received 19 March 2008; revised 26 June 2008; accepted 27 June 2008)

The sound propagation through a rarefied gas is investigated on the basis of the linearized kinetic equation. A plate oscillating in the direction normal to its own plane is considered as a sound wave source. It is assumed a fully established oscillation so that the solution of the kinetic equation depends on time harmonically, while its dependence on the spatial coordinates is obtained numerically. The problem is solved over a wide range of the oscillation speed parameter defined as a ratio of the intermolecular collision frequency to the sound frequency. In order to evaluate the influence of the momentum and energy accommodation coefficients on the solution of the problem, the Cercignani–Lampis scattering kernel is applied as the boundary condition. An analysis of wave characteristics near the source surface shows that they are significantly different from those far from the surface even if the oscillation is slow, i.e., the solution is not harmonic in the space.
 [DOI: 10.1121/1.2967835]

PACS number(s): 43.35.Ae, 43.20.Bi [RR] Pages: 1993–2001

I. INTRODUCTION

The classical theory of sound propagation is based on the continuum mechanics equations,[1,2] which are valid under the two following conditions. First, the characteristic length of fluid flow must be significantly larger than the molecular mean free path. Second, the characteristic time of fluid flow variation should be significantly larger than the molecular mean free time. According to the theory of sound propagation through gases based on the Navier–Stokes equations, the phase speed and attenuation coefficient of a plane wave are constant over the space. In other words, the wave is harmonic in the space and time. However, if the sound frequency is close to the intermolecular collision frequency the second condition of the continuum mechanics applicability is broken. In this case, the sound propagation is qualitatively different from that in the continuum limit and must be considered on the basis of the kinetic equation.

To characterize the speed of oscillation the so-called oscillation parameter is introduced defined as the ratio of the intermolecular collision frequency to that of the sound. According to this parameter three oscillation regimes can be established to the gas flow. First, the low oscillation speed regime in which there are many intermolecular collisions occurring during one cycle of the oscillation; in this regime the classical theory of sound propagation based on the Navier–Stokes equations is used because a gas can be considered being a continuous medium. However, it is worthwhile emphasizing that this is true only outside of the Knudsen layer, which has the order of a mean free path. Inside of the Knudsen layer the classical theory of sound propagation cannot be used. Second, the high oscillation speed regime in which there are very few intermolecular collisions during one cycle of the oscillation; such a situation can happen when the sound frequency is high or when the gas pressure is low. In this regime the intermolecular collisions can be neglected and the problem can be solved analytically on the level of the distribution function. Third, the transitional regime in which both collision and oscillation frequencies have the same order; in such a regime neither the Navier–Stokes equations are applied nor the intermolecular collisions are neglected, but the problem must be solved on the basis of the nonstationary kinetic equation.

The sound propagation through a rarefied gas was investigated both theoretically, see e.g., Refs. 3–18, and experimentally, see, e.g., Refs. 19–23. The authors of Refs. 4, 7, 11, and 15 assumed the wave vector to be constant over the whole regime of the oscillation speed parameter, i.e., the wave is harmonic in the space. However, analytical calculations in the high oscillation speed regime performed in Refs. 3, 9, and 17 showed that the wave vector is not constant, but it depends on the distance from a sound source. In other words, the sound wave propagating through a rarefied gas is not harmonic in the space. A polynomial expansion was used in Refs. 5, 6, 8, and 10 in order to obtain the wave characteristics for arbitrary values of the oscillation speed parameter. More exact solution of the kinetic equation was obtained in Refs. 12, 13, and 18. The direct simulation Monte Carlo method was applied to the sound propagation problem in Ref. 16 for a few values of the oscillation speed parameter.

In the present work the sound propagation through a rarefied gas is considered on the basis of numerical solution of the kinetic equation over the whole range of the oscillation speed parameter. Behaviors of gas density, bulk velocity, temperature, and pressure stress are analyzed using the concepts of the phase speed and attenuation. Since for the sound propagation both viscosity and thermal conductivity are im-

[a]Electronic mail: sharipov@fisica.ufpr.br. URL: http://fisica.ufpr.br/sharipov

portant the kinetic model[24] with the correct value of the Prandtl number is applied and solved by the discrete velocity method.[25] Unlike all previous works, here the scattering kernel by Cercignani and Lampis[26] is used as the boundary condition to the kinetic equation in order to investigate the influence of the gas-surface interaction law on the sound propagation.

II. STATEMENT OF THE PROBLEM

Consider a monoatomic gas occupying a semi-infinite space $x'>0$. An infinite flat plate located in $x'=0$ parallel to the plane yz oscillates harmonically in the x' direction with a frequency ω so that its velocity depends on time t' as

$$U_w(t') = \Re[U_m \exp(-i\omega t')], \tag{1}$$

where $\Re$ denotes the real part of complex quantity and U_m is the velocity amplitude of the plate, which is assumed to be very small when compared with the most probable molecular speed v_m of the gas flow, i.e.,

$$U_m \ll v_m, \quad v_m = \left(\frac{2k_B T_0}{m}\right)^{1/2}, \tag{2}$$

where m is the molecular mass of the gas, T_0 is its equilibrium temperature, and k_B is the Boltzmann constant.

The sound wave generated by the oscillating plate propagates through the gas and causes a gas flow characterized by its bulk velocity U_x, density Δn, and temperature ΔT deviations from their equilibrium values n_0 and T_0, respectively. In practice, a pressure in the wave propagation direction P_{xx} is measured. All these quantities depend on time harmonically and can be written as

$$U_x(t',x') = \Re[\tilde{U}_x(x')\exp(-i\omega t')], \tag{3}$$

$$\Delta n(t',x') = \Re[\Delta\tilde{n}(x')\exp(-i\omega t')], \tag{4}$$

$$\Delta T(t',x') = \Re[\Delta\tilde{T}(x')\exp(-i\omega t')], \tag{5}$$

$$P_{xx}(t',x') = \Re[\tilde{P}_{xx}(x')\exp(-i\omega t')], \tag{6}$$

where $\tilde{U}_x(x')$, $\Delta\tilde{n}(x')$, $\Delta\tilde{T}(x')$, and $\tilde{P}_{xx}(x')$ are complex quantities.

The solution is determined by the oscillation speed parameter defined as

$$\theta = \frac{\nu}{\omega}, \quad \nu = \frac{P_0}{\mu}, \tag{7}$$

where ν is the intermolecular collision frequency, P_0 is the equilibrium pressure, and μ is the shear viscosity of the gas.

For further derivations it is more convenient to introduce the following dimensionless quantities:

$$t = \omega t', \quad x = \frac{\omega}{v_m}x', \tag{8}$$

$$u(x) = \frac{\tilde{U}_x(x)}{U_m}, \quad \varrho(x) = \frac{\Delta\tilde{n}(x)}{n_0}\frac{v_m}{U_m}, \tag{9}$$

$$\tau(x) = \frac{\Delta\tilde{T}(x)}{T_0}\frac{v_m}{U_m}, \quad \Pi(x) = \frac{\tilde{P}_{xx}}{2P_0}\frac{v_m}{U_m}. \tag{10}$$

Since the dimensional quantities given in Eqs. (3)–(6) are complex, the corresponding dimensionless quantities can be written as

$$u(x) = A_u(x)\exp[i\varphi_u(x)], \tag{11}$$

$$\varrho(x) = A_n(x)\exp[i\varphi_n(x)], \tag{12}$$

$$\tau(x) = A_T(x)\exp[i\varphi_T(x)], \tag{13}$$

$$\Pi(x) = A_P(x)\exp[i\varphi_P(x)], \tag{14}$$

where $A_i(x)$ $(i=u,n,T,P)$ are the amplitudes, while $\varphi_i(x)$ $(i=u,n,T,P)$ are the corresponding phases. The attenuation, usually used in acoustics, is introduced here in the dimensionless form for each quantity given in Eqs. (11)–(14), i.e.,

$$a_i = -\frac{\mathrm{d}\ln A_i}{\mathrm{d}x}, \quad i = u,n,T,P. \tag{15}$$

With the help of Eq. (8) the dimension attenuation a_i' is related to the corresponding dimensionless one as $a_i' = a_i\omega/v_m$.

Another quantity widely used in acoustics is the phase speed, which in the dimensionless form reads

$$\vartheta_i = \left[\frac{\mathrm{d}\varphi_i}{\mathrm{d}x}\right]^{-1}, \quad i = u,n,T,P. \tag{16}$$

Then the dimension phase velocity ϑ_i' is given as $\vartheta_i' = \vartheta_i v_m$.

The aim of the present work is to study the sound wave propagation through a rarefied gas on the basis of the kinetic equation. We are going to calculate the attenuation and phase speeds as a function of the oscillation speed parameter θ. To study the role of the gas-surface interaction both noncomplete accommodation and diffuse scattering will be considered.

III. LOW OSCILLATION SPEED REGIME

In the low oscillation speed regime $(\theta \gg 1)$, i.e., when the sound frequency is significantly lower than the intermolecular frequency, the classical theory of sound propagation in gases is valid. According to this theory, any quantity $\psi(x',t')$ given by Eqs. (3)–(6) can be represented as

$$\psi(x',t') = A\exp[i(kx' - \omega t')], \tag{17}$$

where k is the wave number given as[1,2]

$$k = \frac{\omega}{c_0}\left(1 + i\frac{0.7}{\theta}\right), \quad c_0 = \sqrt{\frac{5k_B T_0}{3m}}, \tag{18}$$

where c_0 is the adiabatic sound velocity for a monoatomic gas. Thus, the attenuation a' and phase speed ϑ' are the same for all quantities and given as

$$a' = 0.7\frac{\mu\omega^2}{Pc_0}, \quad \vartheta' = c_0. \tag{19}$$

The corresponding dimensionless attenuation and phase speed take the form

$$a_i = \frac{v_m}{\omega}a' = \frac{0.7}{\theta}\sqrt{\frac{6}{5}}, \quad i = u,n,T,P, \tag{20}$$

$$\vartheta_i = \frac{\vartheta'}{v_m} = \sqrt{\frac{5}{6}}, \quad i = u,n,T,P. \tag{21}$$

IV. KINETIC EQUATION

In order to consider an arbitrary value of the oscillation parameter θ the problem must be solved on the basis of the one-dimensional Boltzmann equation that reads

$$\frac{\partial f}{\partial t'} + v_x\frac{\partial f}{\partial x'} = Q(ff_*), \tag{22}$$

where $f(t',x',\boldsymbol{v})$ is the distribution function that depends on the time t', x' coordinate, and molecular velocity $\mathbf{v}$, and $Q(ff_*)$ denotes the collision integral.[27,28] For our purpose the Shakhov model[29] of the collision integral is most suitable because it describes correctly both mass and heat transfer. In the nonlinear form the model reads

$$Q(ff_*) = \frac{P}{\mu}\left\{f^M\left[1 + \frac{2m(\boldsymbol{Q}\cdot\boldsymbol{V})}{15n(k_BT)^2}\left(\frac{mV^2}{2k_BT} - \frac{5}{2}\right)\right] - f\right\}, \tag{23}$$

where $\boldsymbol{V}=\boldsymbol{v}-\boldsymbol{U}$ is the peculiar velocity and f^M is the local Maxwellian given by

$$f^M = n\left(\frac{m}{2\pi k_BT}\right)^{3/2}\exp\left[-\frac{m(\boldsymbol{v}-\boldsymbol{U})^2}{2k_BT}\right], \tag{24}$$

where n, $\boldsymbol{U}$, T, and $\boldsymbol{Q}$ are the local number density, bulk velocity, temperature, and heat flux vector, respectively. Note that for the one-dimensional problem under question both $\boldsymbol{U}$ and $\boldsymbol{Q}$ have only the x component. All these quantities are calculated via moments of the distribution function as

$$n(t',x') = \int f(t',x',\boldsymbol{v})\mathrm{d}\boldsymbol{v}, \tag{25}$$

$$U_x(t',x') = \frac{1}{n}\int v_x f(t',x',\boldsymbol{v})\mathrm{d}\boldsymbol{v}, \tag{26}$$

$$T(t',x') = \frac{m}{3nk_B}\int (\boldsymbol{v}-\boldsymbol{U})^2 f(t',x',\boldsymbol{v})\mathrm{d}\boldsymbol{v}, \tag{27}$$

$$Q_x(t,x') = \frac{m}{2}\int \boldsymbol{V}^2(v_x - U_x)f(t',x',\boldsymbol{v})\mathrm{d}\boldsymbol{v}. \tag{28}$$

The shear stress is also calculated via the distribution function as

$$P_{xx}(t',x') = m\int (v_x - U_x)^2 f(t',x',\boldsymbol{v})\mathrm{d}\boldsymbol{v}. \tag{29}$$

Since it is assumed the velocity amplitude of the plate U_m to be small when compared with the most probable molecular velocity v_m of the gas, the kinetic equation (22) can be linearized by representing the distribution function as

$$f(t',x',\boldsymbol{v}) = \frac{n_0}{(\sqrt{\pi}v_m)^3}\exp(-c^2)\left\{1 + \Re[h(x,\boldsymbol{c})e^{-it}]\frac{U_m}{v_m}\right\}, \tag{30}$$

where $h(x,\boldsymbol{c})$ is the complex perturbation function. Substituting Eq. (30) into Eq. (22) with Eq. (23) the linearized kinetic equation is obtained as follows:

$$(\theta - i)h + c_x\frac{\partial h}{\partial x} = \theta\left[\varrho + 2c_xu + \tau\left(c^2 - \frac{3}{2}\right) + \frac{4}{15}qc_x\left(c^2 - \frac{5}{2}\right)\right], \tag{31}$$

where

$$\begin{bmatrix}\varrho(x)\\ u(x)\\ \tau(x)\\ q(x)\\ \Pi(x)\end{bmatrix} = \frac{1}{\pi^{3/2}}\int\begin{bmatrix}1\\ c_x\\ \frac{2}{3}c^2 - 1\\ c_x\left(c^2 - \frac{5}{2}\right)\\ c_x^2\end{bmatrix}h(x,\boldsymbol{c})\exp(-c^2)\mathrm{d}\boldsymbol{c}. \tag{32}$$

The last moment is necessary to calculate the pressure tensor by Eq. (10).

The boundary condition on the oscillating plate is given in general form as

$$h^+ = \hat{A}h^- + h_w - \hat{A}h_w, \quad h_w = 2c_x, \tag{33}$$

where $\hat{A}$ denotes the scattering operator, and h^+ and h^- are the perturbations of the reflected and incident gaseous particles on the wall, respectively.

In order to consider a nondiffuse gas-surface interaction law, the scattering kernel proposed by Cercignani and Lampis[26] is employed. This kernel provides a more physical description of the gas-surface interaction law since two accommodation coefficients are introduced, namely, the tangential momentum accommodation coefficient α_t and the energy accommodation coefficient α_n corresponding to the normal molecular velocity. Such a scattering kernel was successfully applied to many rarefied gas dynamics problems[30–33] where experimental values of the accommodation coefficients are reported.

In our notations, the Cercignani–Lampis scattering operator reads

$$\hat{h}^+ = \hat{A}_x\hat{A}_y\hat{A}_zh^-, \tag{34}$$

$$\hat{A}_xh^- = \frac{1}{c_x}\int_{c_x'<0}|c_x'|\exp(c_x^2 - c_x'^2)R_x(c_x' \to c_x) \times h(c_x')\mathrm{d}c_x', \quad c_x \geq 0, \tag{35}$$

$$\hat{A}_i h^- = \int_{-\infty}^{\infty} \exp(c_i^2 - c_i'^2) R_i(c_i' \to c_i) h(c_i') \mathrm{d}c_i', \quad i = y,z, \tag{36}$$

where

$$R_x(c_x' \to c_x) = \frac{c_x}{\pi\alpha_n} \exp\left[- \frac{c_x^2 + (1-\alpha_n)c_x'^2}{\alpha_n} \right] \times \int_0^{2\pi} \exp\left(\frac{2\sqrt{1-\alpha_n} c_x c_x'}{\alpha_n} \cos\phi \right) \mathrm{d}\phi, \tag{37}$$

$$R_i(c_i' \to c_i) = \frac{1}{[\pi\alpha_t(2-\alpha_t)]^2} \times \exp\left[- \frac{[c_i - (1-\alpha_t)c_i']^2}{\alpha_t(2-\alpha_t)} \right], \quad i = y,z. \tag{38}$$

After some algebraic manipulations, the boundary condition given in Eq. (34) is transformed into

$$h^+ = \hat{A} h^- + 2c_x - 2\hat{A}_x c_x. \tag{39}$$

Below, the integrodifferential equations (31) and (32) subject to the boundary conditions (34)–(39) will be solved analytically in the limit of high oscillation speed (θ=0) and numerically for the transitional values of the parameter θ.

V. HIGH OSCILLATION SPEED REGIME

In the high oscillation speed regime the parameter θ tends to zero and Eq. (31) is reduced to

$$-ih + c_x \frac{\partial h}{\partial x} = 0. \tag{40}$$

In the case of the diffuse scattering of gaseous particles on the sound source the problem can be solved analytically. In this case the boundary condition reads

$$h_w^+ = 2c_x + \sqrt{\pi}, \quad c_x > 0. \tag{41}$$

Integrating Eq. (40) with respect to the x coordinate the following solution is obtained:

$$h(x,c_x) = \begin{cases} (2c_x + \sqrt{\pi}) \exp\left(i \dfrac{x}{c_x} \right) & \text{at } c_x > 0 \\ 0 & \text{at } c_x < 0. \end{cases} \tag{42}$$

Then, the moments given in Eq. (32) are obtained as

$$\varrho(x) = \frac{2}{\sqrt{\pi}} I_1(-ix) + I_0(-ix), \tag{43}$$

$$u(x) = \frac{2}{\sqrt{\pi}} I_2(-ix) + I_1(-ix), \tag{44}$$

$$\tau(x) = \frac{4}{3\sqrt{\pi}} I_3(-ix) + \frac{2}{3} I_2(-ix) - \frac{2}{3\sqrt{\pi}} I_1(-ix) - \frac{1}{3} I_0(-ix), \tag{45}$$

$$q(x) = \frac{2}{\sqrt{\pi}} I_4(-ix) + I_3(-ix) - \frac{3}{\sqrt{\pi}} I_2(-ix) - \frac{3}{2} I_1(-ix), \tag{46}$$

$$\Pi(x) = \frac{2}{\sqrt{\pi}} I_3(-ix) + I_2(-ix), \tag{47}$$

where the functions $I_n(z)$ are defined as

$$I_n(z) = \int_0^{\infty} c^n \exp\left(c^2 - \frac{z}{c} \right) \mathrm{d}c. \tag{48}$$

Using the power representation given in Ref. 34 the following limit values of the attenuations and phases of the quantities given in Eqs. (43)–(47) at the surface ($x \to 0$) are obtained:

$$a_n = \frac{\pi^{3/2}}{2(1+\sqrt{\pi})}, \quad a_u = -\left(\frac{1}{\sqrt{\pi}} + \frac{\sqrt{\pi}}{2} \right)^2, \quad a_T = -3\sqrt{\pi}, \quad a_P = 0, \tag{49}$$

$$\varphi_n = \varphi_T = \varphi_P = \varphi_u = 0. \tag{50}$$

In the other limit, when $x \to \infty$, all quantities given in Eqs. (43)–(47) have the same asymptotic behavior for their attenuations, i.e.,

$$a_i = \frac{1}{(4x)^{1/3}}, \quad i = n,u,T,P, \tag{51}$$

while their phase speeds are given by

$$\vartheta_i = \frac{(4x)^{1/3}}{\sqrt{3}}, \quad i = n,u,T,P. \tag{52}$$

VI. TRANSITIONAL REGIME

In the transitional regime ($\theta \sim 1$) the kinetic equation (31) is solved numerically. In order to eliminate the velocity variables c_y and c_z the following new functions are introduced:

$$\Phi(x,c_x) = \frac{1}{\pi} \int h(x,\boldsymbol{c}) \exp(-c_y^2 - c_z^2) \mathrm{d}c_y \mathrm{d}c_z, \tag{53}$$

$$\Psi(x,c_x) = \frac{1}{\pi} \int (c_y^2 + c_z^2 - 1) h(x,\boldsymbol{c}) \exp(-c_y^2 - c_z^2) \mathrm{d}c_y \mathrm{d}c_z. \tag{54}$$

Multiplying Eq. (31) by $\exp(-c_y^2 - c_z^2)/\pi$ and by $(c_y^2 + c_z^2 - 1)\exp(-c_y^2 - c_z^2)/\pi$ then integrating it with respect to c_y and c_z the following equations are obtained:

$$(\theta - i)\Phi + c_x \frac{\partial \Phi}{\partial x} = \theta \left[\varrho + 2c_x u + \tau\left(c_x^2 - \frac{1}{2} \right) + \frac{4}{15} q c_x \left(c_x^2 - \frac{3}{2} \right) \right], \tag{55}$$

$$(\theta - i)\Psi + c_x \frac{\partial \Psi}{\partial x} = \theta \left[\tau + \frac{4}{15} q c_x \right]. \tag{56}$$

The moments defined by Eq. (32) are calculated via the new perturbation functions as

$$\begin{bmatrix} \varrho(x) \\ u(x) \\ \tau(x) \\ q(x) \\ \Pi(x) \end{bmatrix} = \frac{1}{\sqrt{\pi}} \int \left\{ \begin{bmatrix} 1 \\ c_x \\ \frac{2}{3}c_x^2 - \frac{1}{3} \\ c_x\left(c_x^2 - \frac{3}{2}\right) \\ c_x^2 \end{bmatrix} \Phi(x, c_x) + \begin{bmatrix} 0 \\ 0 \\ \frac{2}{3} \\ c_x \\ 0 \end{bmatrix} \Psi(x, c_x) \right\} \exp(-c_x^2) \mathrm{d}c_x. \tag{57}$$

The boundary conditions to Eqs. (55) and (56) are obtained from Eq. (39) and written as

$$\Phi^+ = \hat{A}_x \Phi^- + 2c_x - 2\hat{A}_x c_x, \tag{58}$$

$$\Psi^+ = (1 - \alpha_t)^2 \hat{A}_x \Psi^-. \tag{59}$$

Equations (55) and (56) with the boundary conditions (58) and (59) were solved by the discrete velocity method in which a nonregular distribution of points c_{xi} $(1 \le i \le N_c)$ in the velocity space was chosen so that a higher density of points for small values of c_{xi} and lower density for its large values were used. A regular distribution of points in the x coordinate was introduced as

$$x_k = x_{k-1} + \Delta x, \quad 1 \le k \le N_x, \quad x_0 = 0, \quad \Delta x = \frac{x_{\max}}{N_x}, \tag{60}$$

where N_x is an integer and $x_{\max}$ is the last point corresponding to the distance from the plate in which the perturbation caused by the oscillating plate can be neglected.

Equations (55) and (56) are approximated by the following central finite difference scheme:

$$(\theta - i)\frac{\Phi_{k,i} + \Phi_{k-1,i}}{2} + c_{xi}\frac{\Phi_{k,i} - \Phi_{k-1,i}}{\Delta x} = \frac{\theta}{2}\left[\varrho_k + \varrho_{k-1} + 2c_{xi}(u_k + u_{k-1}) + (\tau_k + \tau_{k-1})\left(c_{xi}^2 - \frac{1}{2}\right) + \frac{4}{15}(q_k + q_{k-1})c_{xi}\left(c_{xi}^2 - \frac{3}{2}\right) \right], \tag{61}$$

$$(\theta - i)\frac{\Psi_{k,i} - \Psi_{k-1,i}}{2} + c_{xi}\frac{\Psi_{k,i} - \Psi_{k-1,i}}{\Delta x} = \frac{\theta}{2}\left[\tau_k + \tau_{k-1} + \frac{4}{15}(q_k + q_{k-1})c_{xi} \right], \tag{62}$$

where

$$\Phi_{k,i} = \Phi(x_k, c_{xi}), \quad \Psi_{k,i} = \Psi(x_k, c_{xi}). \tag{63}$$

The moments ϱ_k, u_k, τ_k, q_k, and Π_k are calculated by the following quadrature formulas:

$$\varrho_k = \sum_{i=1}^{N_c} \Phi_{k,i} W_i, \tag{64}$$

$$u_k = \sum_{i=1}^{N_c} \Phi_{k,i} c_{xi} W_i, \tag{65}$$

$$\tau_k = \frac{2}{3}\sum_{i=1}^{N_c} \left[\Phi_{k,i}\left(c_{xi}^2 - \frac{1}{2}\right) + \Psi_{k,i} \right] W_i, \tag{66}$$

$$q_k = \sum_{i=1}^{N_c} \left[\Phi_{k,i}\left(c_{xi}^2 - \frac{3}{2}\right) + \Psi_{k,i} \right] c_{xi} W_i. \tag{67}$$

$$\Pi_k = \sum_{i=1}^{N_c} \Phi_{k,i} c_{xi}^2 W_i, \tag{68}$$

where W_i is the weight of the corresponding point c_{xi}.

Like our previous work,[35] in order to reduce the number of points in the velocity space the solution of the kinetic equation was split in the case of the diffuse scattering, i.e., at $\alpha_n = 1$ and $\alpha_t = 1$. Therefore, the perturbation functions Φ and Ψ are presented as

$$\Phi = \Phi_0 + \tilde{\Phi}, \quad \Psi = \Psi_0 + \tilde{\Psi}, \tag{69}$$

where the functions Φ_0 and Ψ_0 satisfy the equations

$$(\theta - i)\Phi_0 + c_x \frac{\partial \Phi_0}{\partial x} = 0, \tag{70}$$

$$(\theta - i)\Psi_0 + c_x \frac{\partial \Psi_0}{\partial x} = 0, \tag{71}$$

with the following boundary conditions:

$$\Phi_0 = 2c_x + \nu_0 \quad \text{at } x = 0 \text{ and } c_x > 0, \tag{72}$$

$$\Psi_0 = 0 \quad \text{at } x = 0 \text{ and } c_x > 0. \tag{73}$$

The constant $\nu_0 = \sqrt{\pi}$ is calculated from the boundary condition (58). The integration of Eqs. (70) and (71) with the boundary conditions (72) and (73) yields the expressions

$$\Phi_0(x, c_x) = \begin{cases} (2c_x + \nu_0)\exp\left[-(\theta - i)\dfrac{x}{c_x} \right] & \text{at } c_x > 0 \\ 0 & \text{at } c_x < 0. \end{cases} \tag{74}$$

TABLE I. Parameters of the numerical scheme x_{max}, N_x, and N_c [see Eqs. (60)–(68)] providing the accuracy of 0.1% and penetration depth λ vs oscillation speed parameter θ.

θ	x_{max}	N_x	N_c	λ
0.1	40	40 000	4000	16.44
1	40	40 000	4000	18.16
10	80	80 000	4000	53.63
20	120	120 000	4000	113.40

Substituting Eq. (69) into Eqs. (55) and (56) the following equations are obtained for the functions $\tilde{\Phi}$ and $\tilde{\Psi}$:

$$(\theta - i)\tilde{\Phi} + c_x \frac{\partial \tilde{\Phi}}{\partial x} = \theta \left[\varrho + 2c_x u + \tau \left(c_x^2 - \frac{1}{2} \right) + \frac{4}{15} q c_x \left(c_x^2 - \frac{3}{2} \right) \right], \tag{75}$$

$$(\theta - i)\tilde{\Psi} + c_x \frac{\partial \tilde{\Psi}}{\tilde{x}} = \theta \left[\tau + \frac{4}{15} q c_x \right]. \tag{76}$$

The boundary conditions to Eqs. (75) and (76) are given by

$$\tilde{\Phi}(x, c_x) = -2 \int_{c_x < 0} c_x \tilde{\Phi}(0, c_x) \exp(-c_x^2) \mathrm{d}c_x \quad \text{at } x = 0 \text{ and } c_x > 0, \tag{77}$$

$$\tilde{\Psi}(x, c_x) = 0 \quad \text{at } x = 0 \text{ and } c_x > 0. \tag{78}$$

The moments given by Eq. (57) are also decomposed into two parts as

$$\begin{bmatrix} \varrho(x) \\ u(x) \\ \tau(x) \\ q(x) \\ \Pi(x) \end{bmatrix} = \begin{bmatrix} \tilde{\varrho}(x) \\ \tilde{u}(x) \\ \tilde{\tau}(x) \\ \tilde{q}(x) \\ \tilde{\Pi}(x) \end{bmatrix} + \begin{bmatrix} \varrho_0(x) \\ u_0(x) \\ \tau_0(x) \\ q_0(x) \\ \Pi_0(x) \end{bmatrix}, \tag{79}$$

where the terms $\tilde{\varrho}$, $\tilde{u}$, $\tilde{\tau}$, $\tilde{q}$, and $\tilde{\Pi}$ are calculated by Eq. (57) using $\tilde{\Phi}$ and $\tilde{\Psi}$ instead of Φ and Ψ, respectively. The second terms in the left hand side of Eq. (79) are given by the corresponding expressions (43)–(47) replacing $(-ix)$ with $[(\theta - i)x]$.

Equations (75) and (76) are discretized in the same way as in Eqs. (61) and (62). The splitting of the functions Φ and Ψ allows us to decrease their oscillation in the velocity space c_x and, consequently, to reduce significantly the number of points N_c.

The calculations were carried out with the computational error less than 0.1% estimated by varying the parameters x_{max}, N_x, and N_c. The values of these parameters, which provide such an accuracy, are given in Table I. The penetration depth λ defined as the distance x where the amplitude of the bulk velocity decays up to 1% of the velocity amplitude of the plate U_m is also presented in Table I. As can be seen, the penetration depth λ increases by increasing the parameter θ.

VII. RESULTS AND DISCUSSION

First of all, a comparison of our results corresponding to the diffuse scattering with those reported in the third columns of Tables 1–8 in Ref. 18 was performed. The comparison showed that they coincide with each other within the computational error.

Then, the influence of the gas-surface interaction on the sound propagation was analyzed. The dependence of the amplitude A_P and phase of φ_P on the accommodation coefficients α_n and α_t is presented in Tables II–IV for the three vales of the oscillation parameter θ=0.1, 1, and 10, respec-

TABLE II. Amplitude A_P and phase φ_P based on the pressure tensor at x=0 vs accommodation coefficients α_n and α_t, θ=0.1.

	A_P				$-\varphi_P$			
α_n	α_t=0.25	0.5	1	2	0.25	0.5	1	2
0.25	1.0943	1.0943	1.0943	1.0943	0.014 26	0.014 26	0.014 26	0.014 26
0.5	1.0637	1.0637	1.0637	1.0637	0.013 06	0.013 06	0.013 06	0.013 06
1	1.0066	1.0066	1.0066	1.0066	0.010 85	0.010 85	0.010 85	0.010 85

TABLE III. Amplitude A_P and phase φ_P based on the pressure tensor at x=0 vs accommodation coefficients α_n and α_t, θ=1.

	A_P				$-\varphi_P$			
α_n	α_t=0.25	0.5	1	2	0.25	0.5	1	2
0.25	1.0266	1.0277	1.0286	1.0249	0.098 75	0.100 1	0.101 1	0.096 99
0.5	1.0034	1.0044	1.0052	1.0019	0.090 67	0.091 85	0.092 81	0.089 08
1	0.9587	0.9595	0.9600	0.9575	0.074 82	0.075 80	0.076 61	0.073 49

TABLE IV. Amplitude A_P and phase φ_P based on the pressure tensor at $x=0$ vs accommodation coefficients α_n and α_t, $\theta=10$.

	A_P				$-\varphi_P$			
α_n	$\alpha_t=0.25$	0.5	1	2	0.25	0.5	1	2
0.25	0.8826	0.8736	0.8672	0.8971	0.030 74	0.034 16	0.035 66	0.022 22
0.5	0.8727	0.8648	0.8593	0.8856	0.026 49	0.028 69	0.029 59	0.020 29
1	0.8556	0.8496	0.8455	0.8658	0.015 58	0.016 33	0.016 47	0.012 56

tively. Only the point at the oscillating plate, i.e., $x=0$, where the influence of the gas surface is expected to be most significant, is considered. It can be seen that there is no influence of the tangential momentum accommodation coefficient α_t when $\theta=0.1$, i.e., in the high oscillation speed regime. In the transitional regime, i.e., $\theta=1$ and 10, the amplitude Π_m and phase φ_P weakly depend on the accommodation coefficient α_t. The energy accommodation coefficient α_n affects both amplitude Π_m and phase φ_P, but its influence is within few percents. From these data it can be concluded that the noncomplete accommodation of the gaseous particles on the surface affects weakly the sound characteristics. Further, the results only for the diffuse scattering will be analyzed.

The attenuations a_n, a_u, a_T, and a_P based on the quantities $\varrho(x)$, $u(x)$, $\tau(x)$, and $\Pi(x)$ are presented in Figs. 1–4, respectively, for $\theta=0.1$, 1, 10, and 20. From these plots one can see the following: (i) Only for large values of θ, i.e., 10 and 20, all attenuations tend to the value given by Eq. (20), i.e., $0.7\sqrt{6/5}/\theta$. (ii) For intermediate and small values of θ, i.e., 1 and 0.1, the attenuations do not tend to a constant value. Note, that in the free-molecular regime ($\theta=0$) all attenuations have the asymptotic behavior given by Eq. (51). (iii) Near the sound source, i.e., in the Knudsen layer, the attenuation behaviors of the different quantities are quite different over the whole range of the parameter θ. The temperature attenuation a_T is even negative. Therefore, in the low oscillation speed regime ($\theta \gg 1$), Eq. (20) is valid only outside of the Knudsen layer.

The phase speeds ϑ_n, ϑ_u, ϑ_T, and ϑ_P based on the quantities $\varrho(x)$, $u(x)$, $\tau(x)$, and $\Pi(x)$ are depicted in Figs. 5–8, respectively, for $\theta=0.1$, 1, 10, and 20. From these plots practically the same conclusions as those for the attenuations can be made: (i) Only for large values of θ, i.e., 10 and 20, all phase speed tend to the value given by Eq. (21), i.e., $\sqrt{5/6}$. (ii) For intermediate and small values of θ, i.e., 1 and 0.1, the phase speeds do not tend to a constant value. In the free-molecular regime ($\theta=0$) the phase speeds have the same asymptotic behavior given by Eq. (52). (iii) Near the sound

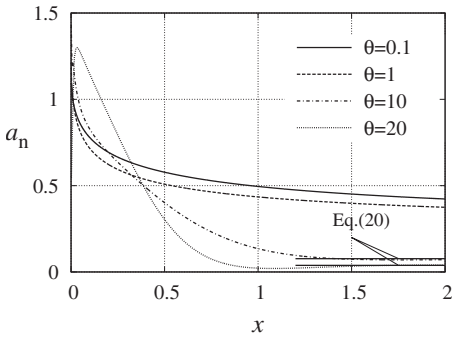


FIG. 1. Attenuation based on the numerical results for density $\varrho(x)$ vs distance x and its asymptotic values at $\theta=10$ and $\theta=20$ calculated by Eq. (20).

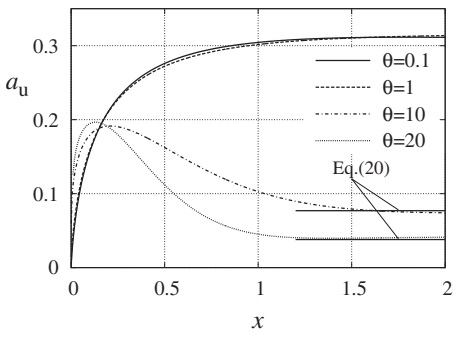


FIG. 2. Attenuation based on the numerical results for bulk velocity $u(x)$ vs distance x and its asymptotic values at $\theta=10$ and $\theta=20$ calculated by Eq. (20).

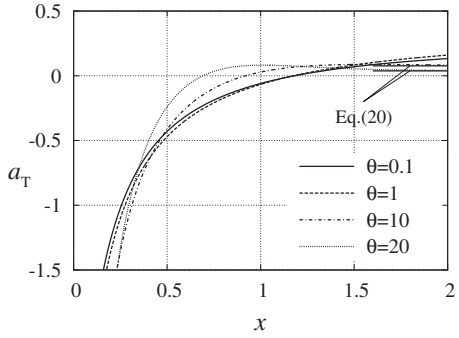


FIG. 3. Attenuation based on the numerical results for temperature $\tau(x)$ vs distance x and its asymptotic values at $\theta=10$ and $\theta=20$ calculated by Eq. (20).

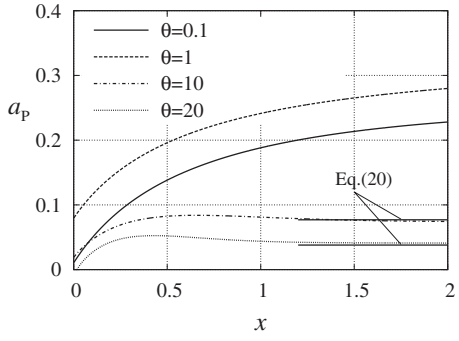


FIG. 4. Attenuation based on the numerical results for pressure tensor $\Pi(x)$ vs distance x and its asymptotic values at $\theta=10$ and $\theta=20$ calculated by Eq. (20).

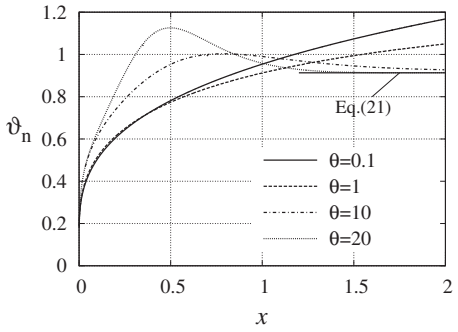


FIG. 5. Phase speed based on the numerical results for density $\varrho(x)$ vs distance x and its asymptotic value given by Eq. (21).

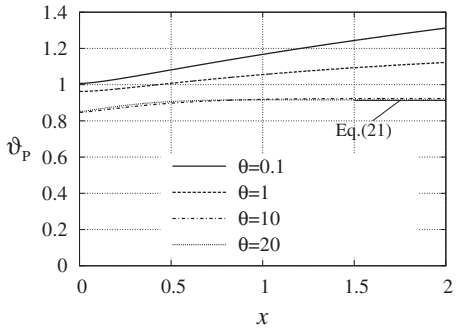


FIG. 8. Phase speed based on the numerical results for pressure tensor $\Pi(x)$ vs distance x and its asymptotic value given by Eq. (21).

source, i.e., in the Knudsen layer, the phase speeds of the different quantities are quite different over the whole range of the parameter θ.

Note that among all quantities showed in Figs. 1–8 those based on the temperature, i.e., a_T and ϑ_P, represent a very sharp variation near the surface. It is explained by the fact that the temperature of the plate is fixed at T_0, while the gas temperature oscillates near the value T_0. In other words, there is a temperature jump near the wall. As was shown by Ewing *et al.*,[23] a consideration of such a jump is important in order to extend an application of the classical theory of sound propagation to lower pressures.

Experimental data on the sound propagation through noble gases for the pressure range from 100 Pa up to atmospheric one are reported in Refs. 19 and 20 where a constant sound frequency was used, but the distance between the transmitter and the receiver was varied. Analogous results were reported by Schotter[22] for a large range of sound frequencies and for two values of the distance x, viz,. $10/\sqrt{\pi}$ and $20/\sqrt{\pi}$. A comparison of our results with the experimental data[22] at $x=20/\sqrt{\pi}$ is shown in Fig. 9. Note that only the quantities based on the pressure tensor were measured in the above mentioned experimental works. It can be seen that the results obtained in the present work are in good agreement with the experimental data[22] over the whole range of oscillation parameter θ. The theoretical results by Marques[15] obtained by the moment method assuming the sound wave to be harmonic in the space are also plotted in Fig. 1. In the range $\theta>0.2$ both the present work and Ref. 15 provide the same results. However, there is a discrepancy of the attenuation for smaller values of the oscillating parameter θ. We believe that even if the moment method is improved by considering a larger number of moments it will be not able to describe the sound propagation for small values of the parameter θ because the wave is not harmonic anymore.

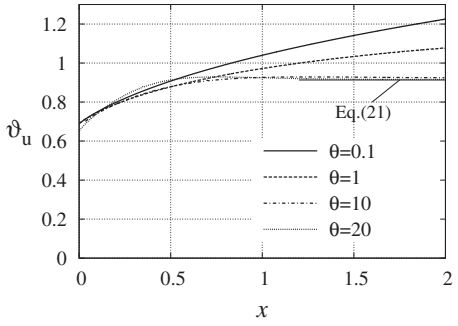


FIG. 6. Phase speed based on the numerical results for bulk velocity $u(x)$ vs distance x and its asymptotic value given by Eq. (21).

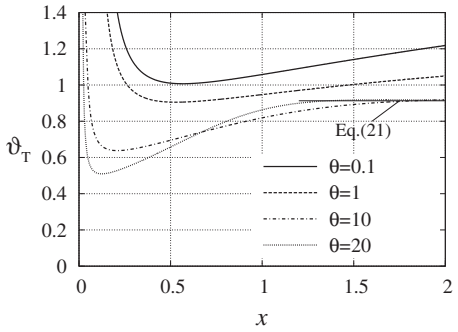


FIG. 7. Phase speed based on the numerical results for temperature $\tau(x)$ vs distance x and its asymptotic value given by Eq. (21).

VIII. CONCLUDING REMARKS

A sound wave propagating through a rarefied gas was calculated over the whole range of the oscillation speed parameter from the free-molecular regime when the oscillation frequency is so high that the intermolecular collisions are neglected up to the hydrodynamic regime when the wave oscillation is so low that the Navier–Stokes equations are applied. The attenuations and phase speeds were calculated on the basis of the gaseous density, bulk velocity, temperature, and pressure tensor. It was found that near the oscillat-

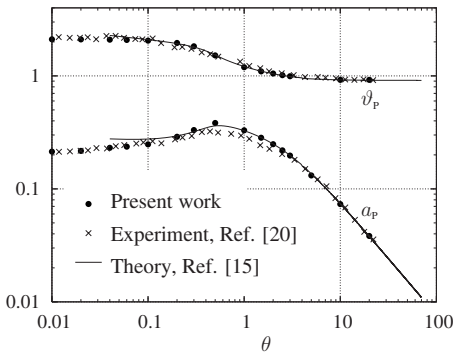


FIG. 9. Comparison of the numerical results for attenuation and phase velocity based on the pressure tensor $\Pi(x)$ at the distance $x=20/\sqrt{\pi}$ with the experimental data (Ref. 22) where the distance was the same and with the theoretical results (Ref. 15) where the distance from the source is assumed to be infinite.

ing plate the attenuations and phase speeds are different for each quantity. Far from the oscillating plate the attenuations and phase speeds tend to constant values only in the low oscillation speed regime. The gas-surface interaction law weakly affects the sound characteristics. The present results are in good agreement with experimental data reported in the open literature.

ACKNOWLEDGMENTS

The authors acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) for the support of their research. The authors also thank Dr. W. Marques Jr. for the discussion on the present work and for his results (Ref. 15) provided in a table.

[1]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1989).
[2]J. Lighthill, *Waves in Fluids* (Cambridge University Press, New York, 1978).
[3]G. Maidanik, H. L. Fox, and M. Heckl, "Propagation and reflection of sound in rarefied gases. I. Theoretical," Phys. Fluids **8**, 259–265 (1965).
[4]L. Sirovich and J. K. Thurber, "Propagation of forced sound waves in rarefied gasdynamics," J. Acoust. Soc. Am. **37**, 329–339 (1965).
[5]D. Kahn and D. Mintzer, "Kinetic theory of sound propagation in rarefied gases," Phys. Fluids **8**, 1090–1102 (1965).
[6]D. Kahn, "Sound propagation in rarefied gases," Phys. Fluids **9**, 1867–1869 (1966).
[7]J. K. Buckner and J. H. Ferziger, "Linearized boundary value problem for a gas and sound propagation," Phys. Fluids **9**, 2315–2322 (1966).
[8]K. Toba, "Kinetic theory of sound propagation in a rarefied gas," Phys. Fluids **11**, 2495–2497 (1968).
[9]K. Toba, "Effect of gas-surface interaction on sound propagation," Phys. Fluids **11**, 507–514 (1968).
[10]F. B. Hanson and T. F. Morse, "Free-molecular expansion polynominals and sound propagation in rarefied gases," Phys. Fluids **12**, 1564–1572 (1969).
[11]C. S. Wang Chang and G. E. Uhlenbeck, "On the propagation of sound in monatomic gases," in *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1970), Vol. V, pp. 43–75.
[12]J. R. Thomas and C. E. Siewert, "Sound-wave propagation in a rarefied-gas," Transp. Theory Stat. Phys. **8**, 219–240 (1979).
[13]S. K. Loyalka and T. C. Cheng, "Sound-wave propagation in a rarefied-gas," Phys. Fluids **22**, 830–836 (1979).
[14]K. Aoki and C. Cercignani, "A technique for time-dependent boundary value problems in the kinetic theory of gases part II. Application to sound propagation," ZAMP **35**, 345–362 (1984).
[15]W. Marques, Jr., "Dispersion and absorption of sound in monatomic gases: An extended kinetic description," J. Acoust. Soc. Am. **106**, 3282–3288 (1999).
[16]N. G. Hadjiconstantinou and A. L. Garcia, "Molecular simulation of sound wave propagation in simple gases," Phys. Fluids **13**, 1040–1046 (2001).
[17]F. Sharipov, W. Marques, Jr., and G. M. Kremer, "Free molecular sound propagation," J. Acoust. Soc. Am. **112**, 395–401 (2002).
[18]R. D. M. Garcia and C. E. Siewert, "The linearized Boltzmann equation: Sound-wave propagation in a rarefied gas," ZAMP **57**, 94–122 (2006).
[19]M. Greenspan, "Propagation of sound in five monatomic gases," J. Acoust. Soc. Am. **28**, 644–648 (1956).
[20]E. Meyer and G. Sessler, "Sound propagation in gases at a high frequency and very low pressure," Z. Phys. **149**, 15–39 (1957).
[21]G. Maidanik and M. Heckl, "Propagation and reflection of sound in rarefied gases. II. Experimental," Phys. Fluids **8**, 266–272 (1965).
[22]R. Schotter, "Rarefied gas acoustics in the noble gases," Phys. Fluids **17**, 1163–1168 (1974).
[23]M. B. Ewing, M. L. McGlashan, and J. P. M. Trusler, "The temperature jump effect and the theory of the thermal boundary layer for a spherical resonator. Speed of sound in argon at 273.16 K," Metrologia **22**, 93–102 (1986).
[24]E. M. Shakhov, "Generalization of the Krook kinetic equation," Fluid Dyn. **3**, 142–145 (1968).
[25]F. M. Sharipov and E. A. Subbotin, "On optimization of the discrete velocity method used in rarefied gas dynamics," ZAMP **44**, 572–577 (1993).
[26]C. Cercignani and M. Lampis, "Kinetic model for gas-surface interaction," Transp. Theory Stat. Phys. **1**, 101–114 (1971).
[27]C. Cercignani, *Theory and Application of the Boltzmann Equation* (Scottish Academic Press, Edinburgh, 1975).
[28]J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases* (North-Holland, Amsterdam, 1972).
[29]E. M. Shakhov, *Method of Investigation of Rarefied Gas Flows* (Nauka, Moscow, 1974).
[30]F. Sharipov, "Application of the Cercignani-Lampis scattering kernel to calculations of rarefied gas flows. I. Plane flow between two parallel plates," Eur. J. Mech. B/Fluids **21**, 113–123 (2002).
[31]F. Sharipov, "Application of the Cercignani-Lampis scattering kernel to calculations of rarefied gas flows. II. Slip and jump coefficients," Eur. J. Mech. B/Fluids **22**, 133–143 (2003).
[32]F. Sharipov, "Application of the Cercignani-Lampis scattering kernel to calculations of rarefied gas flows. III. Poiseuille flow and thermal creep through a long tube," Eur. J. Mech. B/Fluids **22**, 145–154 (2003).
[33]F. Sharipov and G. Bertoldo, "Heat transfer through a rarefied gas confined between two coaxial cylinders with high radius ratio," J. Vac. Sci. Technol. A **24**, 2087–2093 (2006).
[34]*Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*, 9th ed., edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).
[35]F. Sharipov and D. Kalempa, "Oscillatory Couette flow at arbitrary oscillation frequency over the whole range of the Knudsen number," Microfluid. Nanofluid. **4**, 363–374 (2008).

Torsional waves propagation along a waveguide of arbitrary cross section immersed in a perfect fluid

Z. Fan[a] and M. J. S. Lowe
Department of Mechanical Engineering, Imperial College, London SW7 2AZ, United Kingdom

M. Castaings and C. Bacon
Laboratoire de Mécanique Physique, UMR CNRS 5469, Université Bordeaux 1, 33400 Talence, France

(Received 25 January 2008; revised 7 July 2008; accepted 10 July 2008)

Guided torsional waves in a bar with a noncircular cross section have been exploited by previous researchers to measure the density of fluids. However, due to the complexity of the wave behavior in the noncircular cross-sectional shape, the previous theory can only provide an approximate prediction; thus the accuracy of the measurement has been compromised. In this paper, a semianalytical finite element method is developed to model accurately the propagation velocity and leakage of guided waves along an immersed waveguide with arbitrary noncircular cross section. An accurate inverse model is then provided to measure the density of the fluid by measuring the change of the torsional wave speed. Experimental results obtained with a rectangular bar in a range of fluids show very good agreement with the theoretical predictions. Finally, the potentials to use the model for sensor optimization are discussed.
 [DOI: 10.1121/1.2968677]

PACS number(s): 43.35.Bf, 43.35.Yb [RR] Pages: 2002–2010

I. INTRODUCTION

The concept of an ultrasonic "dipstick" is attractive for use in industry for fluid characterizations. The idea is that an ultrasonic wave which propagates in a solid structure can sense the presence and nature of the adjacent fluids. Figure 1 shows an application of a dipstick sensor to measure fluid density. When a torsional wave pulse propagates along a waveguide submerged in a fluid, it interacts at the boundary with surrounding fluid. As a result, the boundary layer of the fluid is alternately accelerated and decelerated. If the waveguide has a noncircular cross section, normal forces are exerted on the surrounding fluid, thus some fluid will be trapped at the corners of the cross section and will affect the propagation of the torsional wave. This mechanism can be attributed to the inertia of the surrounding fluid and is characterized by the density of the fluid. Hence by measuring the speed of propagation of the torsional wave, the density of the fluid can be estimated.

Lynnworth[1] made the first application of such a torsional wave sensor to measure fluid densities. Later, Bau[2] presented a simple quantitative theory to relate the speed of the torsional wave to the density of the surrounding fluid, with a calculation of a two-dimensional, inviscid flow field of the fluid. Based on this theory, various researchers made further applications.[3–5] Kim and Bau[3] analyzed waveguides with various types of cross sections in order to optimize the performance of the sensor. Shepard *et al.*[4] measured the density and viscosity in a variety of fluids, including fluids with high concentration of suspended solids. Smit and Smith[5] made use of a continuous excited torsional wave to allow large cross-sectional dimensions for density measurements. However, due to the complexity of the wave behavior in the noncircular cross-sectional shape, the accuracy of the inversion of the measurements to infer the density of the fluids has been compromised. These authors[2,4,5] have reported the deviation of the measured velocity of torsional waves and theoretical predictions were over 20%. Therefore, better theoretical predictions are required to improve the accuracy of the sensors. In this paper, a finite element (FE) model is applied to predict accurately the wave propagation along a solid bar with a noncircular cross section immersed in a fluid. Then a more precise inverse model relating the group velocity of the torsional wave and the density of the surrounding fluid can be provided.

The modeling work is mainly based on a development of the semianalytical finite element (SAFE) method, addressing the problem of attenuating waves in which the attenuation is caused by leakage from the waveguide into a surrounding fluid. The SAFE method uses a finite element representation of the cross section of the waveguide, thereby enabling arbitrary definition of shapes, together with a harmonic description along the propagation direction.[6–9] The SAFE method was first demonstrated in 1972 for dispersion solutions of solid waveguides of arbitrary geometries.[6] More recently, it was developed to calculate both propagative modes and evanescent modes.[7] Predoi *et al.*[8] have implemented periodic boundary conditions in the SAFE method, which allow the modeling of infinitely wide guides with periodic changes in geometry or material properties along the width. Castaings and Lowe[9] have demonstrated the principles of the SAFE method to study leaky waves which propagate along an elastic waveguide with arbitrary cross section and radiate into a solid of infinite extent. This has been achieved by using an absorbing region to absorb the leaking waves, thus simulating an infinite extent of the solid medium. In this paper, we

[a]Electronic mail: z.fan@imperial.ac.uk

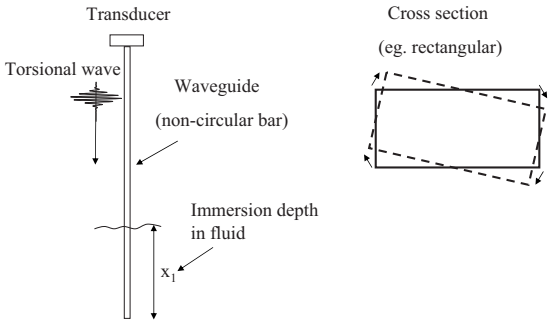


FIG. 1. Schematic of a torsional "dipstick" sensor.

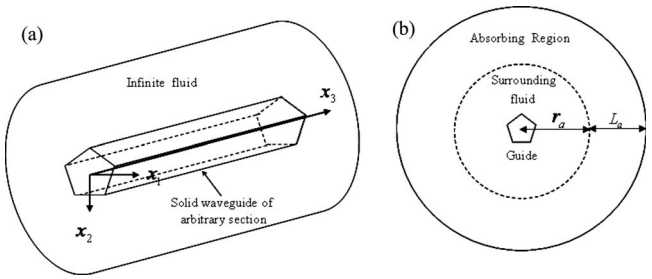


FIG. 2. Schematics of (a) three-dimensional solid waveguide immersed in a fluid and (b) two-dimensional model with absorbing region used for FE simulation of system shown in (a).

present a further development of the modeling capability, addressing the problem of leaky waves propagating along solid waveguides immersed in nonviscous fluids.

This paper starts with a review of the approximate theory[2] and follows with a brief introduction of the SAFE method. In Sec. II, the paper reviews the SAFE method in solids and extends it to solids surrounded by fluids. An absorbing region is also introduced to model fluids of infinite extent. In Sec. IV the model is validated by studying a cylindical steel bar immersed in water, and compared with the results obtained by well-established conventional models[10] which are available for this case of a simple geometry. It is then applied to model a torsional wave propagated along a waveguide with a rectangular cross section that is immersed in a fluid. The velocity change of the torsional wave can be obtained as a function of the density of the fluid, and therefore an accurate inverse model for the density measurements can be provided. In Sec. V experiments are designed to validate the inverse model, showing excellent agreement. General discussions on the optimization of the sensor are presented finally in Sec. VI.

II. PREVIOUS THEORY

Bau's approximate theory suggests that the speed of propagation of torsional waves in a solid, elastic waveguide with a noncircular cross section is inversely proportional to the density of the fluid adjacent to the waveguide.[2] However, in this theory, two assumptions have to be made to achieve the conclusion. The first one is the first-order approximation to the two-dimensional flow field calculation, with which the torsional wave speed c in a solid waveguide can be presented as[2]

$$\frac{C}{C_0} = \left(1 + \frac{\rho_f I_f}{\rho_s I_s}\right)^{-1/2}, \tag{1}$$

where I_s and I_f are the polar moment of inertia[11] of the solid waveguide and the adjacent liquid, respectively, which need to be overcome by the torsional pulse, $C_0 = K(G/\rho_s)^{1/2}$ is the torsional wave speed for a waveguide in vacuum, G is the shear modulus of the solid, $K = (D/I_s)^{1/2}$ is a "shape" factor, D is the torsional rigidity of the cross section, and ρ_s and ρ_f are the densities of the solid and adjacent fluid, respectively.

Kim and Bau[3] have obtained the values of I_f/I_s and K from finite element computation of the flow field around a noncircular cross section. Then a second assumption, $(\rho_f I_f)/(\rho_s Is) \ll 1$, has been made to get the linear relationship:

$$\frac{C}{C_0} \approx 1 - \frac{\rho_f I_f}{2\rho_s I_s}. \tag{2}$$

Thus the density of the fluid ρ_f can be calculated from Eq. (2). However, when I_f/I_s is large or when the density of the solid is close to the density of the fluid, the calculation becomes inaccurate. This can be seen easily, for example, in the case of an aluminum bar with a diamond-shaped cross section which has an aspect ratio 1:3 immersed in alcohol. In this case $I_f/I_s \approx 3$, which can be obtained from Kim's calculation.[3] Thus $(\rho_f I_f)/(\rho_s Is) \approx 1$, and this makes the second assumption invalid.

In addition, any dispersion characteristics of the waves are excluded from this theory, but torsional waves in a geometry with a noncircular cross section are always slightly dispersive, which make the group velocity measured from a pulse-echo experiment slightly different from the phase velocity of the theoretical prediction. This can also lead to deviation between the theory and the experimental measurements.[2,4,5]

In the following sections, an accurate semianalytical finite element model of waves propagating along a noncircular bar immersed in a fluid will be provided, and the dispersion curve (relating frequency and group velocity) of the torsional mode in a certain fluid can be obtained from the model. Therefore, more explicit correlation between the group velocity of the torsional mode at certain frequency and the density of the fluid can be established.

III. SEMIANALYTICAL FINITE ELEMENT (SAFE) METHOD

The schematic of the SAFE model is shown in Fig. 2. The main advantage of the SAFE method compared to a full finite element model is that only the cross section [as shown in Fig. 2(b)], which is normal to the direction of the wave propagation, has to be meshed by finite elements. The waves propagating along the Ox_3 axis are considered to be harmonic.

A. SAFE method in solids

The mathematical model of the SAFE method in solids is based on the three-dimensional elasticity approach. Consequently, the displacement vector in the waveguide can be written:

$$u_i(x_1,x_2,x_3,t) = U_i(x_1,x_2)e^{I(kx_3-\omega t)}, \quad I=\sqrt{-1} \tag{3}$$

in which k is the wave number, $\omega=2\pi f$ is the angular frequency, f being the frequency, t is the time variable, and the subscript $i=1,2,3$. For general anisotropic material, the equation of dynamic equilibrium can be written in the following form of an eigenvalue problem:

$$C_{ikjl}\frac{\partial^2 U_j}{\partial x_k \partial x_l} + I(C_{i3jk}+C_{ikj3})\frac{\partial(kU_j)}{\partial x_k} - kC_{i3j3}(kU_j) + \rho\omega^2\delta_{ij}U_j = 0 \tag{4}$$

with summation over the indices $j=1,2,3$ and $k,l=1,2$. The coefficients C_{ijkl} are the stiffness moduli and δ_{ij} is the Kronecker symbol. In the commercial FEM code[12] used in this study the formalism for eigenvalue problems has the general expression:

$$\nabla\cdot(c\nabla U+\alpha U-\gamma)-\beta\nabla U-aU+\lambda d_a U-\lambda^2 e_a U=0 \tag{5}$$

in which all matrix coefficients are given by Predoi *et al.*[8]

The nature of the solution is thus to find eigenvalues of complex wave number k for chosen values of angular frequency ω. Each solution at a chosen frequency will reveal the wave numbers of all of the possible modes at that frequency; then the full dispersion curve spectrum can be found by repeating the eigenvalue solutions over the desired range of frequencies.

B. SAFE method in perfect fluids

The equation of dynamic equilibrium in the perfect fluid can be written:

$$\nabla\cdot(K_f\nabla P)+\rho\omega^2 p=0 \tag{6}$$

in which p is the pressure of the fluid, and K_f is the compressibility coefficient of the fluid.

When the surface of the fluid is in contact with a deformable solid, we have the boundary equation at the interface:

$$\mathbf{n}\cdot(K_f\nabla p)=\rho\omega^2 K_f\mathbf{n}\cdot\mathbf{u}^{(\text{solid})}, \tag{7}$$

where $\mathbf{n}$ is the outward unit vector of the fluid domain on the interface and $\mathbf{u}^{\text{solid}}$ is the displacement of the interface calculated in the solid domain. For the solid waveguide, the boundary condition involves the pressure p in the fluid by writing that the stress vector is $-p\mathbf{n}$, $\mathbf{n}$ being the outward unit vector of the solid.

For wave propagation along the Ox_3 direction, the pressure of the fluid can be written as

$$p(x_1,x_2,x_3,t)=p(x_1,x_2)e^{I(kx_3-\omega t)}, \quad I=\sqrt{-I}. \tag{8}$$

By combining Eqs. (6) and (7) and comparing with Eq. (5), the pressure can be chosen as the finite element variable and the coefficients become

$$c=K_f, \quad a=-\rho\omega^2, \quad d_a=\alpha=\beta=\gamma=0, \quad e_a=K_f. \tag{9}$$

C. Absorbing region

In order to solve the problem of a solid waveguide immersed in an infinite fluid, an exterior absorbing region is needed to model the surrounding medium.[9] This region, shown in Fig. 2, has the same mass density as the fluid but has damping properties which increase with the distance away from the central axis of the system. To achieve this, the imaginary part of its compressibility coefficient gradually increases according to the following law:

$$K_{fa}=K_f\left[1+I\alpha_1\left(\frac{|r-r_a|}{L_a}\right)^3\right], \quad I=\sqrt{-1}, \tag{10}$$

where K_f represents the compressibility of the liquid, r_a is the inner radius of the absorbing region, L_a is its radial length, and r is the radial position in this absorbing region. α_1 is a coefficient that defines the proportion of the damping at the outer limit of the absorbing region.

By introducing the imaginary part of the compressibility, the propagation wave numbers, which are eigensolutions of the system, become complex ($k=k'+Ik''$). The imaginary parts (k'') represent the attenuation due to leakage from the bar to the infinite fluid. In the previous studies of a leaky waveguide surrounded by a solid of infinite extent,[9] the length of the absorbing region was recommended to be two to three times the biggest wavelength of any radiated wave in the whole frequency range. However, according to our numerical testing results here, we found that if the surrounding material is fluid, the length of the absorbing region can be much shorter. The reason is because the leaking energy carried in the fluid is much smaller than the energy carried in the solid waveguide. Therefore, the relative amount of fluid-borne energy that is reflected back to the waveguide is very small, which can be further reduced by a short absorbing region, and will not affect the eigensolutions of the system. In our numerical testing, the length of the absorbing region was proved to be efficient by a convergence check. When the length was increased we will still obtained the same solutions for the propagating modes in the waveguide.

IV. NUMERICAL CALCULATION

A. SAFE method validation

The first numerical study is to validate the SAFE method for immersed waveguides. A 1 mm radius steel cylinder bar immersed in water is considered. The example is chosen because it can be fully studied by the software DISPERSE,[13] which is based on the global matrix method.[10] (It can also be done by some other treatments.[14,15]) The geometry of the system is shown in Fig. 3. The steel bar is 2 mm in diameter and the surrounding water is modeled by a 4-mm-thick ring having an inner diameter of 2 mm. The absorbing region is

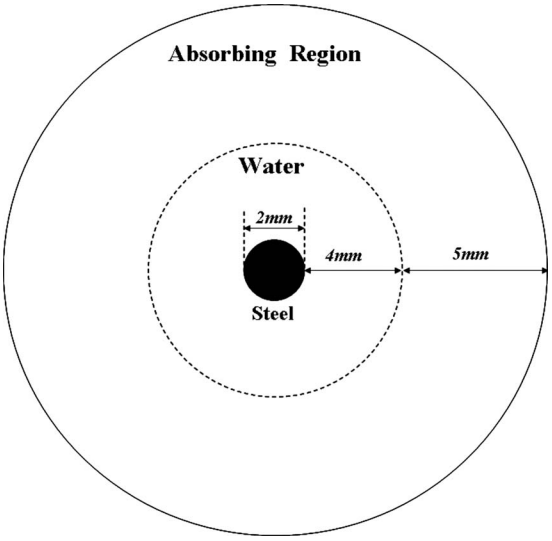


FIG. 3. Schematic of the FE model used for the circular steel bar immersed in water.

modeled by a 5-mm-thick ring having an inner diameter of 10 mm. The material properties are given in Table I. The whole geometry is meshed by 7563 triangular elements of first order (each element has three nodes), which are automatically generated by the software used.[12] The number of degrees of freedom is 14 912.

The system is solved using the SAFE method to find values of the wave number k at different frequencies. For each frequency, several solutions are obtained. For each solution, the amplitude of normal stress in the radial direction T_{rr} is calculated at each nodal position in the solid domain and the pressure p is calculated at each nodal position in the fluid domain and in the absorbing region. These quantities are equal at the border between the solid and fluid according to the imposed boundary condition. Solutions which have higher values of T_{rr} in the solid domain than $-p$ in the fluid domain generally represent modes guided along the bar and radiating in the water, while other solutions represent resonances of the whole system and are unwanted.

Figure 4 shows SAFE solutions at 500 kHz. There are three propagating modes existing at this frequency: the $L(0,1)$ mode, which is shown in Fig. 4(a), the $T(0,1)$ mode, which is shown in Fig. 4(b), and the $F(1,1)$ mode, which is shown in Fig. 4(c). Figure 4(d) shows an unwanted solution that corresponds to a resonance of the absorbing region. [Here $L(n,m)$, $T(n,m)$, and $F(n,m)$ are the longitudinal, torsional, and flexural modes, respectively, where n is an angular index and m is a radial index.[16]] From the figure, it can be seen in the longitudinal mode that the radial normal stress is concentrated in the center of the bar and some energy is radiating to the water; in the torsional mode the radial normal stress is almost zero (theoretically it should be zero, but there is a very small value due to the numerical approximation); in the flexural mode the radial normal stress is symmetric with respect to a diameter of the bar and energy is radiating to the water.

TABLE I. Mechanical properties for materials used in SAFE modeling.

	Density (kg/m^3)	Bulk longitudinal velocity (m/s)	Bulk shear velocity (m/s)
Aluminum	2700	6320	3130
Copper	8900	4700	2260
Steel	7932	5959	3260
Magnesium	1700	5770	3050
Water	1000	1500	
Alcohol	800	1168	

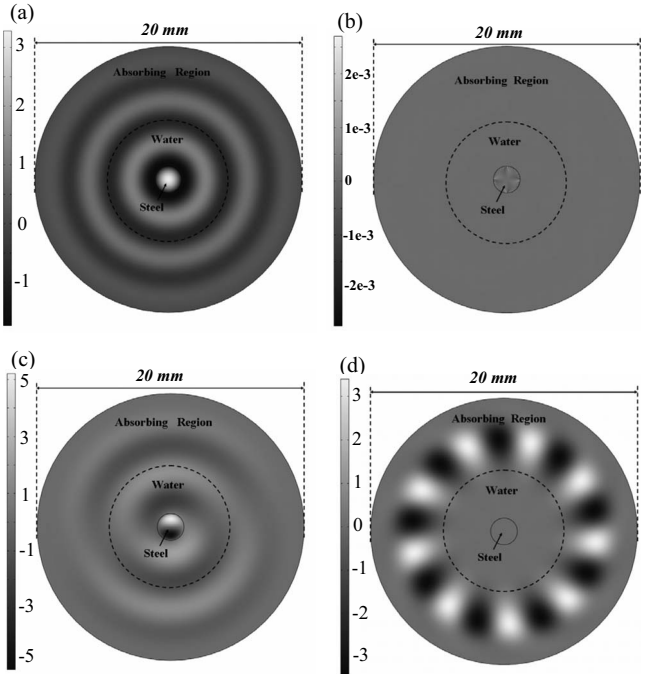


FIG. 4. Cross-section distribution of normal stress in solid and pressure in fluid at 500 kHz for example modal results: (a) $L(0,1)$ mode, (b) $T(0,1)$ mode, (c) $F(1,1)$ mode, and (d) mode resonating in the absorbing region.

The system is solved for 71 frequencies over the range from 100 to 1500 kHz, and solutions which represent the propagating modes are sought according to the above-presented rule. The mode shape information (displacements of x_1, x_2, and x_3 directions) is recorded at each nodal position for each sought eigensolution. By comparing these mode shapes, all the solutions can be classified into the different modes. Figure 5 presents the dispersion curves of wave modes propagating along the steel bar and eventually radiating energy in the infinite water from 100 to 1500 kHz, showing the phase velocity, real wave number, group velocity, and attenuation. The real wave number and the attenuation can be obtained from the eigensolutions directly, while the phase velocity can be calculated by $C_{\mathrm{ph}}=\omega/k'$ and the group velocity is obtained by doing a numerical derivation $C_{\mathrm{gr}}=d\omega/dk'$. Plain lines are predictions made with the DISPERSE software, while circles represent the SAFE solutions obtained with the model.

From Fig. 5 it can be seen that the SAFE predictions have good agreement with DISPERSE results at most of the frequencies. The only disagreement appears at 250–400 kHz of the $F(1,1)$ mode on the group velocity and attenuation curves, which is a result of inefficiency of the absorbing region at these frequencies. According to the Snell–

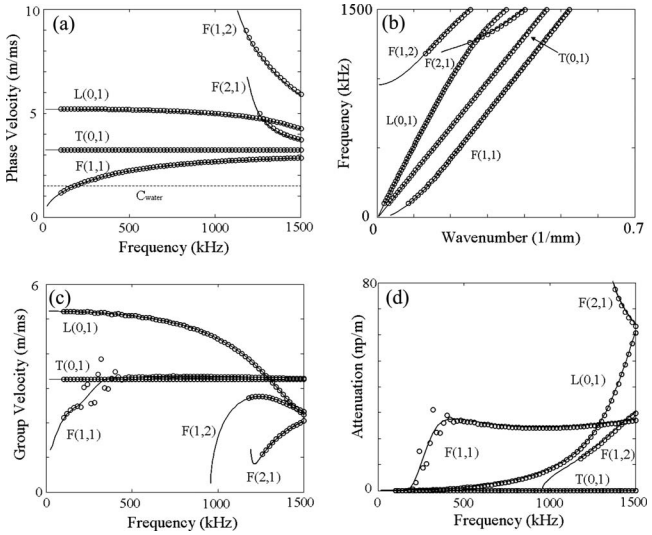


FIG. 5. Dispersion curves of phase velocity (a), wave number (b), group velocity (c), and attenuation (d) of 1 mm radius circular steel cylinder bar immersed in water, predicted by SAFE method (circles) and DISPERSE (lines).

Descartes' law,[17] the angles of radiation, with respect to the normal to the surface of the waveguide, are determined by the following formula:

$$\sin \theta_{\text{rad}} = \frac{C_{\text{water}}}{C_{\text{ph}}} (C_{\text{water}} < C_{\text{ph}}), \tag{11}$$

where C_{water} is the bulk velocity in water and C_{ph} is the phase velocity of the radiating mode. There is no radiation for $C_{\text{water}} > C_{\text{ph}}$. From Fig. 5(a), it can be seen that the phase velocity of the $F(1,1)$ mode at 250–400 kHz is just above the bulk velocity of water, therefore leaky waves have large angles of radiation θ_{rad}, and the absorbing region does not perform well in such cases according to the previous studies.[9] A typical calculation of one SAFE model presented here takes approximately half a minute on a Pentium 4 PC with 1 Gbyte memory. By increasing the length of the absorbing region, the inaccurate frequency range can be reduced, but it will be much more time consuming to solve the model. Since the torsional mode is our main interest and the SAFE method is very accurate for this mode, it is not necessary to increase the length of the absorbing region.

B. SAFE method for rectangular bar immersed in a perfect fluid

This study applies the validated SAFE model to the practical case of a rectangular bar immersed in a fluid. As was shown in the previous model, by the SAFE method one can accurately predict the propagation and leakage of guided waves along an immersed waveguide. Therefore, by performing accurate predictions, such as are undertaken here, it should be possible to construct an inverse relationship so that measurements of the torsional wave could be used to infer the density of the fluid.

The example model shown here is an aluminum bar with rectangular cross section (1.1 mm × 2.2 mm) immersed in alcohol. The length L_a of the absorbing region is 10 mm, which is tested to be long enough to suppress reflections from the outer border for torsional modes at all frequencies. The material properties are shown in Table I. The geometry is meshed by 8376 triangular elements of first order. The number of degrees of freedom is 17 282, and the system can be solved at different frequencies. Figure 6 shows results for the fundamental torsional mode at 50 kHz. Figure 6(a) shows the whole model and Fig. 6(b) shows a zoom of the bar and nearby fluid. The radial stress (with respect to the center of the bar) and pressure in the fluid is displayed as a grey scale, and the displacement in the fluid and the cross section of the bar is plotted by arrows. It can be clearly seen that the fluid is trapped by the corners of the bar, thus it is to be expected that the propagating speed of the torsional mode along the bar should be influenced by the fluid.

The dispersion curves of this case can be generated by repeating the SAFE solution over a desired range of frequencies, and the fundamental torsional mode can be traced by comparing the mode shapes of all the propagating solutions at each frequency. Figure 7 presents the phase velocity dispersion curve of the fundamental torsional mode of the alu-

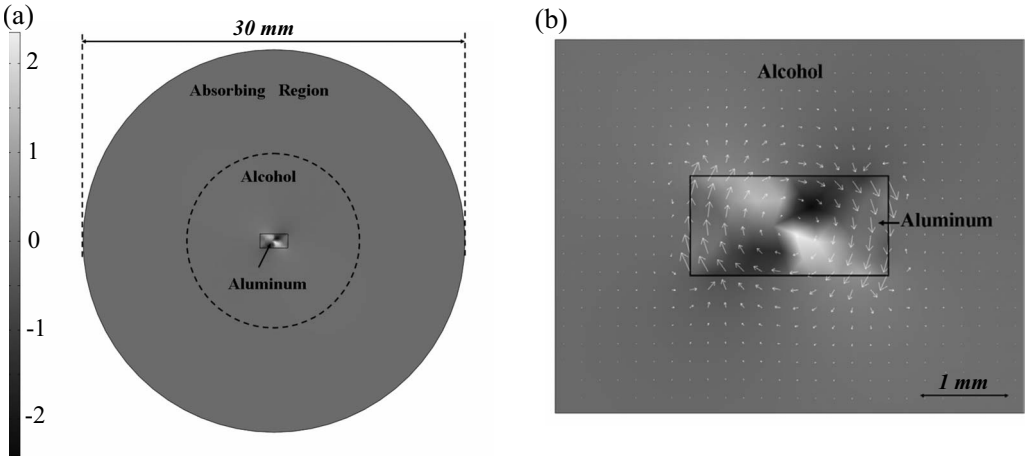


FIG. 6. Theoretical prediction of mode shape of torsional mode of aluminum bar with rectangular cross section (1.1 mm × 2.2 mm) immersed in alcohol at 50 kHz. The radial (with respect to the center of the bar) stress in solid and pressure in fluid is displayed in a grey scale. Arrows indicate displacements in cross section of fluid and the bar.

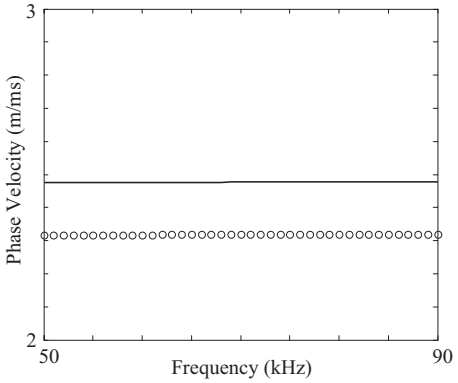


FIG. 7. Phase velocity dispersion curves of aluminum bar with rectangular cross section (1.1 mm × 2.2 mm) immersed in alcohol (circles) and in vacuum (line) predicted by SAFE method.

minum rectangular bar immersed in alcohol from 50 to 90 kHz. For comparison, the phase velocity of the fundamental torsional mode of the same rectangular aluminum bar in vacuum is also plotted in Fig. 7 (line), also generated by the SAFE method but omitting the fluid.[8] It can be seen that the torsional speed of the waveguide with the rectangular cross section decreases when it is immersed in the fluid. In addition, it can be found that this particular mode has almost no dispersion over this frequency range, and is therefore useful for measurements.

From the calculation, we also found that the attenuation of the torsional mode of the aluminum rectangular bar immersed in alcohol is zero, which means there is no leakage from the bar to the fluid in the above-mentioned frequency range. Therefore, the group velocity is identical to the energy velocity,[18] the latter being the quantity which was actually measured in the experiments to follow.

V. EXPERIMENT

A. Experimental setup

An experimental setup was designed to validate the model by exciting the torsional mode in an aluminum rectangular bar immersed in a fluid and measuring its energy velocity (group velocity). Figure 8 shows a schematic of the apparatus. The bar was 450 mm long with rectangular cross section (1.1 mm × 2.2 mm), and it had the same properties as were used for the model. A vessel containing a fluid sample was placed beneath the bar on a table of variable height. By changing the height of the table the bar could conveniently be immersed in the fluid to different depths; the angle between the fluid surface and the axis of the bar was 90°. The signal was sent and received by a pulse generator and receiver unit (Macro Design Ltd., Leeds, UK), a LeCroy 9400A Storage Oscilloscope was used to store the signal and data were then transferred to a computer for processing. A pair of standard shear transducers made by Guided Ultrasonics Ltd. (London, UK) was used to excite the torsional mode with a five-cycle Hanning windowed tone burst. The signal was reflected from the end of the bar and traveled back to the transducer. It was then recorded after 50 averages to reduce the noise.

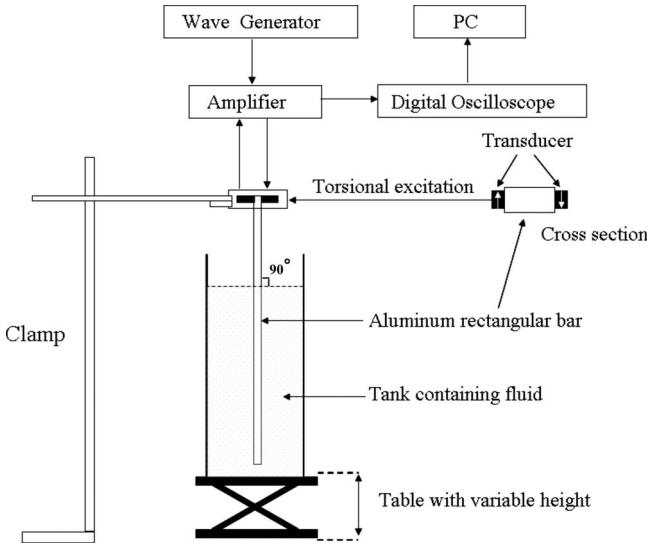


FIG. 8. Experimental setup.

A typical time trace of this bar partly immersed in alcohol is displayed in Fig. 9(a) and for comparison the signal in air in shown in Fig. 9(b). The first and second bar end reflections and reverberation in the fluid are clearly visible. ΔT is the flight time of the wave packages propagating twice the length of the bar, which can be determined by calculating the shift of the Hilbert envelope of the measured signals. The phase of measured signals is inverted after each reflection because one end of the bar was clamped by two transducers

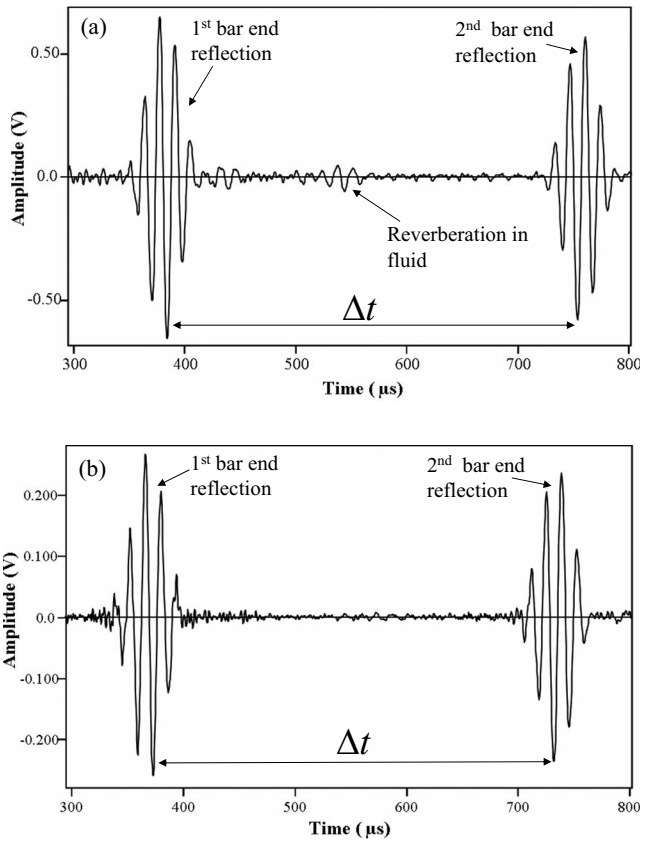


FIG. 9. Time trace at 70 kHz with aluminum rectangular bar immersed in alcohol at 170 mm (a) and in air (b).

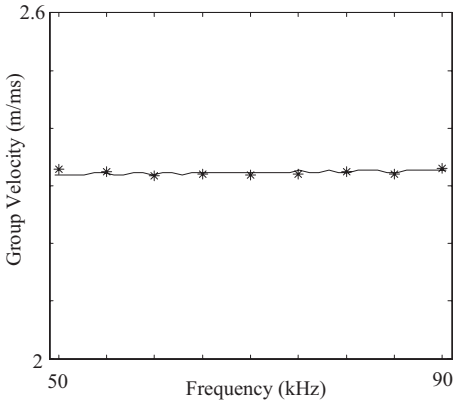


FIG. 10. Measured (stars) and theoretically predicted (line) group velocity of aluminum bar with rectangular cross section immersed in alcohol.

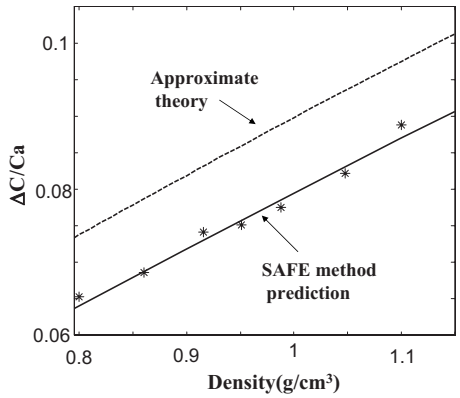


FIG. 11. Measured (stars) and theoretically predicted (line) group velocity of aluminum bar with rectangular cross section immersed in different fluids compared with previous theory prediction (dashed line).

in the measurements. Data for tone bursts at different center frequencies and two different immersion depths were collected. The group velocity C_{gr} was extracted from the measured signals using

$$C_{\mathrm{gr}} = \frac{2(x_2 - x_1)}{(\Delta T_2 - \Delta T_1) + \dfrac{2(x_2 - x_1)}{C_a}}, \tag{12}$$

where $x_2 > x_1$ are two different immersion depths, ΔT_1 and ΔT_2 are the flight times of the wave package propagating twice the length of the bar with different immersion depths. C_a is the group velocity of the torsional mode of the bar in air (which is considered to be a good approximation for group velocity in vacuum C_0).

B. Results

1. Alcohol

The first experiment is to validate the model using pure alcohol and the frequency range from 50 to 90 kHz. The density of a sample of alcohol (800 kg/m^3) was measured by a conventional measurement (by measuring the weight and volume of the liquid) and the aluminum bar properties were evaluated experimentally $\rho = 2700$ kg/m^3, $C_l = 6320$ m/s, $C_s = 3130$ m/s. The temperature was recorded to be 25 °C. Results were extracted from the time traces as described in Sec. V A. Figure 10 shows the measured group velocity of the torsional mode as a function of frequency, and the theoretically predicted curve by the SAFE method; this was obtained by doing a numerical derivation $C_{\mathrm{gr}} = d\omega/dk'$ as we described in Sec. IV. From Fig. 10 we can see that the measured results agree very well with the theoretical predictions.

2. Density measurements

In order to verify the theoretical predictions of the fluid density by the SAFE method and compare with the previous approximate theory, a second experiment was carried out. A few fluid samples with density from 800 to 1100 kg/m^3 were chosen, the variation being achieved by changing the concentration of alcohol and salt with water. The center frequency of the tone burst signal was selected to be 70 kHz. Figure 11 depicts the ratio $(C_a - C_{\mathrm{gr}})/C_a$ as a function of the fluid density. The solid line shows the SAFE model results while the stars are the experiment results. The previous approximate theoretical prediction[2] (dashed lines), according to Eq. (1), is also shown in Fig. 11. It can be seen that the SAFE method predictions agree very well with the measurements, and that this represents a substantial improvement with respect to the approximate model.

VI. POTENTIAL FOR SENSOR OPTIMIZATION

With the SAFE method, one can easily predict the torsional wave speed along bars with any arbitrary cross sections immersed in fluids. Thus the model can be used as a powerful tool to optimize the sensitivity of the dipstick by changing aspect ratio, geometry of the cross section, and material properties of the bar. Compared to the previous simple theory, this can not only improve the accuracy of the density prediction, but can also include the dispersion information. Therefore, one can choose ideal shapes as well as frequencies for measurements according to the calculation.

Figure 12(a) shows the measurement sensitivity of aluminum rectangular bars with aspect ratio from 1:1 to 1:4. Similar inverse models as shown in Sec. V were made by calculating the ratio $(C_a - C_{\mathrm{gr}})/C_a$ as a function of the fluid density. From our calculation, we found that $(C_a - C_{\mathrm{gr}})/C_a$ is almost linear with density, therefore the sensitivity can be presented by the slopes of the lines. From Fig. 12 it can be seen that the sensitivity increases as the aspect ratio increases, thus it would seem desirable to operate with as large

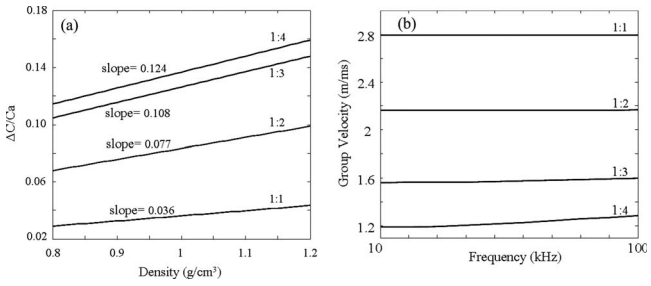


FIG. 12. (a) Measurement sensitivity comparison for aluminum rectangular bars with aspect ratio from 1:1 to 1:4. (b) Dispersion comparison for aluminum rectangular bars with aspect ratio from 1:1 to 1:4.

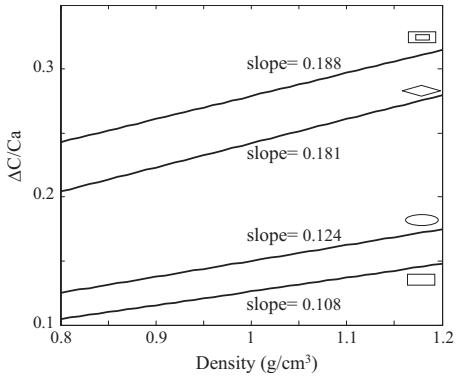


FIG. 13. Measurement sensitivity comparison for aluminum bars of rectangular, elliptical, diamond-shaped, and hollow rectangular cross sections with aspect ratio of 1:3.

an aspect ratio as possible. However, the aspect ratio cannot be increased without limit. Figure 12(b) presents the phase velocity dispersion curves of these aluminum rectangular bars immersed in alcohol from 10 to 100 kHz. It can be seen that as the aspect ratio increases the torsional mode becomes more and more dispersive, thus for bars with large aspect ratios measurements are not practical at some frequencies. This conclusion has also been mentioned in the previous work by Kim and Bau.[3] With the SAFE method the dispersion characteristics of the waves can be accurately quantified, which helps to design the most practical aspect ratio.

For a given ratio, the sensitivity can also be optimized by the choice of geometrical configuration for the cross section.[3] Figure 13 compares the sensitivity of aluminum bar of rectangular, elliptical, and diamond-shaped cross sections with the same aspect ratio of 1:3. In addition, the sensitivity of a rectangular waveguide (1 mm $\times$ 3 mm) with a rectangular hole (0.5 mm $\times$ 1.5 mm) in the middle, which cannot be predicted by the previous theory, is also shown in Fig. 13. The results show that the diamond-shaped cross section outperforms the elliptical one, and the elliptical one has better sensitivity than the rectangular cross section. The hollow rectangular waveguide has similar sensitivity to the diamond-shape cross section with the same aspect ratio.

The sensitivity also changes when different materials of the bar are chosen. Figure 14 compares the sensitivity of bars with diamond-shaped cross sections (with axes 1 and 3 mm)

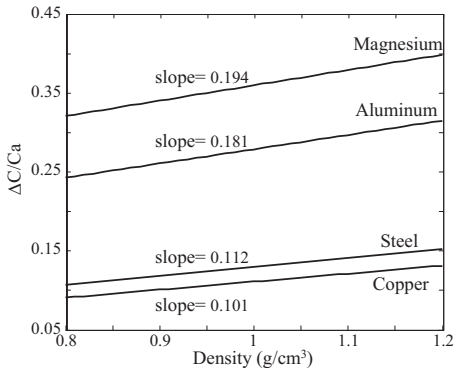


FIG. 14. Measurement sensitivity comparison for bars which are made of copper, steel, aluminum, and magnesium of diamond-shaped cross sections with aspect ratio of 1:3.

which are made of copper, steel, aluminum, and magnesium. The parameters of the materials are shown in Table I. It can be seen that the measurement becomes more sensitive when the density of the solid bar is closer to the fluid, which can also be explained by Eq. (2) from the approximate theory.

VII. CONCLUSIONS

The torsional mode of a noncircular waveguide has previously been employed in fluid density measurements but the accuracy was compromised by the lack of an accurate model. In this paper, the semianalytical finite element method has been extended for the study of solid waveguides immersed in fluids in which the guided waves propagate along the bar and are influenced by the fluid. The prediction of the model may include modes which attenuate by leakage of energy into radiating waves in the fluid. This has been achieved by using established absorbing region modeling techniques in order to absorb the leaking waves and thus simulate an infinite extent of the surrounding fluids. The method has been validated by studying a cylindrical bar immersed in water and comparing with analytical results. An accurate model has thus been developed to enable velocity measurements to be used to determine the density of the fluid. Experiments have been carried out to verify the model on a variety of fluids, showing very good agreement. Optimization of the sensor has been illustrated by adjusting the material and cross-section shape of the solid bar.

It should also be noted that only inviscid fluids were considered in the presented model. However lots of real fluids have some viscosity, which will undoubtedly affect the propagation of the torsional mode on an immersed solid waveguide. A further development of the SAFE model to address the viscous effect is under way.

ACKNOWLEDGMENT

The authors are grateful to the Engineering and Physical Science Research Council (EPSRC) for funding this work.

[1]L. C. Lynnworth, "Slow torsional wave sensors," Proc.-IEEE Ultrason. Symp. **1977**, 29–34 (1977).
[2]H. H. Bau, "Torsional wave senser—A theory," Trans. ASME, J. Appl. Mech. **53**, 846–848 (1986).
[3]J. O. Kim and H. H. Bau, "On line, real-time densimeter—Theory and optimization," J. Acoust. Soc. Am. **85**, 432–439 (1989).
[4]C. L. Shepard, B. J. Burghard, M. A. Friesel, B. P. Hildebrand, X. Moua, A. A. Diaz, and C. W. Enderlin, "Measurements of density and viscosity of one- and two-phase fluids with torsional waveguides," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 536–548 (1999).
[5]C. C. J. Smit and E. D. Smith, "The analysis and results of a continuous wave ultrasonic densitometer," J. Acoust. Soc. Am. **104**, 1413–1417 (1998).
[6]P. Lagasse, "Dispersion of axially symmetric waves in empty and fluid-filled cylindrical shells," Acustica **27**, 317–329 (1972).
[7]V. Damljanovic and R. L. Weaver, "Propagating and evanescent elastic waves in cylindrical waveguides of arbitrary cross section," J. Acoust. Soc. Am. **115**, 1572–1581 (2004).
[8]M. V. Predoi, M. Castaings, B. Hosten, and C. Bacon, "Wave propagation along transversely periodic structures," J. Acoust. Soc. Am. **121**, 1935–1944 (2007).
[9]M. Castaings and M. J. S. Lowe, "Finite element model for waves guided along solid systems of arbitrary section coupled to infinite solid media," J. Acoust. Soc. Am. **123**, 696–708 (2008).
[10]M. J. S. Lowe, "Matrix techniques for modelling ultrasonic waves in mul-

tilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[11]A. C. Ugural and S. K. Fenster, *Advanced Strength and Applied Elasticity* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[12]COMSOL, *User's Guide and Introduction* (Version 3.3 by-COMSOL AB 2007, http://www.comsol.com/, most recently viewed 20th December 2007).
[13]B. N. Pavlakovic, M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "Disperse: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti, (Plenum, New York, 1997), Vol. **16**, pp. 185–192.
[14]P. Marston, "Approximate meridional leaky ray amplitudes for titled cylinders: End-backscattering enhancements and comparisons with exact theory for infinite solid cylinders," J. Acoust. Soc. Am. **102**, 358–369 (1997).
[15]P. B. Nagy, "Longitudinal guided waves propagation in a transversely isotropic rod immersed in fluid," J. Acoust. Soc. Am. **98**, 454–457 (1995).
[16]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, UK, 1999).
[17]B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. **1**.
[18]A. Bernard, M. J. S. Lowe, and M. Deschamps, "Guided waves energy velocity in absorbing and non-absorbing plates," J. Acoust. Soc. Am. **110**, 186–196 (2001).

Shock-induced collapse of a gas bubble in shockwave lithotripsy

Eric Johnsen[a] and Tim Colonius[b]
Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125

(Received 23 April 2008; revised 18 July 2008; accepted 21 July 2008)

The shock-induced collapse of a pre-existing nucleus near a solid surface in the focal region of a lithotripter is investigated. The entire flow field of the collapse of a single gas bubble subjected to a lithotripter pulse is simulated using a high-order accurate shock- and interface-capturing scheme, and the wall pressure is considered as an indication of potential damage. Results from the computations show the same qualitative behavior as that observed in experiments: a re-entrant jet forms in the direction of propagation of the pulse and penetrates the bubble during collapse, ultimately hitting the distal side and generating a water-hammer shock. As a result of the propagation of this wave, wall pressures on the order of 1 GPa may be achieved for bubbles collapsing close to the wall. The wall pressure decreases with initial stand-off distance and pulse width and increases with pulse amplitude. For the stand-off distances considered in the present work, the wall pressure due to bubble collapse is larger than that due to the incoming shockwave; the region over which this holds may extend to ten initial radii. The present results indicate that shock-induced collapse is a mechanism with high potential for damage in shockwave lithotripsy.
 [DOI: 10.1121/1.2973229]

PACS number(s): 43.35.Ei, 43.25.Yw, 43.40.Ng, 43.28.Mw [CCC] Pages: 2011–2020

I. INTRODUCTION

In shockwave lithotripsy (SWL), the most common treatment for kidney stone disease,[1] focused shockwaves are used to pulverize kidney stones. In a typical procedure, several hundreds to thousands of shockwaves are fired at rates of 0.5–2 Hz.[2] In the focal region, a lithotripter pulse consists of a steep compressive shock front followed by a longer duration expansion tail with a tensile (negative) pressure. Since kidney stones can be immersed in urine and possibly in pooled blood, this tension can lead to the formation and growth of bubbles near the stone. Though the precise mechanisms of stone comminution are still debated,[3] the two mechanisms thought to be most important are stress waves propagating within the stone[4–6] and cavitation erosion due to bubble collapse along the stone surface.[7,8] The mechanisms by which cavitation bubbles damage surfaces are complex.[9,10] The bubble collapse is nonspherical, as illustrated by the formation of a re-entrant jet,[11] and shockwaves are generated when the jet impacts the distal side, which in turn leads to a secondary loading of the stone. These processes are further complicated by the formation of bubble clusters or clouds on the surface of the solid.[12] While previous numerical studies considered averaged models for bubble clouds produced in SWL,[13] direct simulation of the detailed nonspherical near-surface bubble collapse under conditions relevant to SWL has not yet been reported.

In SWL, it is clear (at least *in vitro*) that nucleation of bubbles takes place at existing gas nuclei, which become more numerous as the treatment progresses.[12] Due to the tensile portion of the lithotripter pulse, vapor bubbles grow and gather along the surface of the stone before collapsing. This (Rayleigh) collapse[14] of vapor bubbles occurs in SWL at a time of O(100) μs after the passage of the pulse. However, this phenomenon is preceded by the collapse of pre-existing gas nuclei on a time scale of microseconds under the effect of the compressive portion of the shock. Though bubbles undergoing this so-called *shock-induced collapse* (SIC) are initially fewer than those undergoing Rayleigh collapse, the number of gas bubbles is roughly equal to that of cavitation bubbles after a sufficient number of shocks have passed; furthermore, sonoluminescence intensity measurements suggest that higher temperatures are achieved in SIC,[15] thus highlighting the importance of this phenomenon in SWL and motivating the present work. In this paper, we provide the first systematic study of the bubble dynamics and damage potential associated with SIC in SWL. We show that the collapse of even tiny existing bubbles with a diameter of O(10) μm can lead to locally high pressures on the order of 1 GPa on the stone surface. We note that Rayleigh collapse of vapor bubbles has been shown to be important for cavitation erosion;[9,10] however, existing numerical algorithms are unable to simulate satisfactory Rayleigh collapse[16] so that comparisons between SIC and Rayleigh collapse are not currently possible.

Because of the tremendous challenges of resolving the very fast and small scales occurring in bubble collapse both experimentally and computationally, only a limited amount of results has been reported. Experimentally, Rayleigh collapse of a cavitation bubble[9,10,17] has been studied more extensively than SIC.[18–20] Computationally, shock-capturing methods solving the full compressible Euler equations are currently being developed and validated using two-dimensional interface problems so that both shockwaves and

[a]Electronic mail: johnsen@stanford.edu; present address: Center for Turbulence Research, Stanford University, CA 94305-3030.
[b]Electronic mail: colonius@caltech.edu

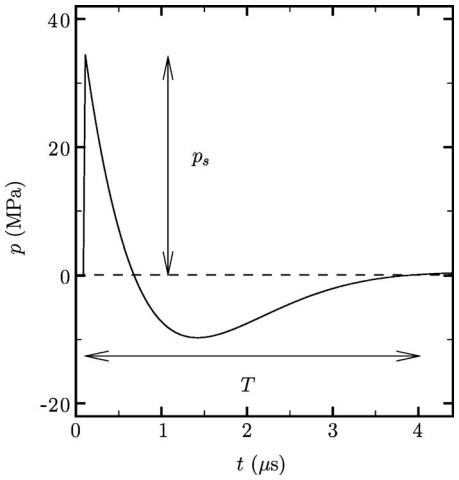


FIG. 1. Pressure waveform of a lithotripter pulse at the focus.

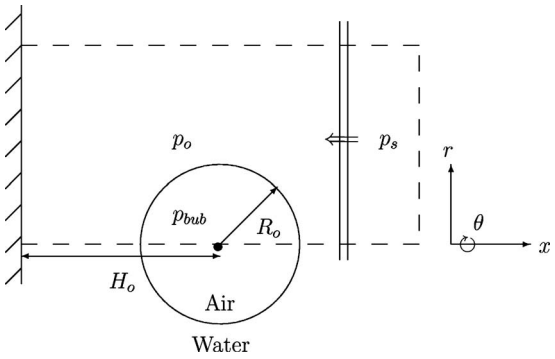


FIG. 2. Schematic of the problem geometry.

interfaces are handled appropriately.[21–23] However, few systematic studies of bubble collapse using such methods have been published at this time.[24–26]

To simulate SIC, a high-order accurate shock- and interface-capturing scheme was developed[27] and is used here to simulate the full flow field generated by the interaction of a lithotripter pulse with a gas bubble near a solid surface. This allows the visualization of the bubble response and the shockwaves generated during the process; in particular, the wall pressure is computed, thus giving a measure of potential damage. A detailed description of the methodology is presented in Sec. II, where the modeling of the lithotripter pulse and the problem setup are first discussed; then, the equations of motion and the numerical method are stated. The overall behavior of the bubble collapse is described in Sec. III. Then, the damage potential of SIC in SWL is examined by performing a parameter study in Sec. IV. Finally, the article ends with a summary of the findings and an outlook for future work.

II. METHODOLOGY

A. Problem setup

The focal region of a lithotripter is characterized by a cigar-shaped area where the wavefront is nearly planar. For simplicity, we assume that the stone is larger than this focal zone, which itself is much larger than the pre-existing spherical bubble. The lithotripter pulse is modeled as a compressive shock front of constant amplitude followed by a long expansion tail that includes a negative (tensile) pressure. The waveform is represented by the following analytical function:[28]

$$p(t) = p_o + 2p_s e^{-\alpha t}\cos\left(\omega t + \frac{\pi}{3}\right), \tag{1}$$

where p_o is the atmospheric pressure. The parameters are chosen to closely match the waveform in a Dornier HM3 lithotripter shown in Fig. 1, with nominal values of $\alpha = 1.48\times10^8$ s^{-1}, $\omega = 1.21\times10^8$ s^{-1}, and $p_s = 35$ MPa; other types of lithotripters generate at least qualitatively similar waveforms. The pulse amplitude, p_s, and characteristic time, T, are defined in the figure; in particular, T is equal to the second zero of the function given by Eq. (1) (i.e., the time when the pressure becomes positive again after the first tensile region). The pulse width is then given by $\sigma = sT$, where s is the speed of propagation of the pulse and is approximately equal to the sound speed in water in the present cases. The values of α and ω lead to a nominal pulse width of $\sigma = 6.75$ mm; in order to understand the dependence of the results on the pulse width, σ is varied in Sec. IV C. Typical peak positive pressures at the focus range from 9 to 114 MPa and negative (tensile) pressures as low as −10 MPa have been measured.[29] We also vary the pulse amplitude in Sec. IV B to assess its impact on bubble collapse. The kidney stone is assumed to have infinite impedance so that all waves are completely reflected with no losses; this results in approximate pressure doubling at the stone surface, which can be understood by considering the reflection of a planar shockwave in water off a solid surface.[16,30] In reality, kidney stones have a finite impedance approximately three to five times that of water.[31,32]

A slice through the center of the computational domain is shown in Fig. 2. The domain consists of a cylinder, along whose centerline an isolated spherical air bubble of radius, R_o, is initially in equilibrium with its surroundings. When present, the wall constitutes the left boundary of the domain. The initial stand-off distance is denoted H_o and the dependence of the wall pressure on this parameter is studied in Sec. IV A. As the simulation progresses, the bubble collapses nonspherically. We record the full flow field at different times and compute an average bubble radius, $R(t)$, based on the bubble volume, and an average stand-off distance, $H(t)$, estimated using the position of the bubble centroid. Appropriate boundary conditions are used at the edge of the cylinder to extend the domain to infinity, while reflecting conditions are employed to represent the wall. The shock described by Eq. (1) is initialized inside the domain and propagates toward the left, impinging the wall normally. The terms proximal and distal denote the near-shock and near-wall sides of the bubble, respectively (i.e., the right and left sides of the bubble in the present setup). The assumption of axisymmetry allows the reduction in the problem from three spatial dimensions to axisymmetric coordinates (cylindrical coordinates with azimuthal symmetry, i.e., there is no θ-dependence), thus greatly alleviating the computational expense; each computational cell consists of an annular region.

In practice, the shock is not necessarily aligned with the stone normal so that the interaction between the shockwave and the bubble is a fully three-dimensional process; the present assumption of axisymmetry constitutes the geometrical configuration for which the bubble collapse is most energetic because of symmetry. A full three-dimensional simulation is required to compute the more general problem, in which case an additional parameter, the angle between the shock and the wall normal, should be introduced.

B. Governing equations and numerical method

The main characteristics of single-bubble collapse in SWL are large nonspherical interface deformations and the generation and propagation of shockwaves. Since interactions between different types of waves and interfaces are the most important flow features, compressibility effects in gases and liquids are primordial. On the other hand, diffusive effects, surface tension, and mass transfer are not expected to play a significant role until the final stages of collapse, rebound, and subsequent growth. Based on these molecular effects, we consider compressible multicomponent flows,[33] which constitute a subset of multiphase flows where the different fluid components, characterized by their respective (constant) ratio of specific heats, are immiscible. The bubble contains only noncondensible gas, which is assumed to behave ideally. We note that the present study focuses on bubbles solely containing *air*. When considering *cavitation* bubbles, mass transfer must be introduced in order to monitor the phase change between vapor and water, though the resulting gas-vapor mixture within the bubble is still expected to behave ideally. The liquid is water and obeys an appropriate equation of state described below. The resulting inviscid and adiabatic flows are governed by the Euler equations, written here in cylindrical coordinates with azimuthal symmetry:

$$\begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix}_t + \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (E+p)u \end{pmatrix}_x + \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (E+p)v \end{pmatrix}_r = -\begin{pmatrix} \frac{\rho v}{r} \\ \frac{\rho uv}{r} \\ \frac{\rho v^2}{r} \\ \frac{(E+p)v}{r} \end{pmatrix}, \tag{2}$$

where ρ is the density, u is the axial velocity, v is the radial velocity, p is the pressure, and E is the total energy. The axial direction is represented by the x-coordinate, the radial direction is represented by the r-coordinate, and time is t; subscripts in Eq. (2) denote differentiation with respect to that variable. These equations are closed by specifying an appropriate equation of state in each fluid. We model water using the stiffened equation of state,[34]

$$p + \gamma B = (\gamma - 1)\left(E - \rho\frac{u^2}{2}\right), \tag{3}$$

where γ and B are empirical constants with values of 6.59 and 4049 atm,[35] respectively. For the air inside the bubble, we can take $B=0$ and $\gamma=1.4$ to obtain the perfect gas relation from Eq. (3). In what follows, the ambient density and sound speed of the water, ρ_L and c_L (along with the initial bubble radius, R_o), are used to nondimensionalize all quantities appearing in the model and results. When quoting dimensional results, we shall use $\rho_L=998$ kg/m^3 and $c_L=1647$ m/s. This value of c_L is approximately 10% larger than usual and is based on the derivation of the stiffened equation of state.[35] However, the impact of this discrepancy on the results is negligible since the shocks considered in the present work are weak (i.e., $M_s=1+\epsilon$, where $0 \ll \epsilon \ll 1$). Because the fluid components are assumed immiscible, interfaces are specified by a discontinuity in the fluid composition characterized by γ and B. Since mass transfer is neglected, interfaces between two fluid components are advected by the flow. The nominal location of the interface between liquid and gas is taken as the value of $\gamma_{\text{int}}=1.42$.

In order to accurately solve the governing equations, we have previously developed a numerical method with the following properties: *high-order accuracy* (good convergence in smooth regions and little dissipation at discontinuities), *conservation* (discrete conservation of mass, momentum, and energy), *shock capturing* (prevention of spurious oscillations at shockwaves), and *interface capturing* (prevention of spurious oscillations at interfaces). From a practical standpoint, it is also desirable that the scheme be computationally efficient and easy to implement. In order to achieve these goals, an existing quasiconservative interface-capturing formulation[36] was extended[27] by implementing a high-order accurate finite volume weighted essentially non-oscillatory (WENO)[37] reconstruction of the average primitive variables and modifying the Harten-Lax-van Leer solver with contact restoration (HLLC)[38] solver to solve appropriate advection equations. This method is implemented here in cylindrical coordinates with azimuthal symmetry on a stretched grid.[16]

C. Nondimensional parameters

Dimensional analysis can be used to minimize the number of (independent) parameters governing the physics of SIC. From the equations of motion and the problem description, the following *three* nondimensional groups can be formed: a nondimensional stand-off distance, H_o/R_o, a nondimensional pulse width, σ/R_o, and the ratio of the maximum (initial) shock pressure to the ambient pressure, p_s/p_o. In the present work, the pressure ratio is varied in the range $p_s/p_o=34-710$, which corresponds to lithotripters with peak pressures in the range 3.4–72 MPa and to shock Mach numbers in the range $M_s=1.003-1.050$. We note that the lower bound nominally corresponds to the (Rayleigh) collapse of a cavitation bubble since in that case $p_s/p_o \to p_o/p_v \approx 34$. Based on the numerical results, it was determined that the effect of the wall on the emitted shockwave is minor for initial stand-off distances greater than five initial radii. Thus, the range $H_o/R_o=1.05-4.5$ is considered. Values less than

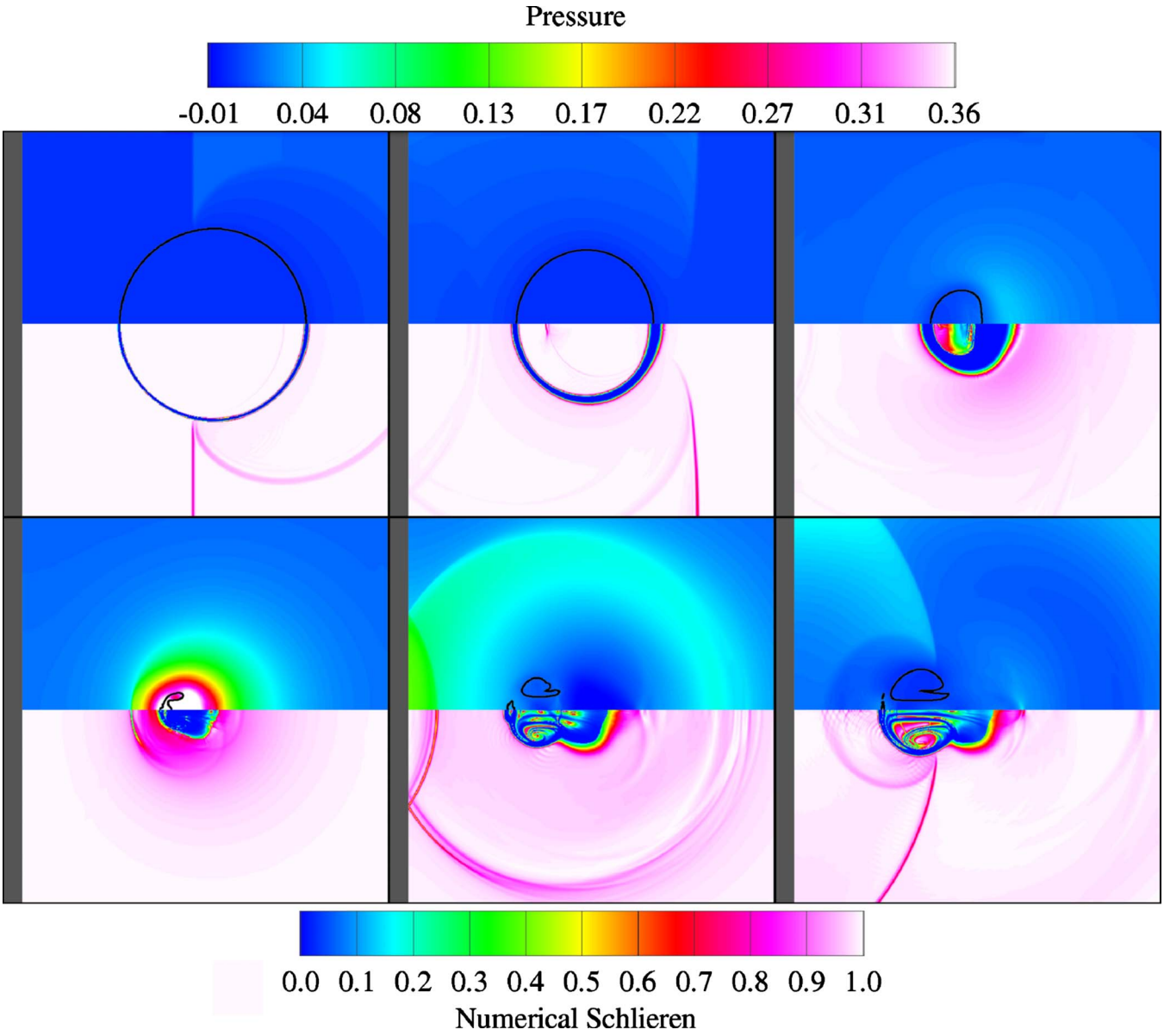


FIG. 3. (Color online) Pressure (top) and numerical schlieren (bottom) contours for SIC in SWL ($p_s/p_o=353$, $H_o/R_o=2.0$, and $\sigma/R_o=135$) at $tc_L/R_o=5.12, 9.77, 11.9, 12.8, 14.0, 15.3$. This figure is in color in the online version.

unity are considered in experiments;[10,17] however, because the bubble is not initially spherical in such a situation, the amount of gas at R_o is not the same so that it is not clear whether meaningful comparisons can be made. Finally, we have considered $67.5<\sigma/R_o<1350$. If the lithotripter pulse is fixed (with $\sigma=6.75$ mm, as quoted in Sec. II A), this range corresponds to bubbles with 5 μm$<R_o<100$ μm, which spans reasonable estimates for measured bubble nuclei in SWL.[39] Conversely, if the bubble size is held fixed (say, with $R_o=10$ μm), then the range corresponds to lithotripter pulses with different expansion durations, with 0.675 mm $<\sigma<13.5$ mm.

III. GENERAL OBSERVATIONS OF SHOCK-INDUCED COLLAPSE

In order to illustrate the general flow features, SIC for a baseline case with $p_s/p_o=353$, $H_o/R_o=2.0$, and $\sigma/R_o=135$ is considered. In physical parameters, this corresponds to a 35 MPa shock impacting a 50 μm radius bubble initially located 100 μm away from the solid surface. A qualitative description of the events is presented in Fig. 3. Slices across the computational domain through the centerline show numerical schlieren[40] (top) and pressure contours (bottom). The numerical schlieren contours have the advantage of allowing both shockwaves and interfaces to be visualized; however, they tend to smear discontinuities. The location of the interface is highlighted in black in the pressure plot and the dark area on the left of each frame denotes the wall. When the left-moving shock hits the bubble, an expansion wave is reflected because of the high impedance mismatch, while a weak shock is transmitted (frame 1). The shock then diffracts off the bubble and later intersects along the axis behind the bubble. Thereafter, the lithotripter pulse reflects off the wall and effectively doubles the local pressure; the initial transmitted shock focuses, not exactly in the center of the bubble but at a location that can be deduced from ray tracing (frame

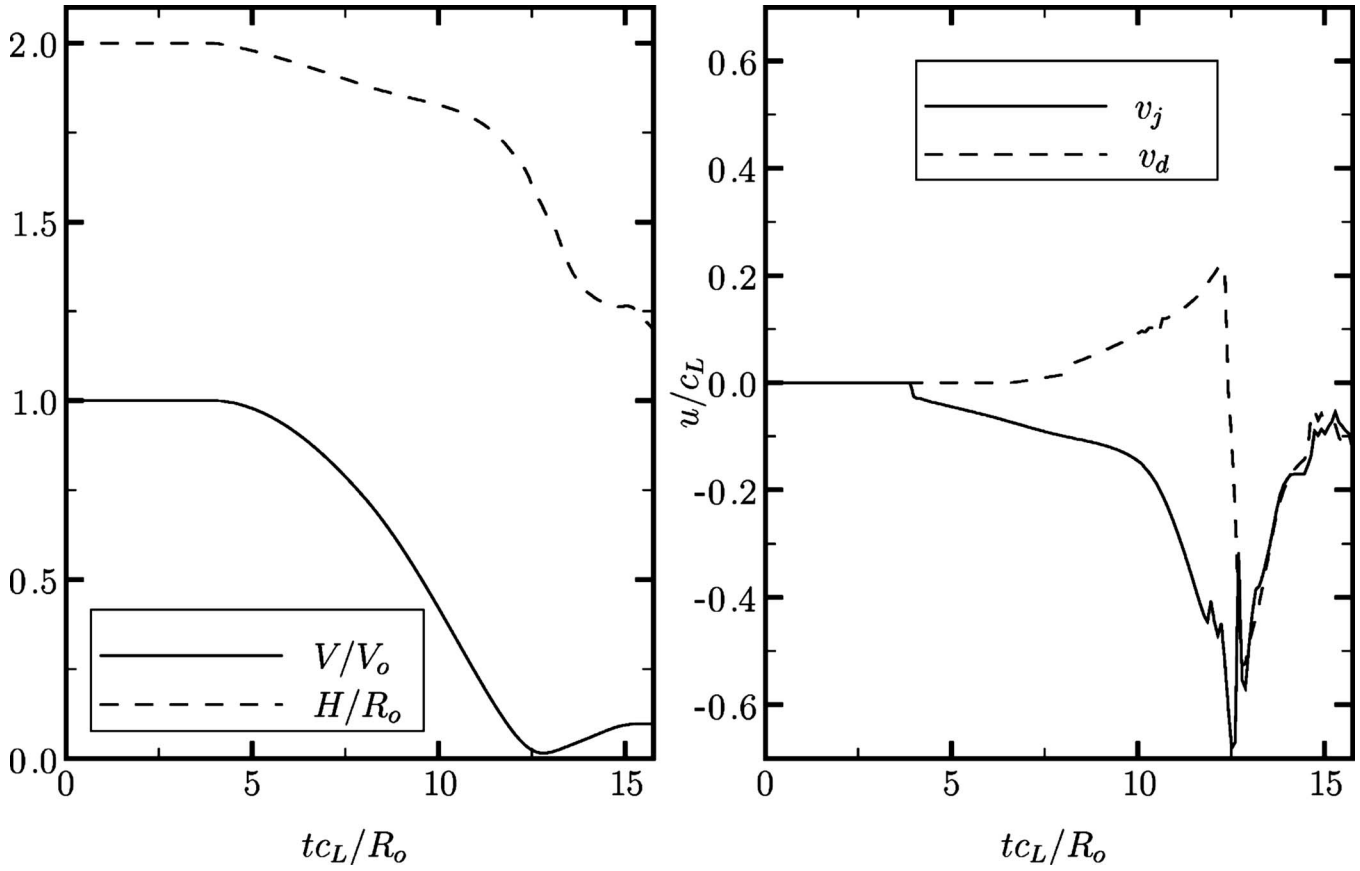


FIG. 4. History of the bubble volume and displacement (left) and velocity of the jet and of the distal side (right) for SIC in SWL (p_s/p_o=353, H_o/R_o=2.0, and σ/R_o=135).

2). We note that another shock is transmitted within the bubble when the incoming pulse interacts with the bubble again after reflection off the wall. The bubble proceeds to collapse nonspherically, while complex wave interactions take place within the bubble (frame 3). The distal side flattens, starts to involute, and eventually takes the form of a re-entrant jet. At collapse, the jet has penetrated the bubble and impacts the distal side; this generates a water-hammer shockwave, which propagates spherically outward (frame 4). The strength of the shock is higher in the direction of the jet. The bubble then takes the shape of a vortex ring and convects toward the wall, while the water-hammer shockwave reflects off the wall back onto the bubble (frame 5). As the bubble expands again, the shock interacts with it and reflects back onto the wall as an expansion wave (frame 6).

A. Bubble dynamics

To better understand the bubble dynamics, the history of the bubble volume, stand-off distance, jet velocity, and velocity of the distal side is plotted in Fig. 4. After the passage of the shock ($tc_L/R_o \approx 4$), the bubble begins to collapse. In the initial stages, the collapse is slow, as observed in the early migration of the bubble toward the wall and in the gradual increase in jet velocity. The external shock reaches the distal bubble side at $tc_L/R_o \approx 6$, and induces it to contract. In the latter stages, high interfacial velocities are achieved; the bubble collapses to a very small size and accelerates toward the wall. The jet eventually impacts the distal side, causing a large deceleration of the interface; this occurs slightly before the bubble reaches its minimum volume. The shockwaves generated by the impact of the jet onto the distal side and by the achievement of the minimum volume are difficult to distinguish.[20] After collapse, the bubble still migrates toward the wall.

Upon the impact of the jet onto the distal side of the bubble, a water-hammer pressure is generated. For the impact of a liquid jet onto a liquid surface, the water-hammer equation simplifies to

$$p_{wh} \approx \frac{\rho_L c_L |v_j - v_d|}{2}, \tag{4}$$

where v_j is the jet velocity ($v_j \approx 1120$ m/s) and v_d is the velocity of the distal side ($v_d \approx 364$ m/s). The local sound speed and density at the moment of impact are 1330 m/s and 1400 kg/m^3 so that the computed water-hammer pressure is 1.5 GPa. In the simulations, the local pressure at the jet is 1.8 GPa, giving reasonable agreement. This analysis illustrates that very high local velocities and pressures are achieved in the liquid during the process. Comparisons with experimental findings are provided for the jet velocity in Sec. IV C.

B. Wall pressure

The wall pressure is an important quantity indicative of the damage potential of bubble collapse. Figure 5 shows the history of the wall pressure at different locations along the

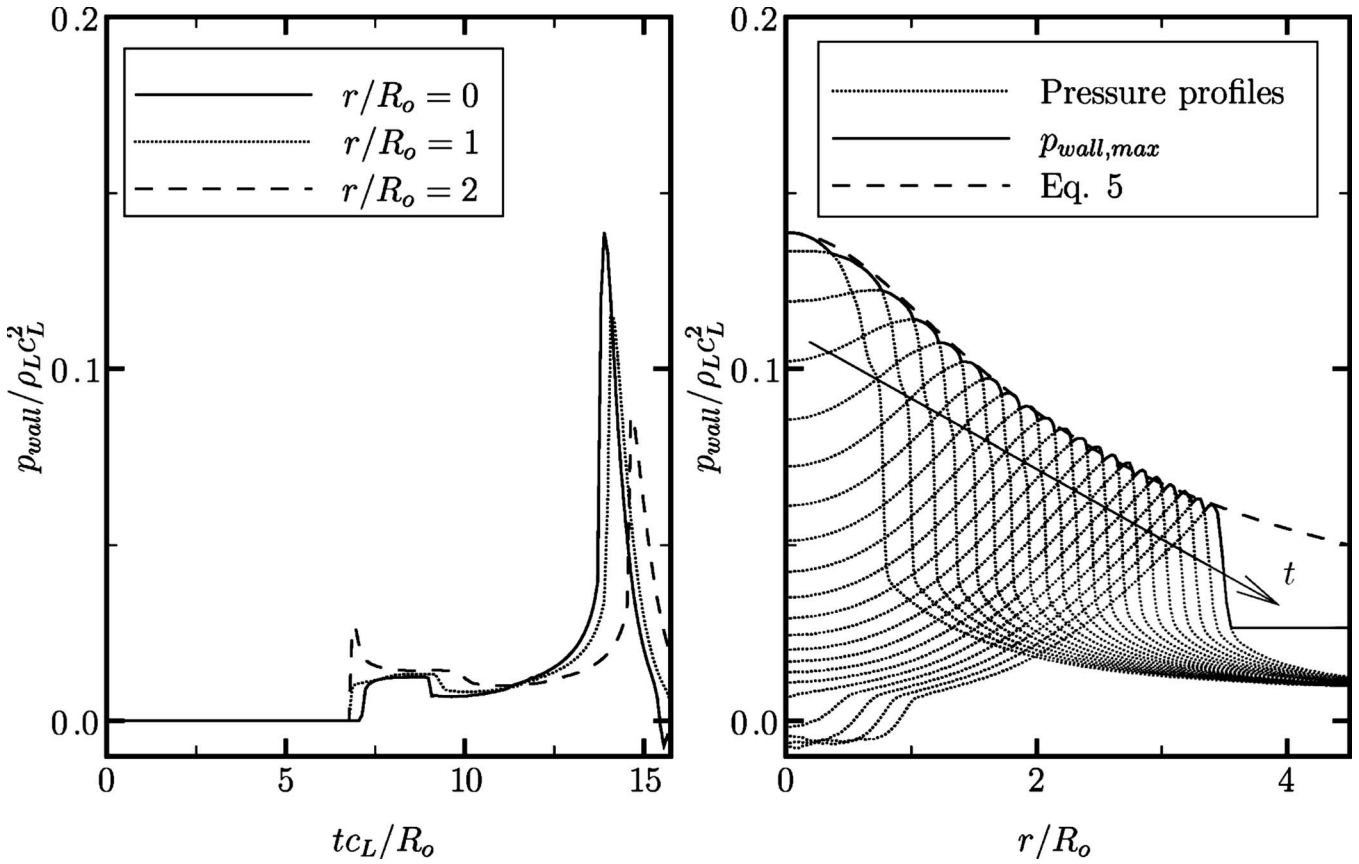


FIG. 5. History of the wall pressure (left) and pressure profiles along the wall (right) for SIC in SWL (p_s/p_o=353, H_o/R_o=2.0, and σ/R_o=135).

wall and pressure profiles along the wall at different times. First, the lithotripter pulse hits the wall at $tc_L/R_o \approx 7$. The pressure along the centerline (r/R_o=0) is slightly lower and delayed compared to other locations because, when the initial pulse impacts the bubble, a portion of the wave is reflected and the external shock has to diffract around the bubble. In other words, the bubble shields the wall by some amount that depends on H_o/R_o. The shock then reflects off the wall and impacts the bubble again. Because of the impedance mismatch, the amplitude inverts so that, upon the interaction with the resulting expansion wave, the wall pressure decreases at $tc_L/R_o \approx 9$. The large pressure rise then observed is caused by the water-hammer shock generated upon bubble collapse; hence, the maximum pressure due to bubble collapse (approximately 380 MPa) is much larger than that due to the pulse in the present case; this is further discussed in Sec. IV A. Later in the wall pressure history, negative pressures (tension) are achieved due to the reflection of the water-hammer shock onto the bubble and the subsequent inversion in the amplitude.

Because the water-hammer shock propagates spherically outward, the pressure is inversely proportional to the radial distance from the origin of the shock.[41,42] Thus, using basic geometry, the wall pressure is given by

$$p_{\text{wall}}(y) = \frac{a}{\sqrt{H_c^2 + y^2}} + b, \tag{5}$$

where y is the radial coordinate along the wall, H_c is the distance from the collapse location to the wall for a given H_o/R_o, and a and b are constants that can be determined if at least two pressure measurements are known. Equation (5) is the dashed curve in Fig. 5, which agrees very well with the computational results. This equation explains why the shock resulting from the collapse of a bubble far away (large H_c) looks essentially planar when it impacts the wall. Using Eq. (5), the radius of the area over which the wall pressure is larger than that of the pulse is given by

$$L_r = \sqrt{\left(\frac{a}{p_s - b}\right)^2 - H_c^2}. \tag{6}$$

In this particular case, $L_r/R_o \approx 10$, meaning that the area over which the pressure due to bubble collapse is larger than that of the lithotripter pulse is 100 times larger than the initial projected area of the bubble. Although bubbles undergoing SIC are initially small ($R_o \approx 10$ μm) and collapse to an even smaller size, the area over which they act is much larger ($R_{\text{area}} \approx 100$ μm). Thus, only a few nuclei are needed to generate sizable surfaces of high pressure along the stone.

As illustrated by the present flow visualizations, the large wall pressure observed in Fig. 5 is caused by the shock-wave emitted during bubble collapse, not by the impact of the jet onto the distal side. For the latter phenomenon to be important, the bubble must be located very close to the wall initially (i.e., $H_o/R_o < 1$), as is the case in prior experiments of Rayleigh collapse.[10,17]

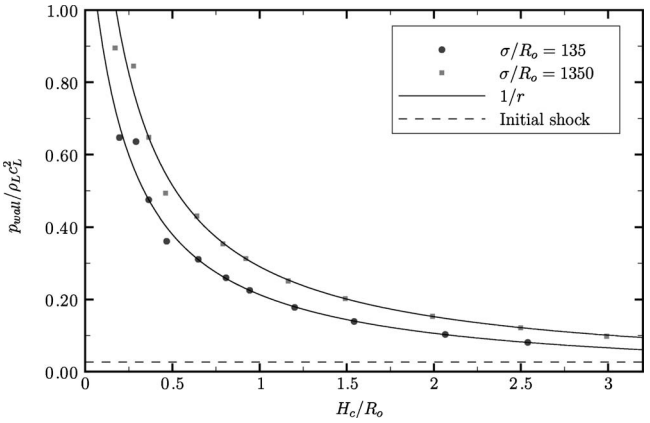


FIG. 6. Wall pressure along the centerline as a function of the initial stand-off distance for two different bubble sizes (p_s/p_o=353).

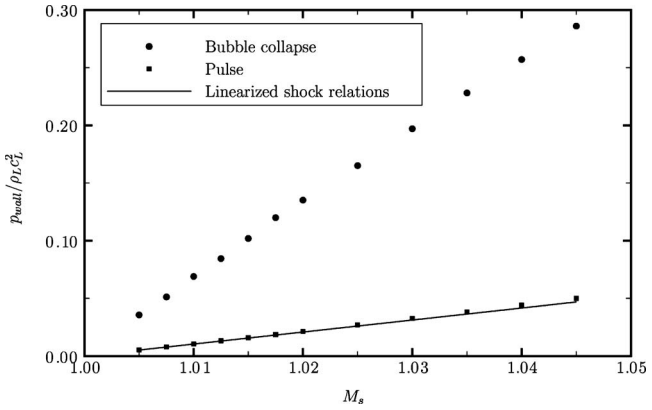


FIG. 7. Wall pressure due to the bubble collapse and to the pulse as a function of the amplitude of the shockwave for H_o/R_o=2.0 and σ/R_o =135.

IV. DAMAGE POTENTIAL OF SHOCK-INDUCED COLLAPSE

As a measure of the damage potential of SIC in SWL, the maximum wall pressure generated by bubble collapse is considered in the following parametric study. The dependence on the initial stand-off distance, pulse amplitude, and pulse width is studied.

A. Dependence on the initial stand-off distance

The distance at which bubbles are likely to cause damage is of great importance for practical purposes. The dependence of the maximum wall pressure along the centerline on the stand-off distance for p_s/p_o=353 and σ/R_o=135, 1350 is shown in Fig. 6 to illustrate the effect of small and large isolated bubbles in SWL (holding σ fixed). The pressure due to the pulse is also included (dashed line); it was compiled by running the simulation with no bubble. Because the bubble migrates toward the wall by some amount dependent on H_o/R_o, the stand-off distance at collapse is considered since this is when the shockwave is emitted. The same number of computational points is used across the bubble for each σ/R_o. Because the extent of the potential damage scales with the initial bubble radius, the pressure is averaged over the first ten cells in the case σ/R_o=1350 so that the area over which the pressure is recorded is the same in both cases. As expected, bubbles close to the wall generate a higher pressure (up to 2.4 GPa), which is inversely proportional to distance from the origin, as remarked in Sec. III B. The maximum pressure due to bubble collapse is thus much larger than that of the incoming pulse for the range of H_o/R_o considered in the present study; in the case of $\sigma/R_o \to \infty$, the collapse of a bubble within $H_o \approx 8R_o$ generates a wall pressure higher than that of the lithotripter pulse. Hence, bubbles within this distance show the potential of SIC for surface erosion.

In addition, the results show that the smaller bubble generates higher wall pressures, thus showing a higher potential for damage; however, the extent of the damage scales with the initial radius. This phenomenon is due to the fact that the high pressure of the compressive part of the pulse is exerted over a longer time for a bubble that is small compared to the pulse width. As a generalization, the results indicate that the collapse of a bubble that is *large* compared to the pulse width (e.g., cavitation bubble) is more gentle than that of a *small* bubble (e.g., gas nucleus). The dependence of the bubble dynamics on the pulse width is investigated in Sec. IV C. Experiments of SIC (Ref. 43) have measured wall pressures up to approximately 11 MPa for SIC, though lower amplitude shockwaves were used.

B. Dependence on the pulse amplitude

The dependence of the wall pressure on the shock amplitude and initial stand-off distance is useful when considering shock propagation through a bubble cloud near a solid surface. Figure 7 shows the dependence of the maximum wall pressure along the centerline on the Mach number of the pulse, M_s, for SIC with H_o/R_o=2.0 and σ/R_o=135. The Mach number is related to the pressure ratio across the shock.[16] The wall pressure due to bubble collapse increases linearly with increasing shock amplitude, with a slope steeper than that of the wall pressure due to the pulse. The measurements of the wall pressure due to the pulse match the linearized shock relations

$$\frac{p_{\text{wall,pulse}}}{\rho_L c_L^2} = 2\left(\frac{p_o}{\rho_L c_L^2} + \frac{4}{\gamma+1}\epsilon\right), \tag{7}$$

where the factor of 2 accounts for pressure doubling and $M_s = 1+\epsilon$, with $0 < \epsilon \ll 1$.

In bubble clouds, the damage due to the bubbles closest to the surface could be assessed by combining the present analysis with a model for shock propagation through bubbly mixtures.[44,45] As the shock propagates through the cloud, it becomes attenuated, such that the bubbles nearest to the wall only feel a fraction of the original shock amplitude. Thus, if the initial stand-off distance and the shock amplitude at that location in the cloud are known, the results of this section can be used to predict the potential damage due to the SIC of a single bubble, which can then be averaged over a given area to represent the damage due to multiple bubbles. It

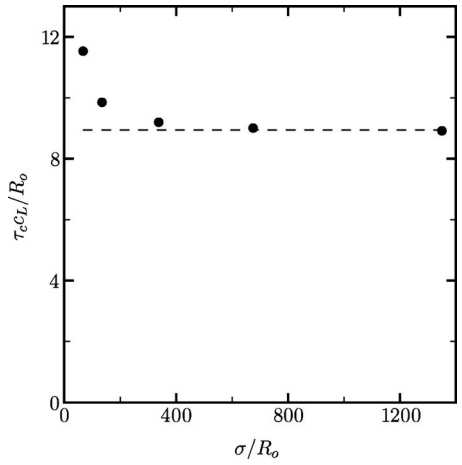


FIG. 8. Collapse time as a function of the pulse width for free-field SIC in SWL (p_s/p_o=353).

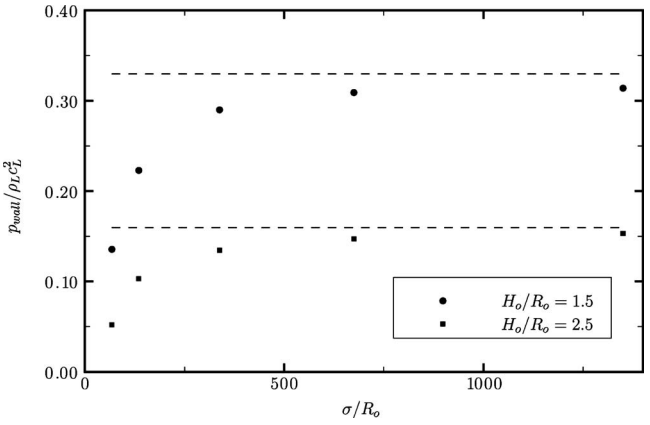


FIG. 10. Wall pressure along the centerline as a function of pulse width for different initial stand-off distances in SWL with p_s/p_o=353.

should be noted that the present analysis applied to gas bubbles, while clouds consist of vapor bubbles in most applications.

C. Dependence on the pulse width

The finite width of the lithotripter pulse has important consequences. Changing this quantity leads to two possible interpretations: a variable pulse width (which is a property of the lithotripter) and a variable initial bubble radius (which depends on the bubble population), as discussed in Sec. IV A. In order to understand the effect of the pulse width in SWL, Fig. 8 shows the collapse time, τ_c, and Fig. 9 depicts the maximum interfacial velocity (jet and distal side) and the water-hammer pressure for *free-field collapse* with p_s/p_o =353. The dashed line represents the limiting case of $\sigma/R_o \to \infty$ (i.e., a stepwise increase in pressure, followed by no expansion). For the parameters considered in the present study, the collapse time occurs well before the arrival of the negative tail of the pulse (at $tc_L/R_o \approx 23$). Thus, the dynamics of collapse are only affected by the compressive portion of the pulse. Yet, the results show that the dynamics are sensitive to the pulse width. As the pulse width is increased, the behavior of the bubble tends to that resulting from the interaction with a shockwave that has infinite width; the data for σ/R_o=675 are already close to the asymptotic value.

The behavior of the maximum wall pressure in SIC near a solid surface follows a similar trend, as seen in Fig. 10, though the asymptote is achieved for larger values of σ/R_o. This phenomenon may be understood by the fact that increasing the pulse width subjects the bubble to a high pressure for a longer time, as shown schematically in Fig. 11. An alternate view is to hold σ fixed so that the bubble radius is the variable, as discussed previously; similarly, a smaller bubble is exposed to the high pressure for a longer time.

In experiments of SIC of a gas bubble near an aluminum foil,[19] a pressure ratio of $p_s/p_o \approx 650$ is used. The pulse width in this case is $\sigma \approx 10$ mm and the range is σ/R_o =11.4–20.8, which corresponds to large bubbles. A trend similar to the present results is observed for the collapse time, which falls in the range $\tau_c c_L/R_o$=6.51–9.36, as a function of the pulse width (or initial bubble radius). The collapse time decreases with increasing pulse width, eventually asymptoting to the value represented by a stepwise increase in pressure. Jet velocities of approximately 600–700 m/s are achieved in this range of σ/R_o; the present simulations yield a jet velocity of 660 m/s for p_s/p_o=353 and σ/R_o=67.5 (the smallest value of σ/R_o considered here). Although the parameters are different, similar orders of magnitude and trends

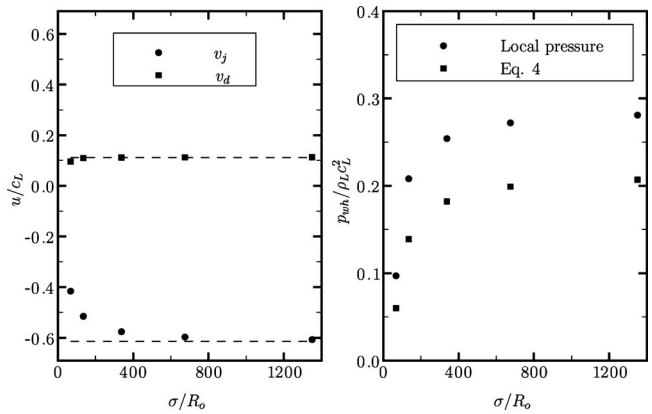


FIG. 9. Maximum velocity of the jet and of the distal side (left) and measured and computed water-hammer pressure (right) as a function of the pulse width for free-field SIC in SWL (p_s/p_o=353).

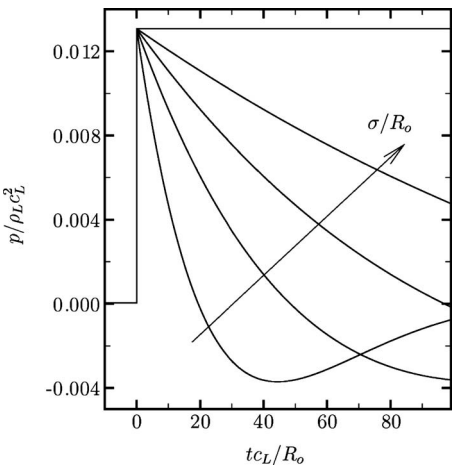


FIG. 11. Waveform for increasing σ/R_o.

are achieved in the simulations. In free-field experiments with lower pulse amplitude[18] ($p_s/p_o=100-200$), jet velocities up to 150 m/s are attained for $R_o=50$ μm, while jet velocities of 360–700 m/s are achieved in the simulations. In these experiments, however, the rise time is much slower than in the simulations, and measured values constitute lower bounds limited by the resolution of the laboratory equipment.

V. CONCLUSIONS

We numerically simulated the shock-induced collapse (SIC) of a single air bubble in shockwave lithotripsy (SWL) using a high-order accurate quasiconservative shock- and interface-capturing scheme. The maximum wall pressure due to bubble collapse is considered as an indication of potential damage.

Flow visualizations allow a detailed examination of the bubble dynamics. A re-entrant jet forms during collapse and impacts the distal side, thereby generating a water-hammer shock. This wave propagates spherically outward and hits the neighboring wall; in the range of stand-off distances considered in the present work, this phenomenon (and not the impact of the jet onto the wall) leads to a high wall pressure. After collapse, the bubble takes the form of a vortex ring and convects toward the wall as it rebounds.

The findings show that wall pressures on the order of 1 GPa may be achieved in SIC for bubbles located close to the wall initially. This maximum pressure decreases with initial stand-off distance and pulse width (i.e., when the bubble is large compared to the pulse width) and increases with pulse amplitude. For the stand-off distances considered in the present work, the wall pressure due to bubble collapse is larger than that due to the incoming shockwave; this applies to bubbles within approximately eight initial radii from the wall. The region along the wall over which this holds may extend to ten initial radii, thus showing that, even though the size of a bubble is small, the collapse of just a few such bubbles would lead to a significant area of high pressure on the stone surface. Though shock-induced bubble collapse had not yet been systematically investigated in the context of SWL, the present results indicate that this phenomenon has a high potential for damage.

Because the development of numerical methods capable of simulating shockwaves and interfaces is still in their infancy, only a limited range of the physics is included. An implementation of surface tension, dissipative effects, and mass transfer especially would allow a better representation of the phenomenon. Then, the Rayleigh (and shock-induced) collapse of cavitation (i.e., vapor) bubbles could be simulated. An extension of the method to three dimensions would further allow the computation of more complex geometries. Finally, by coupling the present method to a solid mechanics simulation, the wave propagation within the solid and the actual damage caused by bubble collapse could be predicted. This application is not only relevant to SWL but to the cavitation community as a whole.

ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful discussions with members of the PPG group (PI Andrew Evan), in particular, Dr. Michael Bailey, Dr. Robin Cleveland, Dr. Wayne Kreider, Dr. Yura Pishchalnikov, and Dr. Oleg Sapozhnikov. This work was supported by NIH Grant No. PO1 DK043881 and by ONR Grant No. N00014-06-1-0730.

[1]NIH, "Kidney stones in adults," on the WWW, URL: http://kidney.niddk.nih.gov (Last viewed April 21, 2008).
[2]R. F. Paterson, D. A. Lifshitz, J. E. Lingeman, A. P. Evan, B. A. Connors, N. S. Fineberg, J. C. Williams, and J. A. McAteer, "Stone fragmentation during shock wave lithotripsy is improved by slowing the shock wave rate: Studies with a new animal model," J. Urol. (Baltimore) **168**, 2211–2215 (2002).
[3]W. Eisenmenger, "The mechanisms of stone fragmentation in ESWL," Ultrasound Med. Biol. **27**, 683–693 (2001).
[4]R. O. Cleveland and O. A. Sapozhnikov, "Modeling elastic wave propagation in kidney stones with application to shock wave lithotripsy," J. Acoust. Soc. Am. **118**, 2667–2676 (2005).
[5]M. Lockandwalla and B. Sturtevant, "Fracture mechanics model of stone comminution in ESWL and implications for tissue damage," Phys. Med. Biol. **45**, 1923–1940 (2000).
[6]S. Zhu, F. G. Preminger, and P. Zhong, "The role of stress waves and cavitation in stone comminution in shock wave lithotripsy," Ultrasound Med. Biol. **28**, 661–671 (2002).
[7]A. J. Coleman, J. E. Saunders, L. A. Crum, and M. Dyson, "Acoustic cavitation generated by an extracorporeal shockwave lithotripter," Ultrasound Med. Biol. **13**, 69–76 (1987).
[8]L. A. Crum, "Cavitation microjets as a contributory mechanism for renal calculi disintegration in ESWL," J. Urol. (Baltimore) **140**, 1587–1590 (1988).
[9]A. Philipp and W. Lauterborn, "Cavitation erosion by single laser-produced bubbles," J. Fluid Mech. **361**, 75–116 (1998).
[10]Y. Tomita and A. Shima, "Mechanisms of impulsive pressure generation and damage pit formation by bubble collapse," J. Fluid Mech. **169**, 535–564 (1986).
[11]M. S. Plesset and R. B. Chapman, "Collapse of an initially spherical vapour cavity in the neighbourhood of a solid boundary," J. Fluid Mech. **47**, 283–290 (1971).
[12]Y. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, J. C. Williams, R. O. Cleveland, T. Colonius, L. A. Crum, A. P. Evan, and J. A. McAteer, "Cavitation bubble cluster activity in the breakage of kidney stones by lithotripter shockwaves," J. Endourol. **17**, 435–446 (2003).
[13]M. Tanguay, "Computation of bubbly cavitating flow in shock wave lithotripsy," Ph.D. thesis, California Institute of Technology, Pasadena, CA (2004).
[14]L. Rayleigh, "On the pressure developed in a liquid during the collapse of a spherical cavity," Philos. Mag. **34**, 94–98 (1917).
[15]T. J. Matula, P. R. Hilmo, M. R. Bailey, and L. A. Crum, "*In vitro* sonoluminescence and sonochemistry studies with an electrohydraulic shockwave lithotripter," Ultrasound Med. Biol. **28**, 1199–1207 (2002).
[16]E. Johnsen, "Numerical simulation of non-spherical bubble collapse," Ph.D. thesis, California Institute of Technology, Pasadena, CA (2007).
[17]A. Vogel, W. Lauterborn, and R. Timm, "Optical and acoustic investigations of the dynamics of laser-produced cavitation bubbles near a solid boundary," J. Fluid Mech. **206**, 299–338 (1989).
[18]C. D. Ohl and R. Ikink, "Shock-wave-induced jetting of micron-size bubbles," Phys. Rev. Lett. **90**, 214502 (2003).
[19]A. Philipp, M. Delius, C. Scheffczyk, A. Vogel, and W. Lauterborn, "Interaction of lithotripter-generated shock waves with air bubbles," J. Acoust. Soc. Am. **93**, 2496–2509 (1993).
[20]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).
[21]C. H. Chang and M. S. Liou, "A robust and accurate approach to computing compressible multiphase flow: Stratified flow model and AUSM$^+$-up scheme," J. Comput. Phys. **225**, 840–873 (2007).
[22]X. Y. Hu, B. C. Khoo, N. A. Adams, and F. L. Huang, "A conservative interface method for compressible flows," J. Comput. Phys. **219**, 553–578 (2006).
[23]R. R. Nourgaliev, T. N. Dinh, and T. G. Theofanous, "Adaptive

characteristics-based matching for compressible multifluid dynamics," J. Comput. Phys. **213**, 500–529 (2006).
[24]X. Y. Hu and N. A. Adams, "Shock-induced collapse of bubbles in liquid," in Proceeding of the 26th International Symposium on Shock Waves, Goettingen, Germany (2007).
[25]A. R. Jamaluddin, "Free-Lagrange simulations of shock-bubble interaction in extracorporeal shock wave lithotripsy," Ph.D. thesis, University of Southampton, Southampton, UK (2005).
[26]S. Nagrath, K. Jansen, R. T. Lahey, Jr., and I. Akhatov, "Hydrodynamics simulation of air bubble implosion using a level set approach," J. Comput. Phys. **215**, 98–132 (2006).
[27]E. Johnsen and T. Colonius, "Implementation of WENO schemes for compressible multicomponent flow problems," J. Comput. Phys. **219**, 715–732 (2006).
[28]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[29]A. J. Coleman and J. E. Saunders, "A survey of the acoustic output of commercial extracorporeal shock wave lithotripters," Ultrasound Med. Biol. **15**, 213–227 (1989).
[30]P. A. Thompson, *Compressible-Fluid Dynamics* (McGraw-Hill, New York, 1972).
[31]R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, M. Lokhandwalla, J. A. McAteer, and B. Sturtevant, "Design and characterization of a research electrohydraulic lithotripter patterned after the Dornier HM3," Rev. Sci. Instrum. **71**, 2514–2525 (2000).
[32]P. Zhong, C. J. Chuong, and G. M. Preminger, "Propagation of shock waves in elastic solids caused by cavitation microjet impact. II: Application in extracorporeal shock wave lithotripsy," J. Acoust. Soc. Am. **94**, 29–36 (1993).
[33]K. M. Shyue, "An efficient shock-capturing algorithm for compressible multicomponent problems," J. Comput. Phys. **142**, 208–242 (1998).
[34]F. Harlow and A. Amsden, "Fluid dynamics," LANL Monograph Technical Report No. LA-4700, Los Alamos National Laboratory.
[35]J. P. Cocchi, R. Saurel, and J. C. Loraud, "Treatment of interface problems with Godunov-type schemes," Shock Waves **5**, 347–357 (1996).
[36]R. Abgrall, "How to prevent pressure oscillations in multicomponent flow calculations: A quasi conservative approach," J. Comput. Phys. **125**, 150–160 (1996).
[37]J. S. Jiang and C. W. Shu, "Efficient implementation of WENO schemes," J. Comput. Phys. **126**, 202–228 (1996).
[38]E. F. Toro, M. Spruce, and M. Spears, "Restoration of the contact surface in the HLL-Riemann solver," Shock Waves **4**, 25–34 (1996).
[39]A. J. Coleman, M. J. Choi, J. E. Saunders, and T. G. Leighton, "Acoustic emission and sonoluminescence due to cavitation at the beam focus of an electrohydraulic shock wave lithotripter," Ultrasound Med. Biol. **18**, 267–281 (1992).
[40]J. J. Quirk and S. Karni, "On the dynamics of a shock-bubble interaction," J. Fluid Mech. **318**, 129–163 (1996).
[41]S. Fujikawa and T. Akamatsu, "Effects of the non-equilibrium condensation of vapour on the pressure wave produced by the collapse of a bubble in a liquid," J. Fluid Mech. **97**, 481–512 (1980).
[42]R. Hickling and M. S. Plesset, "Collapse and rebound of a spherical bubble in water," Phys. Fluids **7**, 7–14 (1964).
[43]A. Shima, Y. Tomita, and K. Takahashi, "The collapse of a gas bubble near a solid wall by a shock wave and the induced impulsive pressure," Proc. Inst. Mech. Eng. **198C**, 81–86 (1984).
[44]M. Kameda and Y. Matsumoto, "Shock waves in a liquid containing small gas bubbles," Phys. Fluids **8**, 322–335 (1996).
[45]M. Kameda, N. Shimaura, F. Higashino, and Y. Matsumoto, "Shock waves in a uniform bubbly flow," Phys. Fluids **10**, 2661–2668 (1998).

Angular influence on the scattering of fundamental shear horizontal guided waves by a through-thickness crack in an isotropic plate

P. Rajagopal and M. J. S. Lowe[a)]
Department of Mechanical Engineering, Imperial College, London SW7 2AZ, United Kingdom

(Received 26 November 2007; revised 14 July 2008; accepted 15 July 2008)

The angular influence on the scattering of cylindrical-crested waves of the fundamental shear horizontal (*SH*0) guided mode by through-thickness cracks in an isotropic plate is studied in the context of array imaging using ultrasonic guided waves. Finite element simulations are used to obtain trends which are subject to analytical study and experimental confirmation. The influence of the incidence angle on reflection behavior is first studied in terms of two complementary cases, that of normal incidence and that of specular reflection at various oblique incidence angles. The normal incidence study suggests that for a given incidence angle, the peak reflection is concentrated around the specular direction, while the oblique incidence studies show that maximum specular reflection occurs in the case of normal incidence. The variation of diffraction with both the angle of incidence and that of monitoring is then taken up and this shows that when the first diffraction from the crack edges can be separated, its angular dependence can be obtained from literature on similar bulk elastic wave scattering problems. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2968697]

PACS number(s): 43.35.Zc, 43.20.Mv [YHB] Pages: 2021–2030

I. INTRODUCTION

Transducer array based imaging methods to improve the resolution of ultrasonic guided wave inspection are a subject of continuing investigation.[1,2] Such tools will be valuable in the inspection of inaccessible regions of structures, and a reduction in range as compared to screening applications[3–5] is acceptable in order to achieve this. In this context, it is important to understand the interaction of guided waves with defects, some key questions being how the transducer, defect location, and defect extent impact the accuracy of such methods. To this end we are studying the general shorter range scattering of the fundamental shear horizontal guided wave (*SH*0) mode by ideal but finite cracks in thin isotropic plates. The attractiveness of the *SH*0 mode, requirements of cylindrical wave fronts, and studies focusing on understanding the spatial behavior of scattering from through-thickness cracks were presented in previous work by the authors.[6] A low-frequency point excitation of in-plane force was used to simulate the expected behavior of individual transducer elements. As shown in Fig. 5(a) of Ref. 6, this leads[5,7] to the *S*0 and the *SH*0 modes being generated primarily parallel and perpendicular to the excitation direction, respectively. The scattering behavior was studied in terms of the influence of the crack length on the specular reflection and diffraction fields.

This earlier work showed that at low frequencies both reflection and diffraction fields are important. For small cracks, the specular reflection is enhanced by contributions from the diffracted field: This may be seen as focusing of energy by the crack in the backscattered direction. A simple diffraction model assuming uniform shear sources on the insonified face is found to adequately represent cracks up to moderate lengths, and thus provides an easy means of estimating the far field of the waves. Further, the diffracted field itself consists of multiple components, the direct diffraction from the crack tips (primary diffraction) and the radiation from Rayleigh-like[8–10] waves traveling along the crack faces (secondary diffraction). Figure 5(b) of Ref. 6 illustrates these concepts.

Here we extend this line of investigation and study the angular profile of scattering when the *SH*0 mode is incident at through cracks. Essentially, we wish to study how the scattering at different angles of observation varies due to different angles of incidence and crack lengths. The findings from the previous studies[6] helped identify the minimum number of independent parameters to be considered. Both the source and the receiver are thus placed in the far field of the largest cracks to be studied in a given model, in order to avoid spatial fluctuations in the scattered field. The crack length is stated in terms of the *SH*0 wavelength at the center frequency of the incident tone burst. The largest cracks studied were five wavelengths long so a distance of 15 wavelengths was sufficient. For the reflection studies, the parameters are further reduced by considering two complementary configurations involving the incidence and the observation angles, which together help one to understand the larger picture. For the diffraction studies, only cracks that are long enough to allow the separation of primary and secondary diffraction from each other are considered. As in Ref. 6, the studies are carried out using finite element (FE) analysis and general results are validated with experiments, while the dif-

[a)]Electronic mail: m.lowe@imperial.ac.uk

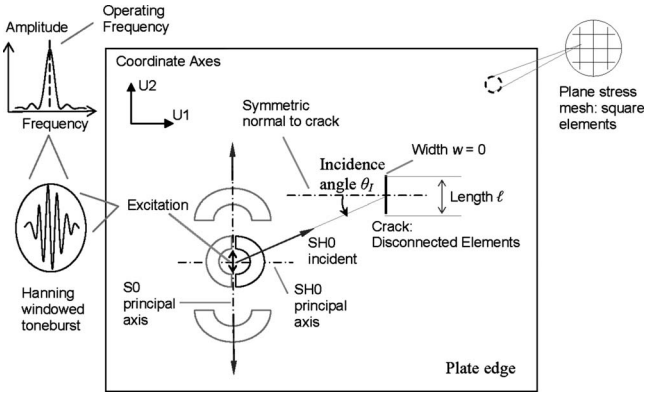


FIG. 1. Details of the configuration studied. The *SH*0 mode is generated such that a line from the wave source bisects the crack face at a required angle.

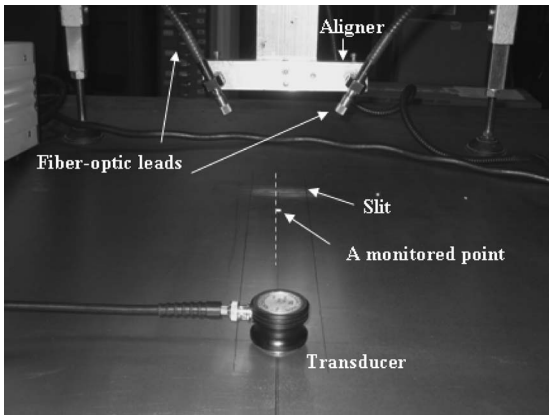


FIG. 2. Close-up photograph of the experimental setup: The transducer, the slit, the aligner, and the laser optical fiber leads can be seen.

fraction modeling is verified by comparison with results from similar problems in bulk elastic wave scattering literature.

In what follows, we first briefly describe the procedure for FE simulations and experiments. Studies examining the angular behavior of the reflection and diffraction are then presented and results are discussed in light of the physics of scattering. Finally, these insights are used to consider the implications for imaging applications.

II. METHODS

The configuration studied and the general procedure for FE simulations and experiments are broadly the same as that in Ref. 6 where they have been described in detail. For a self-contained presentation within this paper, we quickly summarize the key features.

A. Studied configuration and general procedure for FE simulations

Figure 1 illustrates the configuration studied and the procedure for FE simulations. A finite length planar through-thickness crack of vanishing width is located in a thin isotropic plate. The length of the crack is along the *U*2 direction. A point excitation[7] of in-plane force vibrating along the *U*2 direction is applied uniformly through the plate thickness and along a line which bisects the crack face at a given angle: We call this the incidence angle. This generates incident circular-crested waves of the *SH*0 mode traveling principally in the *U*1 direction and those of the *S*0 mode, principally in the *U*2 direction. Away from their principal directions both modes exhibit a cosine angular decay of the displacement amplitude. At the low frequency thickness of 0.1 MHz mm chosen for the study, only these two in-plane modes can be present even after scattering. The *S*0 mode travels much faster than the *SH*0 mode and this fact is used to help time gate its effects and study the scattering due to the latter in isolation. The *A*0 mode does not occur in this problem since it is not generated at the source and the through-thickness symmetry of the crack means that there is no mode conversion into it by the scattering. We then seek to study the angular profile of the *SH*0 mode reflected back from the crack face and diffracted around the crack tips (or edges). As shown in Ref. 6, this problem is equivalent to that of the scattering phenomena in the sagittal plane when bulk shear verticle (*SV*) waves interact with striplike internal cracks.

The problem was modeled by FE simulations in the two-dimensional domain with assumption of plane stress, implemented using the commercial package ABAQUS.[11] The plate thickness is assumed to be 1 mm, so the frequency thickness at an input frequency of say f_0 MHz would simply be f_0 MHz mm. As demonstrated in Ref. 6, for through-thickness cracks, the results obtained at f_0 would be valid at any other frequency f_1 provided the dimensions are appropriately scaled by a factor of f_1/f_0. The mesh consisted of perfectly square elements, allowing for around 16 elements per *SH*0 wavelength, respecting the spatial discretization limit required for accurate modeling of wave propagation.[12] Cracks of vanishing width were created by disconnecting nodes along adjacent elements. Material properties of aluminum were used for the reflection studies, while those of steel were chosen for the study of diffraction so as to facilitate comparison with the literature consulted, although there is only a small difference in the results obtained using these two properties. Absorbing layers with increasing damping (see Castaings *et al.*,[13] Castaings and Bacon,[14] and Drozdz *et al.*[15] for details on the use of absorbing boundaries in FE analysis) were used around the edges of the plate in order to isolate the scattered *SH*0 waves.

The excitation consisted of a five cycle Hanning windowed tone burst centered at 0.1 MHz applied as a force vibrating in the *U*2 direction at a single node. This generated the *S*0 and *SH*0 modes, with their principal directions parallel and perpendicular, respectively, to the applied force. Explicit time integration with a constant time step respecting the stability limit simulated the propagation of the modes. The default condition on the faces of the disconnected elements representing the crack is that of zero stress. Thus, the scattering from an ideal open crack was simulated.

B. General procedure for experiments

Figure 2 shows a close-up photograph of the experimental setup. A pointlike excitation was achieved by means of a

plane wideband piezoelectric shear transducer (Panametrics V301, 0.5 MHz center frequency) coupled to the plate through a thin (0.2 mm thickness) and small (3 mm diameter, roughly 1/3 that of typical wavelength) brass disk. The source signal consisting of a five-cycle Hanning windowed tone burst was generated by a Wavemaker (Macro Design Ltd., UK) instrument and centered at the required frequency.

The detection was achieved using a laser interferometer with dual differential fiber optic lines (Polytec OFV 2700) aligned at an angle of 30° to the surface of the plate, so that the difference between their signals gives the in-plane surface displacement. The aligner holding the two optical fiber leads is in turn attached to a translatable rotary arm. The interferometer reads in-plane displacements perpendicular to the direction in which the arm points, and thus the *SH*0 mode propagating along that direction is picked up when the rotational axis of the arm is vertically above the source point. Small patches (3 mm^2) of a thin reflective tape (about 0.1 mm thickness) were attached to the surface of the plate to enhance the optical backscatter from the laser beams at each of the required monitoring positions. Since these patches have low acoustic impedance and are much thinner in comparison to the wavelengths used as well as the plate thickness, they do not modify the wave propagation properties of the plate in any significant measure. The quality of each measurement was improved by applying a bandpass filter and taking an average over 500 acquisitions.

Measurements were carried out on 0.5-mm-wide slits of lengths 8 and 65 mm cut in 1250 mm^2 standard aluminum (commercial purity, AA1050) plates of 1 mm thickness. The center frequency of excitation was varied in order to obtain different values of slit length in terms of the wavelength. The excitation and detection positions were also varied at each center frequency, so that they were located at a certain constant number of wavelengths away from the slit.

III. SPECIFIC STUDIES AND RESULTS

A. Angular influence on reflection

We examine the angular influence on *SH*0 reflection through two complementary sets of studies. First, we consider the case of normal incidence and study the angular profile of the total reflected field, both specular and diffuse, as the crack length increases. (In this context, by "specular" we mean the direction of mirrorlike reflection as given by Snell's law and the term "diffuse" refers to all the directions in which the waves can be reflected in the backscatter direction). We then study the behavior of just the specular reflection at various other angles of incidence, again with increasing crack length. By an intuitive superposition of the findings from these two sets of studies, we will be able to picture the overall angular behavior of the reflected field. For these reflection studies, angles are defined in a counterclockwise sense from the crack face symmetric normal *to a ray pointing to or from its center* along a required direction. For simplicity, we refer to large angles, say $\alpha=360°-\theta$ defined in this way as $-\theta$ instead, measuring clockwise from the symmetric normal. These details are shown in Fig. 3.

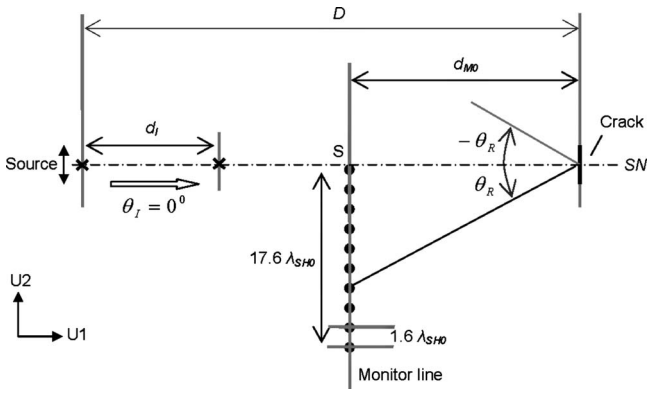


FIG. 3. Geometry for the study of reflection at angles θ_R for an incidence angle $\theta_1=0°$.

The reflection behavior was studied in terms of a frequency domain ratio of the resultant displacement of the reflected signal to that of the signal incident at the center of the crack face. The signals were compensated for beam spreading such that the ratio from a straight edge would be unity. Thus,

$$\text{Reflection ratio} = \frac{|R(\omega)|\sqrt{D+d_M}}{|I(\omega)|\sqrt{d_I}}, \tag{1}$$

where $R(\omega)$ and $I(\omega)$ are the frequency spectra of the reflected and incoming signals, respectively, D is the distance between the source and the center of the crack, d_M is the distance from the crack center to the point where the reflection is monitored, and d_I is the distance from the source to the point where the incident signal is monitored. (See Fig. 3 for an example illustration).

Here it must be mentioned that the compensation for beam spreading was chosen to be based on a straight edge rather than, say, a point source, because the former is commonly realized in practice. Due to this choice, as shown in Ref. 6, although the reflection ratio will ultimately reach unity if the specular reflection for symmetric normal incidence is monitored at a single point while the crack length is increased, it will not remain constant if we vary the point of measurement while maintaining a constant finite crack length. Under the latter condition, in the far field, the ratio decays continuously with distance of measurement. However, since our interest here is in understanding the physical principles rather than obtaining a rigorous quantitative parameter that would hold under different conditions, the convenient definition in Eq. (1) is sufficient.

1. Study of diffuse reflection due to normal incidence

Figure 3 illustrates the geometry used for the study of reflection at various diffuse angles θ_R for incidence angles $\theta_I=0°$. In the FE models, the excitation was located at a distance $D=32\ \lambda_{SH0}$ (*SH*0 wavelengths at the center frequency of the tone burst) away from the crack and such that the principal axis of the generated *SH*0 waves was aligned with SN, its symmetric normal. As this is a normal incidence study, the FE time snapshots from our previous work in Fig. 5(a) of Ref. 6, showing mode generation and interaction with

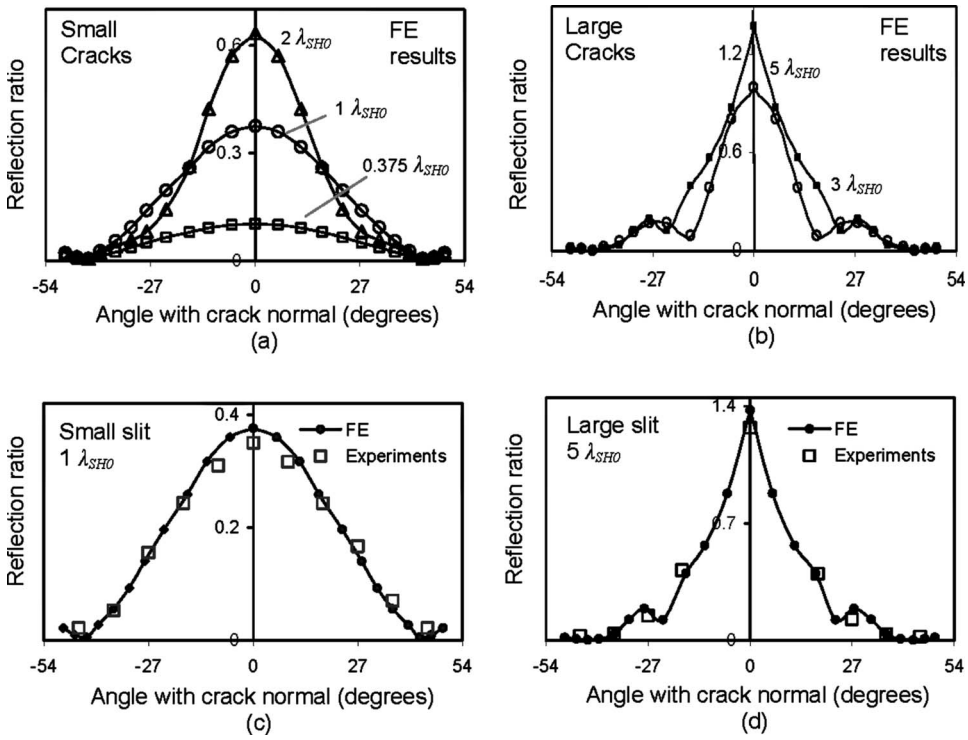


FIG. 4. (a), (b) Plots of reflection ratio from FE simulations with θ_R for $\theta_1=0°$, for small (0.375, 1, and 2 λ_{SH0}) and large (3 and 5 λ_{SH0}) cracks, respectively. (c), (d) Comparison of FE results with experiments for two slit length cases of 1 and 5 λ_{SH0}.

the crack, are illustrative of this present case as well. The $U2$ displacement of the incident waves was monitored along their principal axis at a distance d_I from the source: This is the only displacement component here, so it directly gives the total magnitude I. Reflected $SH0$ waves were monitored along one-half of a vertical line at a distance $d_{M0}=15\ \lambda_{SH0}$ from the crack and bisected by its symmetric normal at a point we call S. Starting from a point along this line at a distance of $17.6\ \lambda_{SH0}$ from S, nodal displacements R_{U2} in the $U2$ direction were monitored at intervals of $1.6\ \lambda_{SH0}$ all the way down to S. This yielded monitoring angles in the range $\theta_R=0°-50°$ in intervals of 3°–6°. The resultant displacement R is related to R_{U2} as $R\cos(\theta_R)=R_{U2}$, so for each monitored point the resultant displacement values were obtained from the $U2$ values as $R=R_{U2}/\cos(\theta_R)$. The values in the range from −50° down to 0° were assumed to be identical to these due to the symmetry in the configuration. The reflection ratio was then calculated using Eq. (1) with $d_M=d_{M0}/\cos(\theta_R)$.

Experiments were carried out on slits of length $1\ \lambda_{SH0}$ and $5\ \lambda_{SH0}$ achieved using the required center-frequency values on slits of lengths 8 and 65 mm, respectively. Measurements were taken using the same method as for FE simulations, at the appropriate excitation and monitoring distances in terms of λ_{SH0} at the operating center frequency. The incident $SH0$ waves were monitored by positioning the rotary arm along the symmetric normal to the slit. Reflection values were measured by allowing the arm to move along a straight line and monitoring at regular intervals yielding angles spanning the range of θ_R from −45° to 45°. The reflection ratio was then calculated using the same procedure as that for the FE predictions.

Figures 4(a) and 4(b) show the plot of reflection ratio with θ_R obtained from FE simulations for small and long cracks, respectively, and Figs. 7(c) and 7(d) present comparisons with experiments for two slit length cases. There is a very good agreement between the FE and experimental results.

The first observation is that as the crack length increases, most of the reflected energy is seen to be concentrated in a narrow region around the backscatter direction. An explanation for such behavior follows from arguments we presented in Ref. 6, that in the configuration studied, the insonified face of cracks can be taken to behave like a simple ideal shear transducer focusing energy back. It is known that the field from a circular piston transducer is increasingly focused with increasing piston radius (see, for instance, Ref. 16, pp. 347–348). We do discern such a trend from Fig. 4: As the crack length increases, the reflection tends to become increasingly concentrated around the direction of the crack's symmetric normal. We will discuss this idea in some more detail in Sec. IV A.

2. Study of specular reflection at various incidence angles

Figure 5 illustrates the geometrical details for the study of specular reflection at angles $\theta_R=-\theta_I$ for different incidence angles θ_I. In the FE models, excitation in the $U2$ direction was applied at a node along a vertical line at a fixed distance $x_I=19\ \lambda_{SH0}$ from the crack face and skewed from its symmetric normal by a certain distance y_I. For each angle of incidence θ_I, $y_I\approx x_I\tan(\theta_I)$ and angles in the range 0°–55° were considered. Figure 6 shows snapshots of the contour of

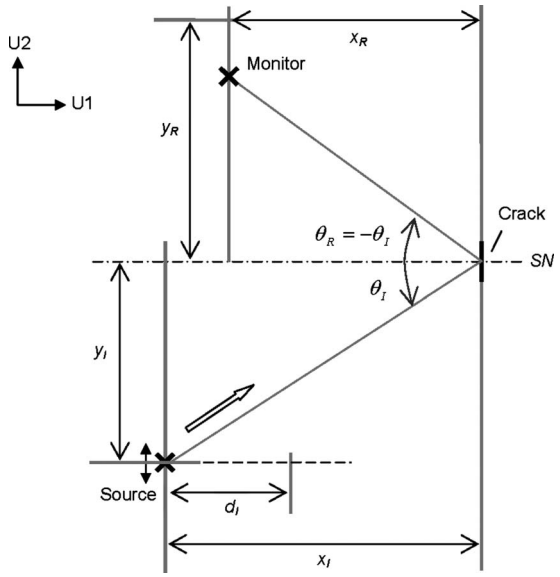


FIG. 5. Geometrical details for the study of specular reflection at oblique incidence angles.

magnitude of resultant displacement from FE simulations, showing the interaction of *SH*0 mode with the crack at an oblique angle of incidence. The total magnitude I of the incident *SH*0 waves was obtained for a point along their principal axis at a distance d_I from the source. For specularly

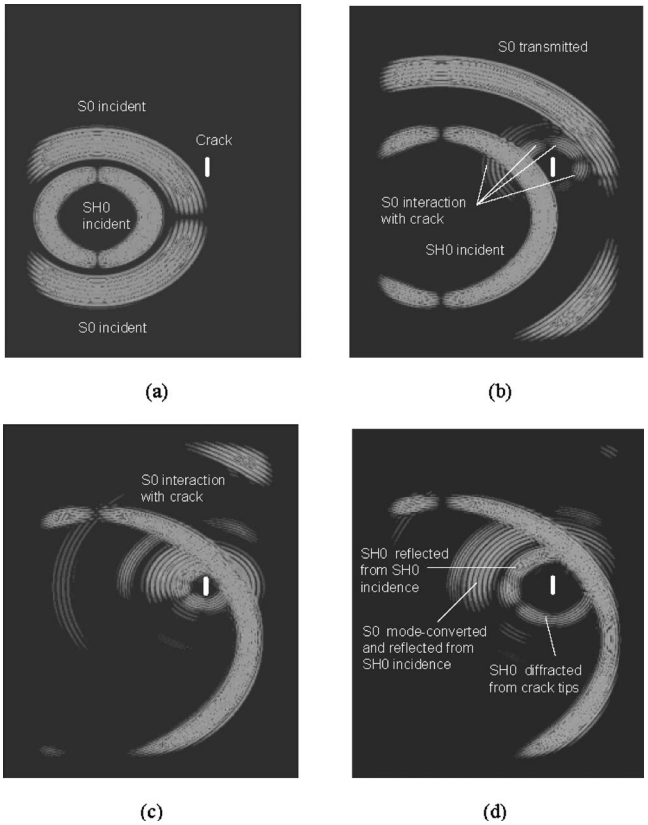


FIG. 6. Snapshots of the contour of resultant displacement magnitude from FE simulations for an oblique incidence angle. Due to absorbing boundaries present all around the actual plate, the interaction of the waves with the crack are isolated. (a) The generation of the waves. (b) The weak interaction of the *S*0 mode with the crack. (c), (d) The interaction of the SH0 mode with the crack.

reflected waves at angles $\theta_R = -\theta_I$, nodal displacements R_{U2} in the $U2$ direction were monitored along a vertical line at a distance $x_R = 16\ \lambda_{SH0}$ from the crack and a distance $y_R \approx x_R \tan(\theta_I)$ from its symmetric normal. The resultant displacement values were then obtained as $R = R_{U2}/\cos(\theta_R)$. The values in the range from −55° to 0° were assumed to be identical to these due to the symmetry in the configuration. The reflection ratio was then calculated using Eq. (1) with $D = x_I/\cos(\theta_I)$ and $d_M = x_R/\cos(\theta_I)$.

Experiments were again carried out on slits of 1 λ_{SH0} and 5 λ_{SH0} achieved using the required center-frequency values on slits of lengths 8 and 65 mm, respectively. Measurements were taken using exactly the same procedure for the FE simulations, with appropriate excitation and monitoring positions in terms of λ_{SH0}. The reflection ratio was calculated using Eq. (1) and a similar procedure as that for the FE predictions.

Figure 7(a) shows FE predicted values of the reflection ratio plotted with $\theta_R = -\theta_I$ and Figs. 7(b) and 7(c) present comparison with experiments for two slit length cases. We find a fairly good agreement between FE and experimental results except at large angles of incidence, where the crack tip (multiple) diffraction contributes strongly to the reflected field but the tip of the experimental notch does not represent the crack tip very well. Also, the presence of unwanted reflections from the plate edges makes it difficult to take measurements at these angles.

From the results we see that as the crack length increases, the highest specular reflection occurs at normal incidence, while almost no reflection is observed at a constant incidence angle of around 30° to the crack normal. In order to understand this behavior, Fig. 7(a) also overlays the reflection coefficient predicted for plane *SV* wave reflection from a straight edge. When a plane *SV* wave is incident on a straight edge, the reflected energy is repartitioned into *SV* and longitudinal (L) waves traveling at angles given by Snell's law. As the angle of incidence increases, more and more energy is delivered to the reflected L wave. This happens until the first critical angle θ_{crI} (around 29° for aluminum) at which the L wave begins to graze through the surface of the edge and almost all the reflected energy goes into the L wave. Beyond this angle, energy is again diverted to the reflected SV wave and the reflection coefficient rises with the incidence angle until it settles at a steady value. What we see is that as the crack length increases the reflection behavior from a finite crack increasingly tends toward that of the straight edge case. However, in the finite crack case, there is an interference characteristic of the presence of diffraction components which are not much separated in time. We will further discuss these findings and also their implications for imaging applications in Sec. IV A.

B. Angular influence on diffraction

The problem of the scattering of bulk *SV* waves from semi-infinite internal strip cracks is analytically solvable and exact solutions have been published:[17,18] because of the equivalence of the physics involved, in Ref. 6 we used these results to obtain a validation of our diffraction modeling. In

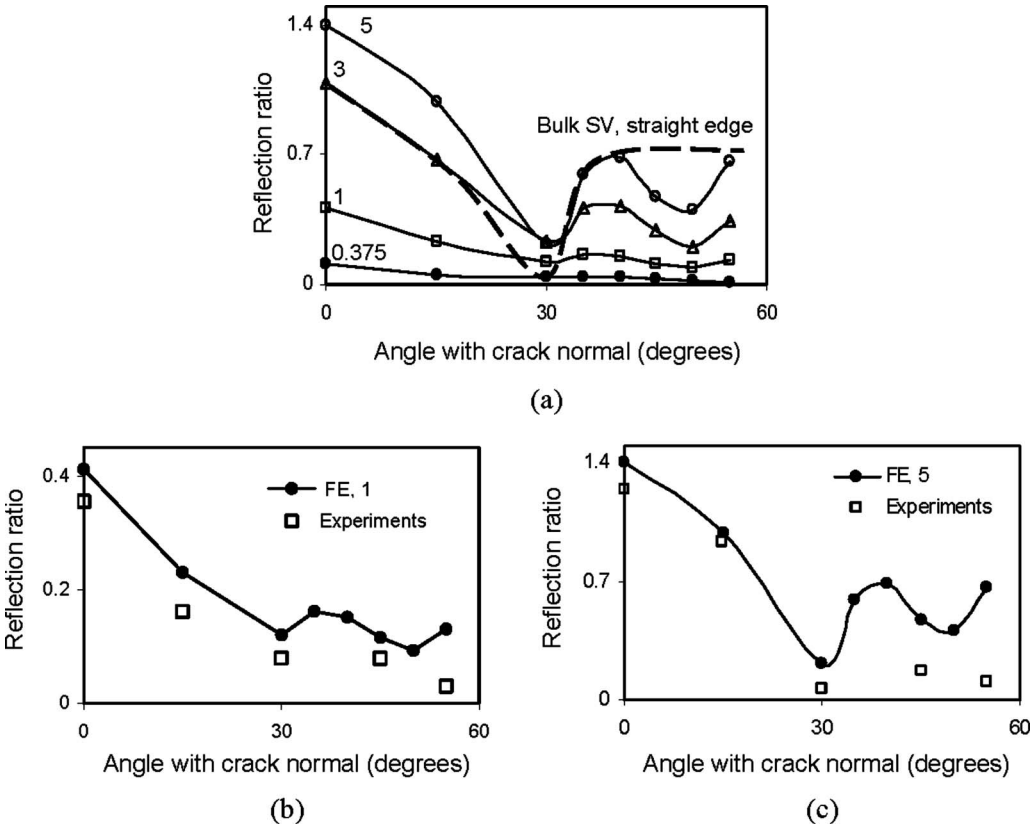


FIG. 7. (a) FE predicted values of the reflection ratio plotted at $\theta_R=-\theta_1$; the crack lengths are indicated in terms of λ_{SH0}. (b), (c) Comparison with experiments for two slit length cases of 1 and 5 λ_{SH0}.

studying the angular influence on the diffracted field, here we revisit this equivalence on the premise that the primary diffraction is just what would result from the tip of a semi-infinite crack. So we first take up the case of a semi-infinite crack, where diffraction consists solely of the primary component from the single tip, and examine its angular behavior. We then consider finite cracks which are of sufficient length to allow separation of the primary and secondary diffraction fields, and again study the former.

For both these studies, the validation of diffraction modeling as well as the study of its nature was performed by comparison of the FE results with those from Scruby *et al.*[18] for the equivalent bulk *SV* wave problem. For this purpose we make use of a diffraction coefficient

$$|F_s(\theta,\beta)| = \frac{|\psi_D|}{|\psi_I|}\sqrt{\frac{r_M}{\lambda_S}}, \tag{2}$$

where ψ_D is the frequency-domain amplitude of the monitored diffracted signal and ψ_I that of the wave incident at the crack tip, λ_S is the wavelength of bulk shear waves, r_M is the distance from the crack tip to the observation point, and θ, β are the angles of incidence and diffraction at the crack tip measured in a counterclockwise sense from the crack face [see Fig. 8(a) for an illustration].

Finally we will briefly discuss the case of shorter cracks where such separation cannot be achieved.

For all the sets of diffraction studies, the angles are defined counterclockwise from the crack fact t*o a ray pointing to or from the crack tip in the required direction* as shown in Fig. 8(a).

1. Diffraction from single-tipped or "semi-infinite" cracks

The FE model created for this purpose consisted of a long single-tip crack exposed to *SH*0 waves incident at various angles θ_I, as shown in Fig. 8(a). The *U*1 amplitude of the diffracted *SH*0 waves D_{U1} is monitored at a range of angles θ_D around the crack tip and the resultant amplitude obtained as $\psi_D=D_{U1}/\cos(\theta_D)$. Equation (2) is then used to calculate the diffraction coefficient $|F_s(\theta,\beta)|$. Figures 8(b)–8(f) show the values of the coefficient for five different incidence angle cases compared with the theoretical predictions by Scruby *et al.* In all cases we see that the amplitude of diffraction can be obtained to a very good accuracy from the bulk *SV* wave semi-infinite crack solution.

2. Diffraction from long finite cracks

Thus we would expect that even for finite cracks, if we can isolate the primary diffraction at each of the two crack tips, its behavior must correspond to that of bulk *SV* wave values for appropriate angles of incidence at a semi-infinite crack. From this perspective, using a five-cycle Hanning windowed tone burst excitation with a 100 kHz center frequency, a crack longer than 7 λ_{SH0} is suitable and a value of 10 λ_{SH0} was used for the FE studies. Details of the model used are shown in Fig. 9. A point source vibrating in the *U*2

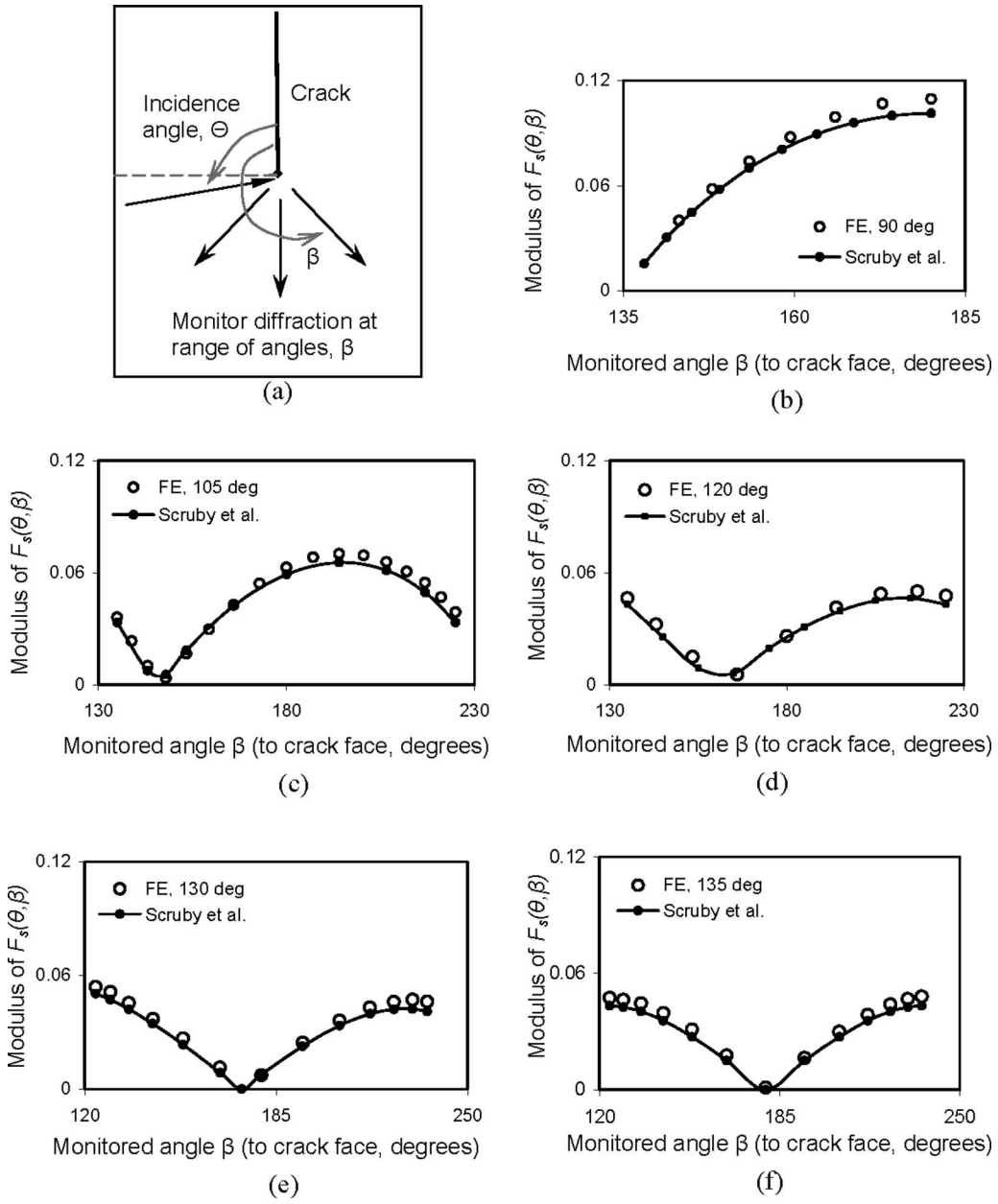


FIG. 8. (a) Representation of the model generated for diffraction validation. (b)–(f) Comparison of results from FE simulations with the theoretical calculations presented by Scruby *et al.* (Ref. 18) for incidence angle cases of 90°, 105°, 120°, 130°, and 135°, respectively.

direction parallel to the crack face is located along its symmetric normal at a distance $D=32\ \lambda_{SH0}$ from it. This causes the incidence of *SH*0 waves at the two crack tips (labeled Tip 1 and Tip 2) at angles θ_{I1} and θ_{I2} with the crack face. The *U*2 displacement of the incident waves I_{U2} was monitored at a distance d_0 from the source along their principal axis: The amplitude incident at each crack tip n ($n=1$ or 2) situated at a distance $d_{In}=D/\sin(\theta_{In})$ from the source would then be[7]

$$\psi_I = \frac{I_{U2}\sqrt{d_0}}{\sqrt{d_{In}}}\sin(\theta_{In}).$$

The *U*1 component of the diffracted signal D_{U1} was monitored at a range of angles θ_{Dn} around both the crack tips ($n=1$ and 2) at radial positions r_M, with the resultant displacement obtained as $\psi_{Dn}=D_{U1}/\cos(\theta_{Dn})$.

Equation (2) is then used to calculate the modulus of the diffraction coefficient and Fig. 10 compares it with the results from bulk *SV* wave scattering literature.[18] The FE results for angles around both the crack tips are mutually identical as predicted by, and in good agreement with, the theoretical calculations, thus confirming our expectation. The small deviations between the two are because of the presence of unwanted signals which arrive together with the primary diffraction. We thus note that even for finite cracks, the angular pattern as well as the amplitude of the primary diffraction can be estimated quite well from the theoretical results for bulk *SV* waves.

3. Diffraction from shorter finite cracks

In the case of shorter cracks where the primary and secondary diffraction cannot be separated, the diffraction behav-

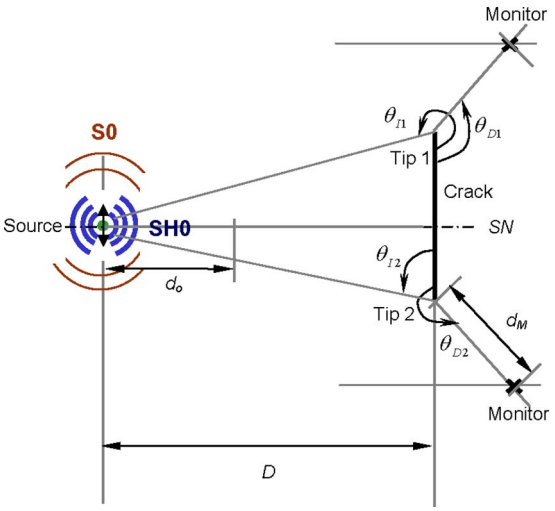


FIG. 9. (Color online) Geometrical details of the model used for study of diffraction from finite cracks.

ior will be modulated by the presence of these additional components. Consistent trends may not exist in this case and due to the presence of multiple interfering components it would be advisable to work at a single frequency. As an illustration of a typically mixed signal, we present a time trace of diffraction from the tip of a 5.25 λ_{SH0} long crack, for a normally incident *SH*0 wave in Fig. 11(a); Fig. 11(b) shows the Fourier transform of just the mixed diffraction signal: Interference due to multiple components can be observed.

IV. DISCUSSION

A. Understanding reflection behavior

Trends similar to those of the results for guided *SH* wave reflection from through cracks presented in Sec. III A are known for the analogous problem of bulk shear wave scattering from striplike cracks (see, e.g., Danilov[19] and more recently, Caleap *et al.*[20]). Here we seek a physical explanation for this behavior with the help of the analysis and arguments presented in Ref. 6, invoking a Kirchhoff-type approximation. A key insight from this analysis was that cracks that are small compared to the distance from the source could be considered to behave as if uniform point shear

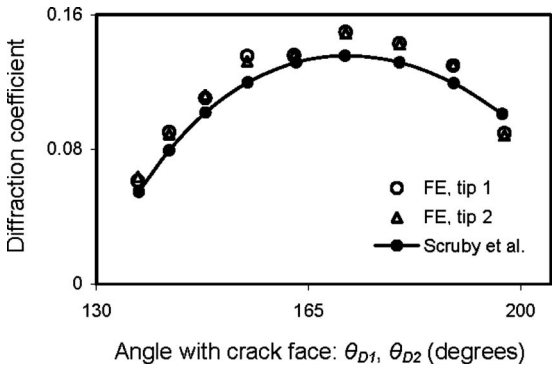


FIG. 10. Modulus of diffraction coefficient from FE simulations of *SH*0 waves incident symmetrically at a 10 λ_{SH0} long through-thickness crack, for the primary diffraction at both the crack tips, compared with theoretical results from bulk *SV* wave scattering literature (Ref. 18).

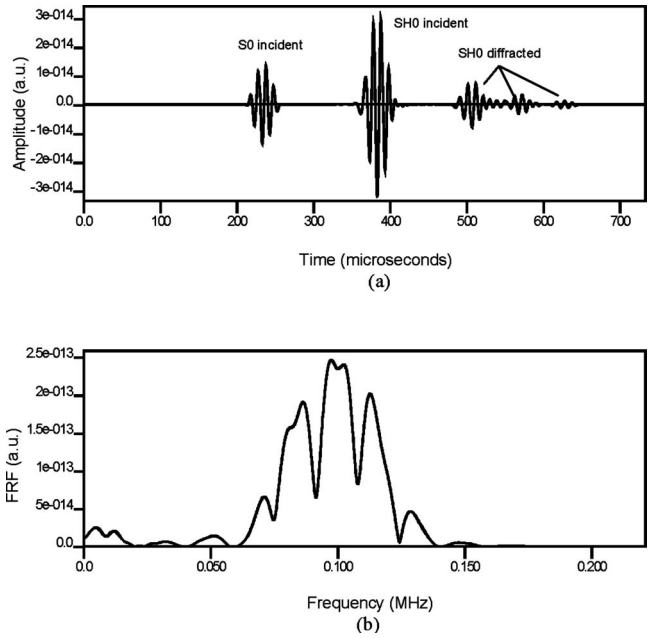


FIG. 11. (a) Time trace of diffraction from the tip of a 5.25 λ_{SH0} long crack for a normally incident *SH*0 wave. (b) Fourier transform of just the mixed diffraction components.

sources were placed along the insonified face—this was shown to be valid for cracks as long as $7\lambda_{SH0}$ for the range of distances of interest to us.

The observed angular behavior of reflection for normal incidence, as noted in Sec. III A 1, with a narrowing down of the field at increasing crack lengths, does conform to such a model. Here, we make use of FE simulations to further quantitatively compare the behavior of this simple Kirchoff approximation based model with the reflection results from earlier full-FE studies. One face of the crack was loaded with vibrating parallel point forces and the generated amplitude was monitored at a range of angles around the crack face, using the same procedure as described in Sec. III A 1. This was repeated for cracks of lengths 1, 3, and 5 λ_{SH0}, and for each crack length the monitored amplitude was normalized to the peak value: Figure 12 compares these with similarly normalized values for the same crack lengths from the results of Fig. 4. We observe a very good agreement, revealing that the angular trend of reflection at normal incidence noted earlier does indeed arise from the crack exhibiting an ideal shear-source-like behavior.

The trend of reflections at other angles of incidence considered in Sec. III A 2 can also be understood from a Kirchhoff approximation framework. The main contribution to the specular reflection field can be said to be occurring from the central region of the crack, where there is very little interaction between the crack faces. Therefore, incident waves would reflect from the crack face locally as if it were an infinite straight edge and thus the reflection ratio would approach that from such an edge for increasing crack lengths. We indeed discern such a trend from the results presented in Fig. 7.

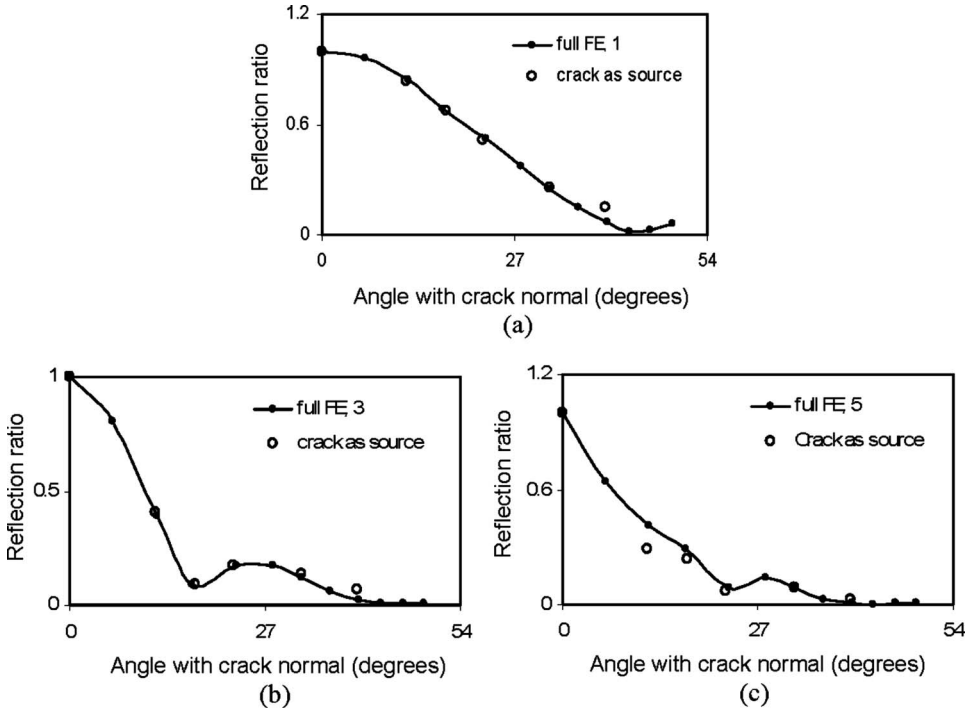


FIG. 12. Comparison of amplitudes from part-FE simulation of crack face loaded with parallel point forces and reflection results from Fig. 4. In both cases, the values are normalized with the peak value, giving a unit maximum. (a), (b), (c) Results for cracks of lengths 1, 3, and 5 λ_{SH0}, respectively.

B. Implications for imaging

We infer from the results of the normal incidence study that *SH*0 mode reflection from planar through cracks is likely to be concentrated in a narrow angular band around the specular direction. The oblique incidence studies show that the specular reflection itself is strongest at normal incidence, but drops rapidly at other angles, reaching a distinct minimum around the material's first critical angle. These factors are likely to play a key role when defining the limits for the spatial resolution achievable by different imaging methods.

Provided that the crack and signal length are appropriate to allow the primary component to be isolated, the variation of diffraction with the angular position can be predicted from known solutions from bulk *SV* wave scattering literature. This can potentially yield important additional information such as the crack size and orientation, which would be useful for more precise imaging. For smaller cracks, the presence of multiple interfering components means that a single frequency for calculations would be important.

V. CONCLUSIONS

In this paper, we studied the angular profile of the scattering when cylindrical crested low frequency *SH*0 waves interact with through-thickness cracks. The interaction was studied in terms of both the reflection from the crack face and the diffraction at the crack tips. From two complementary sets of studies, it is seen that the reflection is mainly concentrated in a narrow angular band around the specular direction and that it approaches that from a straight edge for large cracks. The specular reflection itself varies quite rapidly with the incidence angle, and is strongest at normal incidence. If the crack is long enough to allow isolation of the primary tip diffraction at the operating frequency, a good estimate of its amplitude and angular variation can be obtained from known solutions of bulk *SV* wave canonical scattering problems. These conclusions will help in defining limits for the spatial resolution achievable by imaging methods and possibly in improving imaging precision by yielding additional information about the crack. The next and ongoing step in this research involves part-through instead of through-thickness cracks.

[1]J. Davies, F. Simonetti, M. J. S. Lowe, and P. Cawley, "Review of synthetically focused guided wave imaging techniques with application to defect sizing," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2006), Vol. 25, pp. 142–149.
[2]J. Davies and P. Cawley, "The application of synthetically focused imaging techniques for high resolution guided wave pipe inspection," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2007), Vol. 26, pp. 681–688.
[3]P. Cawley and D. Alleyne, "The use of Lamb waves for the long range inspection of large structures," Ultrasonics **34**, 287–290 (1996).
[4]D. Alleyne, B. Pavlakovic, P. Cawley, and M. J. S. Lowe, "Rapid long range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96 (2001).
[5]P. Wilcox, "Omni-directional guided wave transducer arrays for the rapid inspection of large areas of plate structures," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 699–709 (2003).
[6]P. Rajagopal and M. J. S. Lowe, "Short range scattering of the fundamental shear horizontal guided wave mode normally incident at a through-thickness crack in an isotropic plate," J. Acoust. Soc. Am. **122**, 1527–1538 (2007).
[7]P. Wilcox, "Modeling the excitation of Lamb and SH waves by point and line sources," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2004), Vol. 23, pp. 206–213.
[8]A. A. Oliner, "Waveguides for acoustic surface waves: A review," Proc. IEEE **64**, 615–627 (1976).
[9]A. K. Gautesen, and J. D. Achenbach, "On the existence of surface waves

guided by a slit," SIAM J. Appl. Math. **35**, 301–306 (1978).
[10]B. K. Sinha, "Some remarks on propagation characteristics of ridge guides for acoustic surface waves at low frequencies," J. Acoust. Soc. Am. **56**, 16–18 (1974).
[11]*ABAQUS v6.6 Analysis User's Manual*. For information on ABAQUS FE package see http://simulia.com/products/abaqus_fea.html (date last viewed, 10 July 2008).
[12]D. N. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[13]M. Castaings, C. Bacon, B. Hosten, and M. V. Predoi, "Finite element predictions for the dynamic response of thermo-viscoelastic material structures," J. Acoust. Soc. Am. **115**, 1125–1133 (2004).
[14]M. Castaings and C. Bacon, "Finite element modeling of torsional wave modes along pipes with absorbing materials," J. Acoust. Soc. Am. **119**, 3741–3751 (2006).
[15]M. Drozdz, M. J. S. Lowe, E. Skelton, and R. Craster, "Modeling bulk and guided wave propagation in unbounded elastic media using absorbing layers in commercial FE packages," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2007), Vol. 26, pp. 87–94.
[16]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, UK, 1999).
[17]J. A. Ogilvy and J. A. G. Temple, "Diffraction of elastic waves by cracks: Application to time-of-flight inspection," Ultrasonics **21**, 259–269 (1983).
[18]C. B. Scruby, K. R. Jones, and L. Antoniazzi, "Diffraction of elastic waves by defects in plates: Calculated arrival strengths for point force and thermoelastic sources of ultrasound," J. Nondestruct. Eval. **5**, 145–156 (1986).
[19]V. N. Danilov, "Scattering of transverse waves by strip-shaped cracks," Russ. J. Nondestruct. Testing **38**, 582–592 (2002) [translated from Defektoskopiya **8**, 31–43 (2002)].
[20]M. Caleap, C. Aristégui, and Y. C. Angel, "Further results for antiplane scattering by a thin strip," J. Acoust. Soc. Am. **122**, 1876–1879 (2007).

Verification of chaotic behavior in an experimental loudspeaker

Joshua D. Reiss
Department of Electronic Engineering, Queen Mary, University of London, Mile End Road, London E14NS, United Kingdom

Ivan Djurek[a] and Antonio Petosic[b]
Department of Electroacoustics, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, Zagreb, Croatia

Danijel Djurek[c]
Alessandro Volta Applied Ceramics, (AVAC) Laboratory for Nonlinear Dynamics, Kesten brijeg 5, Remete, Zagreb, Croatia

(Received 24 January 2008; revised 23 April 2008; accepted 4 July 2008)

The dynamics of an experimental electrodynamic loudspeaker is studied by using the tools of chaos theory and time series analysis. Delay time, embedding dimension, fractal dimension, and other empirical quantities are determined from experimental data. Particular attention is paid to issues of stationarity in a system in order to identify sources of uncertainty. Lyapunov exponents and fractal dimension are measured using several independent techniques. Results are compared in order to establish independent confirmation of low dimensional dynamics and a positive dominant Lyapunov exponent. We thus show that the loudspeaker may function as a chaotic system suitable for low dimensional modeling and the application of chaos control techniques.
 [DOI: 10.1121/1.2967843]

PACS number(s): 43.38.Ja, 43.25.Ts, 43.25.Rq [AJZ] Pages: 2031–2041

I. INTRODUCTION

Loudspeakers are the most variable elements in any audio system and are responsible for marked audible differences between otherwise identical sound systems. The performance of loudspeakers (i.e., their accuracy in reproducing a signal without adding distortion) is significantly poorer than that of other audio equipments. For example, harmonic distortion in a typical loudspeaker can be 100 to 1000 times greater than that of amplifiers.[1] The frequency response of a loudspeaker is often referenced as being within ±3 dB of perfect linearity (and many speaker designs fall further outside this range), whereas that of an amplifier may vary by less than 0.1 dB. Thus an improved understanding of the dynamics of a loudspeaker is of great importance since a better loudspeaker design may yield a significantly better performance in an audio playback system.

In a dynamic loudspeaker, sound is typically reproduced by the movement of a current-carrying object in a magnetic field due to the Lorentz force. An electrodynamic loudspeaker, as depicted in Fig. 1, is a type of a dynamic loudspeaker in which the magnetic field is produced by a permanent magnet. The audio signal from the amplifier, which is a time-varying voltage, causes a current to flow through a tightly wound coil of wire called the voice coil. The voice coil is attached to a stiff but light cone and connected to a fixed chassis by means of a flexible membrane. The spider, a weblike membrane, suspends the voice coil in the center of the magnet and provides part of its restoring force. The surround connects the cone to the speaker chassis and keeps nonaxial motion to a minimum so that the cone functions approximately as a rigid piston. The interaction of the coil and the driver's magnetic system generates a mechanical force that causes the coil, and thus the attached cone, to move back and forth. This creates pressure waves in the air, which reproduce the original audio signal coming from the amplifier as sound.

A simplified model of an electrodynamic loudspeaker operates as a driven damped harmonic oscillator.[2] The dynamics of the displacement of the membrane, x, is given by the second order time-dependent ordinary differential equation

$$M\frac{d^2x}{dt^2} + R_M\frac{dx}{dt} + kx = BlI_0\cos(\omega t), \tag{1}$$

where M is the combined mass of the membrane and voice coil, k is the elastic stiffness, B is the magnetic field, l is the length of the voice coil wire, and I_0 represents the drive current with frequency $f_0=\omega_0/(2\pi)$. The friction term R_M consists of internal membrane friction and friction coming from vibration in air. Note that this model is not used in the analysis of the experiment but is given in order to show the need for higher order nonlinear terms to accurately represent the dynamics.

Both an inhomogeneous magnetic field and the excursion-dependent stiffness of the suspension account for the nonlinear behavior of the transducer. Thus, all coefficients in Eq. (1) are dependent on displacement x. The mass M was measured by weighing the detached vibrating parts and was found to be 13.1 g. However, due to the elastic

[a]Electronic mail: ivan.djurek@zg.t-com.hr
[b]Electronic mail: antonio.petosic@fer.hr
[c]Electronic mail: danijel.djurek@zg.t-com.hr

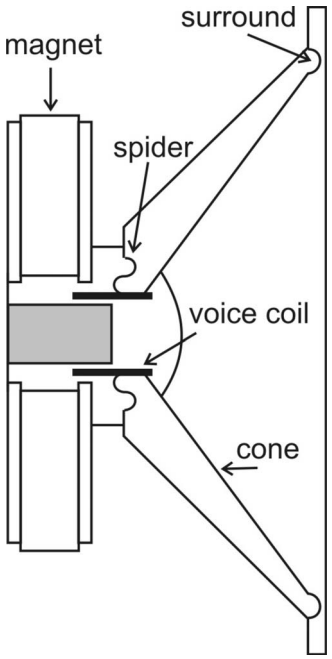


FIG. 1. Diagram of an electrodynamic loudspeaker.

properties of the membrane, which spans between the voice coil and the suspension on the perimeter, M should be replaced by an effective mass, M_{eff}. M_{eff} was evaluated from the Q-factor of the loudspeaker[3,4] and, as expected, was found to be dependent on displacement. For higher driving currents, the parameter Bl decreases with displacement nearly symmetrically relative to a zero position.[5] In these experiments, $Bl=3.9$ T m and reduces to ~ 3.6 T m for a membrane displacement of ± 6 mm, and corresponding nonlinearities can be neglected. In most commercially available loudspeakers, the elastic stiffness has a quadratic dependence on displacement. A systematic approach to describing such a vibrating system, using the Duffing equation, was provided by Woafo,[6] where it was shown that related models of electromechanical transducers may exhibit chaos. However, Woafo also noted that a linear version is used to describe loudspeakers and that nonlinear terms, which are known to cause subharmonics, are often neglected.[6,7]

The stiffness was recorded in static measurements by the use of calibrated loads and corresponding measurements of the membrane deflections and may be given as approximately $k=m+nx+px^2$, where $m=480$ N m^{-1}, $n=31$ N m^{-2}, and $p=7.5$ N m^{-3}. Due to the viscoelastic contribution of the membrane polymer material, this disagrees substantially with the stiffness calculated in the dynamic regime from the simple resonance formula $\omega=k/M$, even in the regime of comparatively small drive currents ($I_0<100$ mA). Thus higher order corrections coming from the dynamics of a vibrating viscoelastic body[8,9] must be included.

When dealing with moderately low driving currents, a significant agreement was found between a model based on static measurements and the experimental data, especially for an estimation of frequency dependence and electrical impedance under a variety of different operating regimes.[10] Although simple models can be used to identify loudspeaker parameters, at high drive currents ($I_0>1$ A) they generally do not account for the rich nonlinear dynamic behavior that might be observed.[11–13]

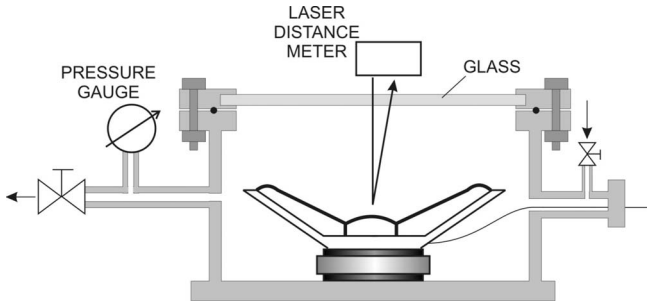


FIG. 2. Experimental setup with a loudspeaker and a laser distance meter.

Evidence of a possible chaotic behavior in a loudspeaker was first observed by Wei *et al.* in 1986,[14] where the appearance of subharmonics and broadband spectra at various drive frequencies and voltages was noted. Recent work strengthened this conjecture,[15–17] with the observation of hysteresis and period doubling when the loudspeaker was driven at low frequencies. This, in general, is not observed in the models. Thus it is important to verify chaos and adjust the models accordingly.

In this work, we analyze time series from an experimental electrodynamic loudspeaker system. We use a variety of techniques from the chaotic time series analysis[18] to show that the system is indeed chaotic and exhibits low dimensional dynamics suitable for further analysis and the implementation of chaos control techniques. By quantifying the nonlinear behavior, we also provide empirical observations, which may be used to refine the modeling and design of loudspeakers.

II. EXPERIMENTAL SETUP

In these experiments, a low frequency loudspeaker was used with a resonant frequency, recorded in air, of $f=38$ Hz, driving current of $I_0=10$ mA, factor of $Bl=3.9$ T m, membrane diameter of $2R=16$ cm, rated rms power of 60 W, and nominal impedance of 8 Ω. According to the manufacturer's data, the voice coil inductance is $L=0.9$ mH, and the contribution of the inductive part ωL to the loudspeaker's impedance can be neglected for driving frequencies $f<100$ Hz.

The experimental setup is depicted in Fig. 2. The loudspeaker was placed in a stainless steel chamber. Air pressure within the chamber was measured by the use of an absolute capacitive gauge with an ultimate resolution of 0.01 mbar. A glass window on the top of the chamber ensured the transparency to the light beam from the laser distance meter, which measured the vibration amplitude with an accuracy of 2 μm and a sampling frequency of ~ 1 kHz (a similar measurement apparatus was used to analyze nonlinear vibrations of a loudspeaker in Wei *et al.*[14]). The analog/digital (A/D) converter resolution of the laser distance meter was 8 bits. This allowed the acquisition of 128 amplitude levels in both up and down vibration directions. In order to check the possible influence of the chamber wall friction on the course of

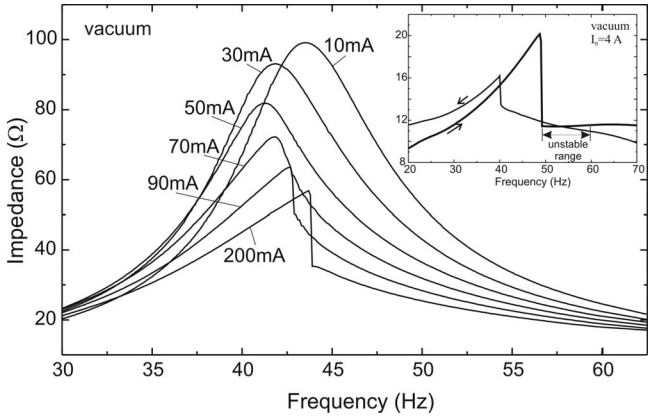


FIG. 3. Impedance measured *in vacuo* for various driving currents. The inset shows the hysteresis of the cutoff frequency for positive and negative frequency sweeps.

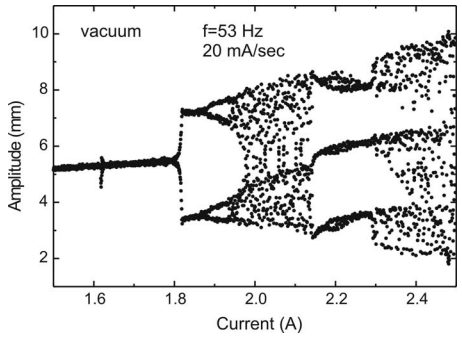


FIG. 4. A bifurcation diagram of vibration amplitudes recorded *in vacuo* for a fixed driving frequency of 53 Hz and a sweeping driving current of 20 mA/s.

measurements, the impedance and vibration amplitude data were recorded at 1 bar air pressure in a closed chamber and compared to those evaluated in free laboratory atmosphere. In a frequency range near 50 Hz, recorded data, notably impedance, showed no significant difference.

For impedance and amplitude measurements, the loudspeaker was connected to an audio amplifier with a rated power of 300 W via a series resistor 0.44 Ω. A rather small resistor value was used because of the possibility of driving higher currents. However, small resistance gave rise to increased influence of the back electromotive force. A satisfactory compromise was found with the total voltage swing across the loudspeaker, clipping not included, of ±50 V, which, in turn, provided driving currents $I_0=4-5$ A. For driving currents $I_0<100$ mA, the back electromotive force is comparable to the friction term in Eq. (1), while for higher currents intrinsic friction increased and became the dominant contribution to the impedance.

The loudspeaker vibration amplitude dependent on frequency was measured for various driving currents in an evacuated space and in normal atmospheric pressure (1 bar).[19] The data recorded *in vacuo* are shown in Fig. 3. By an increase in the driving current, the resonance curve became more and more distorted until an amplitude downturn (cutoff) appeared near $f=43.5$ Hz for $I_0=200$ mA. This current indicated the starting value for an identification of the chaotic regime. The inset in Fig. 3 depicts the hysteretic property of the cutoff effect. That is, cutoff frequencies differ for positive and negative frequency sweeps.

The onset of the amplitude cutoff was followed by a subsequent frequency sweep, which gave rise to the erratic vibration amplitude. This unstable range extended up to 54 Hz. This was chosen as the fixed frequency for an evaluation of the appearance of subharmonics that precede the chaotic state. An important feature of the unstable range was the relatively small change in the impedance with increasing driving current. In these experiments, loudspeaker impedance stayed at ~11 Ω, irrespective of whether the system was operated in air or in an evacuated chamber. This, in turn, meant that the driving system could be considered as a current source, even in the case when the amplifier was used as a voltage source.

A bifurcation diagram, as shown in Fig. 4, was produced by fixing the frequency at $f=53$ Hz, sweeping the driving current at $I_0=1.5-2.5$ A at a rate of 20 mA/s, and recording the amplitude of the displacement of the loudspeaker. At $I_0 \sim 1.81$, a first bifurcation pitchfork appeared, which was followed by multiple period doublings, until at $I_0=2.15$ A the characteristic period 3 window appeared[20] and the system vibrated with three amplitudes. The existence of period doubling and a period 3 window was a strong indicator of chaos.

However, the period 3 window was observed only in an evacuated space, and data from a loudspeaker operating in an evacuated chamber are unsuitable for time series analysis techniques. This is primarily because the voice coil bonding agent evaporates when in a vacuum, which, in turn, changes the resonant frequency during the course of the measurements. In addition, for a long time and heavy duty loudspeaker operation, it is important to remove heat from the loaded voice coil, and this is more easily accomplished in an air atmosphere.

Measurements in 1 bar air were performed in a closed chamber since this minimized parameter drift due to free air convection in the laboratory. The driving frequency was fixed at 56 Hz, and the driving current was increased up to values when higher harmonics appeared as a result of the nonlinear restoring term in Eq. (1). Excerpts of the time series waveforms representing various driving currents are given in Fig. 5. Figure 5(a) shows the time-dependent vibration amplitude at the starting driving current $I_0=2.4$ A when the recorded signal shows a nearly sinusoidal behavior. Figures 5(b)–5(d) show new vibration amplitudes (marked with triangles), which appear with increasing driving current. The corresponding averaged spectra over the whole time series are given in Fig. 6. Figure 6(a) shows the expected behavior, with a fundamental frequency corresponding to the drive frequency. The first subharmonic appeared at 28 Hz, as depicted in Fig. 6(b). A further increase in the driving current resulted in a new subharmonic at 14 Hz [see Fig. 6(c)] and a subsequent appearance of broadband behavior, [Fig. 6(d)].

The bifurcation diagram for measurements of vibration amplitude in air is shown in Fig. 7 and was produced in the same manner as Fig. 4. Vertical lines indicate values of driving currents for which the Feigenbaum ratio $\delta_1/\delta_2=4.669$ is fulfilled.[21] A period 3 window cannot be seen, but this is not a contra-indicator of chaos. Noise in the data acquisition sys-

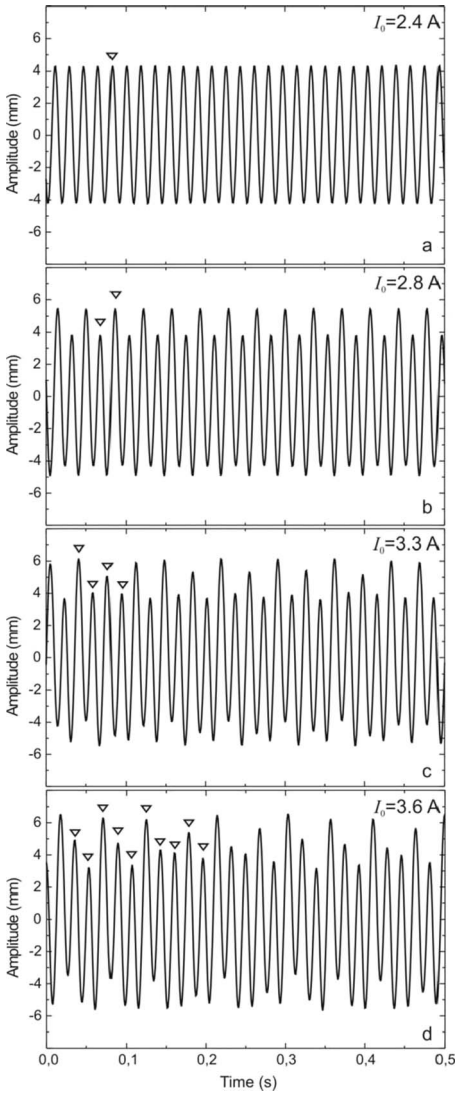


FIG. 5. Short time series plots of the vibration amplitude recorded in 1 bar air for a fixed driving frequency of 56 Hz and various driving currents. The triangles indicate the appearance of a broadband behavior.

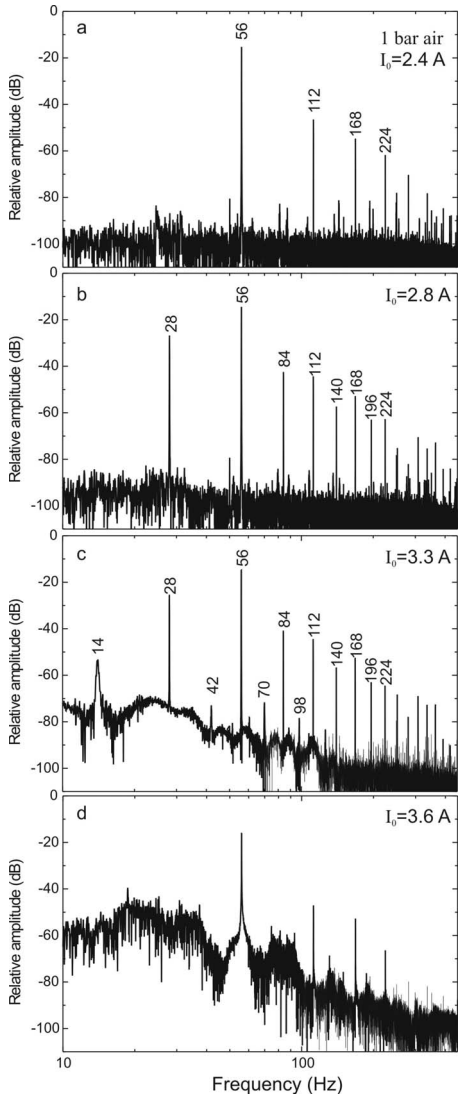


FIG. 6. Average spectrum of vibration amplitude recorded in 1 bar air for various driving currents corresponding to those in Fig. 5.

tem may obscure the window, and the existence of such a window is not considered a necessary condition. Whereas heating the voice coil makes long term measurements difficult for the evacuated loudspeaker experiment, reverberation and air circulation added noise to the short term measurements used in generating the bifurcation diagram of Fig. 7.

Leaving the loudspeaker to operate at a driving frequency of 45 Hz, the driving current was selected at a value slightly below 2.8 A, at which point the first subharmonic became attenuated. This indicated the starting point for recording the vibration amplitude included in a time series analysis. A rather low frequency of 45 Hz was selected because the dynamics appeared less susceptible to parameter drift and nonstationary behavior in this range.

The time series analysis presented in the following sections is derived from a 247 392 point experimental flow data set. The data were recorded with a 16 bit resolution, although this is further limited by the 8 bit resolution of the laser distance meter. The sample rate was 1024 Hz, so that the data set is just over 4 min long and there are approximately 22.76 samples per drive period. In the results that follow, units are not typically given on the measurements since they have been scaled and transformed by the data acquisition system.

III. DATA ANALYSIS AND RESULTS

To analyze a one dimensional experimental data set using chaotic time series analysis techniques, it is necessary to

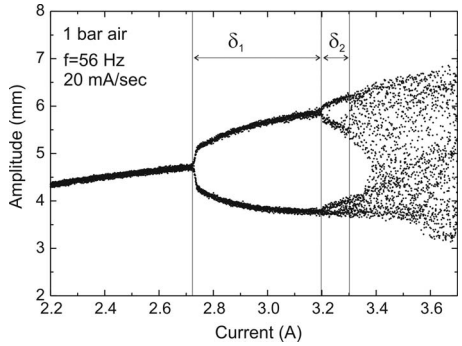


FIG. 7. A bifurcation diagram of vibration amplitudes recorded in 1 bar air for a driving current swept at 20 mA/s and a fixed driving frequency of 56 Hz.

transform the data using phase space reconstruction techniques. If only one variable from the system can be observed, $\mathbf{X}=\{X(1),X(2),\ldots,X(N),\ldots\}$, then a D dimensional time series of length N, $\mathbf{Y}=\{Y(1),Y(2),\ldots,Y(N)\}$ is constructed from the original time series using a delay d as follows:

$$\mathbf{Y}(n)=(X(n),X(n+d),\ldots,X(n+(D-1)d)), \tag{2}$$

where we have assumed that $\mathbf{X}$ contains at least $N+(D-1)d$ data points. If the time between samples represents one period of data, then $\mathbf{X}$ represents a time series generated from a map, or a Poincaré section of the system, in which case the delay d used to generate $\mathbf{Y}$ is usually set to 1.

Assuming that the time series is stationary, that is, the parameters that govern the dynamics are not significantly changing over time, then with sufficient data and the appropriate choice of the delay parameter d and the embedding dimension D, $\mathbf{Y}$ will accurately represent the dynamical behavior of the system. Once $\mathbf{Y}$ has been constructed, then further analysis of this multidimensional time series may be used to estimate various quantities related to the structure of the phase space, such as the dimensionality of the attractor, characterization of the chaotic behavior or lack thereof, and identification of chaotic orbits.

In the following subsections, we construct delay coordinate embeddings from the scalar time series and then use this technique to analyze the data and quantify the dynamical system properties. We will use the notation introduced above to describe the original scalar time series, $\mathbf{X}=\{X(1),X(2),\ldots,X(N)\}$, and a delay coordinate embedding of the time series, $\mathbf{Y}=\{Y(1),Y(2),\ldots,Y(N)\}$.

A. Nonstationarity and long term dynamics

A few simple tests that would identify strong drifts in the data were first performed. Sliding windows of varying lengths were applied to the data, and statistical quantities such as mean, maximum, minimum, and standard deviations were computed for each window. If the data were truly stationary, then these quantities would remain constant throughout the data. The results of the drift in the maximum and minimum values of a one second window (1024 data points) are depicted in Fig. 8.

It can be seen that the dynamics of the system are not entirely stationary. For instance, the maximum value undergoes an upward trend, particularly near the beginning of the time series. This nonstationarity was also confirmed by the measurement of other statistical quantities such as the skewness and kurtosis for windowed data.

When the dynamics change over time in an experimental system, it is often difficult to determine the cause. The behavior may be caused by long term dynamics, which are inherent to the system, or by a simple transient before settling into some behavior. However, this fluctuation is quite small in relation to the full extent of the data (for instance, the variation in the maximum value is less than 2% of the full scale of the data). Thus, although it may affect the results of chaotic time series analysis methods, it is still small enough such that the data are acceptable for analysis.

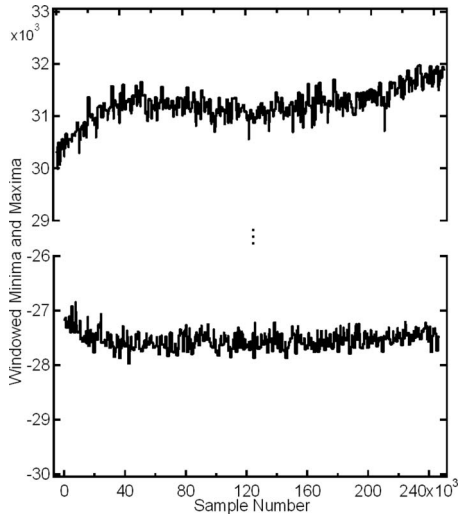


FIG. 8. Nonstationary behavior of the time series. Plotted are estimates of the maximum and minimum values for overlapping windows of 1024 samples (1 s) from the experimental data.

B. Poincaré sections

A common technique in a chaotic time series analysis is to generate a Poincaré section, with one point per period, from data sampled at a much higher frequency than the drive frequency. Given that the system has a drive frequency of 45 Hz, a natural Poincaré section would be to sample the system at the drive frequency. Since this was not possible due to the limited sampling frequencies of the data acquisition system, we considered several techniques for extracting a Poincaré section. These included the peak amplitude values, their second derivatives, times between peaks, and times at which the amplitudes cross a fixed value. Figure 9 depicts the Poincaré section plots using extracted peak amplitudes and extracted times between zero crossings in the flow data. Both techniques successfully capture the dynamics, although the use of zero crossings appears slightly less noisy.

C. Embedding parameters

A reasonable value for the delay may be suggested either by the first zero crossing of the autocorrelation function or by the first minimum of the mutual information function[22,23] as either value is plotted as a function of delay. For the time series data, given a delay d, the autocorrelation is found simply from

$$R(d)=\frac{1}{(N-k)\sigma^2}\sum_{n=1}^{N-k}[X(n)-\mu][X(n+d)-\mu], \tag{3}$$

where μ is the mean and σ^2 is the variance of the data.

The mutual information of two random variables, $a\in A$ and $b\in B$, is given by

$$I(A;B)=\sum_{a\in A}\sum_{b\in B}p(a,b)\log\frac{p(a,b)}{p(a)p(b)}, \tag{4}$$

where the convention $0\log 0=0$ is used. In the case of the mutual information between a time series and a delayed version of itself, a represents a range of values for $X(n)$ and b denotes a range of values for $X(n+d)$. These values must be

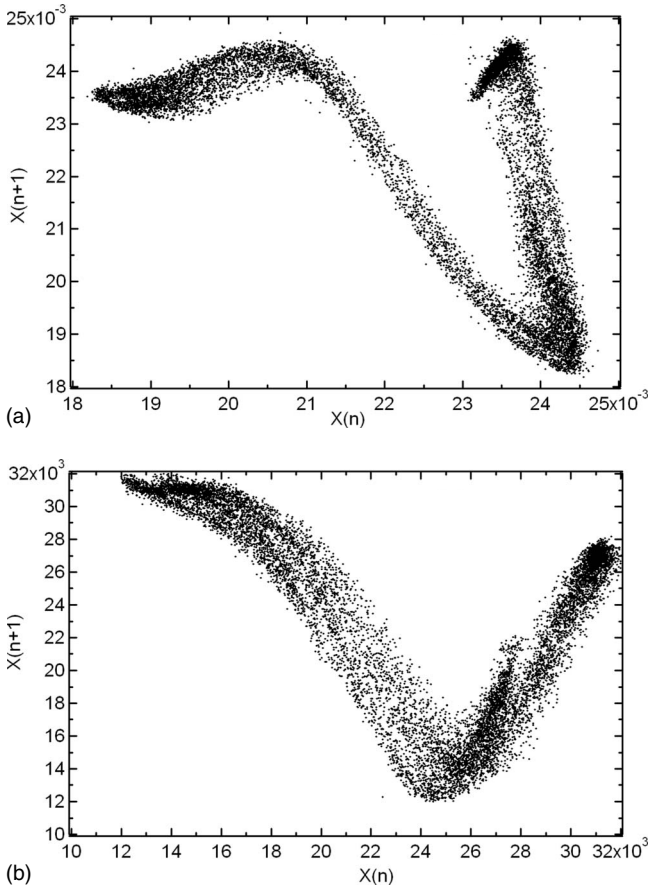


FIG. 9. Plot of the waveform for two methods of sectioning the data. The top plot depicts peak amplitudes and the bottom plot depicts time intervals between zero crossings.

chosen so as to provide a reasonable approximation to the mutual information of the underlying dynamical systems generating $X(n)$ and $X(n+k)$. Here, the mutual information was calculated efficiently using a method described in Reiss *et al.*,[24] which partitions the range of values for $X(n)$ and $X(n+k)$ recursively until there is no more hidden structure.

The mutual information often gives a better value because it takes nonlinear correlations into account. However,

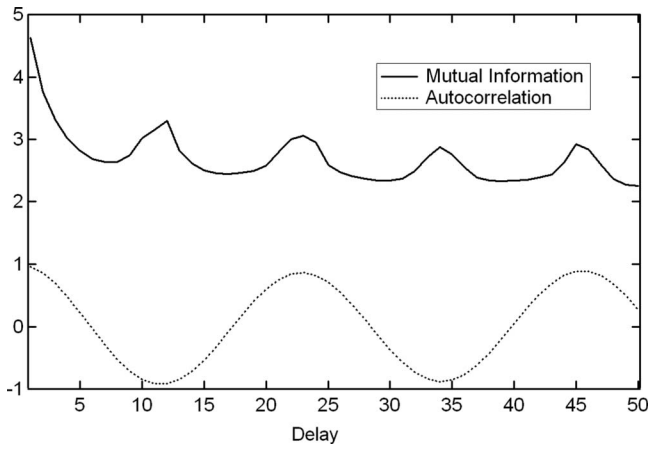


FIG. 10. Two techniques for estimating an embedding delay from the experimental data. The first method uses the first minimum of the mutual information function, and the second uses the first zero crossing of the autocorrelation function. The methods suggest a delay between 6 and 7.

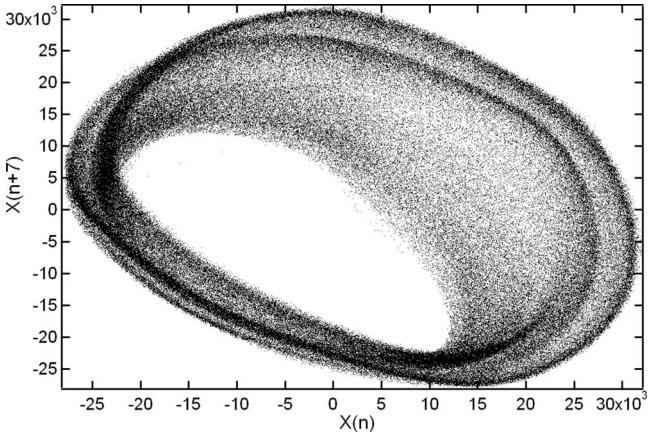


FIG. 11. A two dimensional plot of the experimental data with a delay of 7.

for the loudspeaker data, the mutual information function and the autocorrelation function were in strong agreement. As shown in Fig. 10, the autocorrelation suggests a delay of 6 and the mutual information suggests a delay of 7. This was in agreement with visual inspection since two and three dimensional plots revealed the most structure near these values of delay (see Fig. 11). Unfortunately, they also reveal a complexity or noise dependence that makes the fine scale structure very difficult to detect.

A modified form of the false nearest neighbor (FNN) algorithm[25] was chosen as the primary technique for determining the embedding dimension. The modification is intended to take into account stochastic phenomena, which result in FNNs occurring regardless of the embedding dimension. An appropriate embedding dimension is found when the percentage of FNNs drops to a constant value. As shown in Fig. 12, the percentage of false neighbors approaches a constant value with an embedding dimension of 5 for the flow data and an embedding dimension of 4 for either sectioned data set. This is in agreement with the observation that a Poincaré section should have one less dimension than the original data.

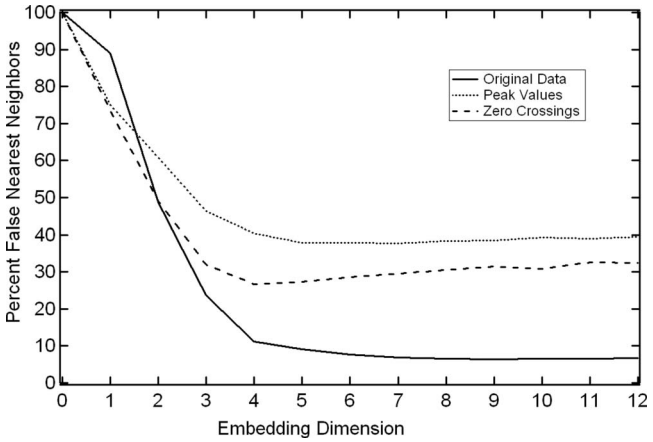


FIG. 12. Results of the FNN routine as applied to the flow data. An appropriate embedding dimension is found when the percentage of false near neighbors drops to a constant value. This indicates that the embedding dimension should be at least 4.

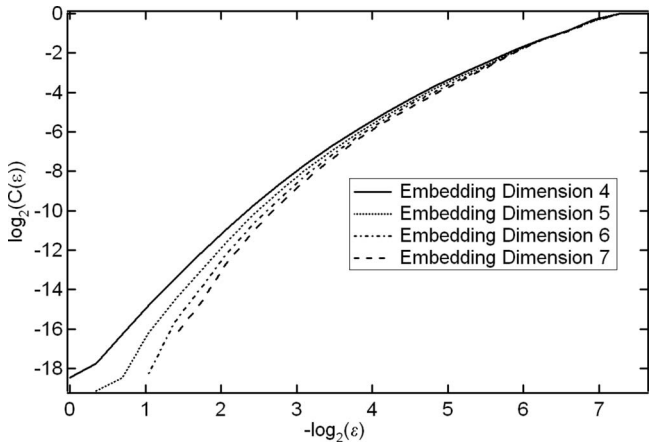


FIG. 13. Plot of the log correlation function vs log distance. With sufficient low noise data, the slope of the plot may provide the correlation dimension.

D. Fractal dimension

The dimensionality of a chaotic attractor is typically a noninteger value. That is, the attracting region of the phase space will not completely fill out a region of that space. In this section we use several different methods to estimate the dimensionality of the loudspeaker data, which further give an idea of the complexity of the underlying dynamics.

The correlation dimension is found by constructing a function $C(\varepsilon)$, which is the probability that two arbitrary points from the delay coordinate embedding are closer together than a distance ε,

$$C(\varepsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{i-1} H(\varepsilon - |Y(i) - Y(j)|), \tag{5}$$

where H is the Heaviside unit step function. The correlation dimension of an experimental time series is then given by

$$D = d \log(C)/d \log(\varepsilon) \tag{6}$$

in the limit $\varepsilon \to 0$ and $N \to \infty$. The correlation dimension may be estimated by the slope of the curve $\log[C(\varepsilon)]$ vs $\log(\varepsilon)$. A noninteger result for the correlation dimension indicates that the data are probably fractal. For too low or too high ε values, the results are inaccurate, so the slope must be measured in the midrange of the curve. A good value should be in the region where measurements of the dimension are most stable.

The Grassberger–Proccacia algorithm,[26,27] was used to estimate fractal dimensions. The results of $\log_2[C(\varepsilon)]$ vs $\log(\varepsilon)$ for the peak values from the original 45 Hz data are depicted in Fig. 13. The correlation dimension cannot be accurately estimated since there is not one significant region where the slope remains constant. This is because estimates of correlation dimension using the Grassberger–Proccacia algorithm are highly susceptible to noise and data set size. A limited data set size reduces the region of the plateau and increases uncertainty, and the presence of noise implies that an accurate measurement can only be obtained for large ε. For higher dimensional data these problems are aggravated since minimal noise and exponentially more data are required to identify the plateau region. In addition, data with a high sample rate may exhibit strong correlations that skew the estimates. The approximations due to noise, data set size, nonstationarity, and so on are inherent in the data set. However, the Grassberger–Proccacia algorithm also uses an approximation to the definition of correlation dimension. Therefore, we attempted an alternative technique that allows an estimation of the multiple definitions of the fractal dimension.

The delay coordinate embedded data can be gridded into n dimensional boxes of equal length ε, such that all vectors lie within these N boxes. If a box is labeled i, then it has an associated probability, $P_i(\varepsilon)$, that a vector on the attractor will reside within this box. The generalized entropies, $H_0, H_1, \ldots$, are defined in terms of the probabilities of vectors occupying the boxes. For $q = 0, 1, 2 \ldots$,

$$H_q(\varepsilon) = \frac{1}{1-q} \ln \sum_{i=1}^{N(\varepsilon)} P_i^q(\varepsilon), \quad q \neq 1,$$

$$H_1(\varepsilon) = -\sum_{i=1}^{N(\varepsilon)} P_i(\varepsilon) \ln P_i(\varepsilon). \tag{7}$$

In the following we keep the conventions common in dimension definitions and use the natural logarithm as opposed to log base 2. The generalized dimension of order q is then defined as

$$D(q) = -\lim_{\varepsilon \to \infty} \frac{H_q(\varepsilon)}{\ln \varepsilon}. \tag{8}$$

Under this definition, $D(0)$, $D(1)$, and $D(2)$ are the box counting dimension, the information dimension, and the correlation dimension, respectively. We also have the property that if $p > q$, then $D(p) \leqslant D(q)$.

Once the generalized entropies have been determined, there are two ways to approximate the generalized dimensions for time series data. The first is if ε is sufficiently small such that the limit is approximately correct and we have enough data to get an accurate measurement for $H_q(\varepsilon)$. In that case we may use

$$D(q) = D_q(\varepsilon) = -H_q(\varepsilon)/\ln \varepsilon. \tag{9}$$

However, the preferred method is to simply look at the slope of a plot of $H_q(\varepsilon)$ vs $\ln(\varepsilon)$ since this has a quicker rate of converge to the limit of $\varepsilon \to 0$. We should mention that further information on theoretic properties can be determined from the analysis of this sorting, such as the generalized mutual information of high dimensional data, $I_n(X_1, X_2, \ldots, X_n)$, or the estimation of the metric entropy.[28]

For a large box size, the box counting dimension varies widely from the others since the box counting dimension $D(0)$ is more susceptible to errors. It is also a poor quantity to use since it says nothing about the density of the attractor, only about its shape. However, the box counting dimension and all the others converge in the midregion before diverging slightly and then dropping to zero (due to data set size). It is this midregion that parallels the plateau region of the Grassberger–Proccacia algorithm.[26,27]

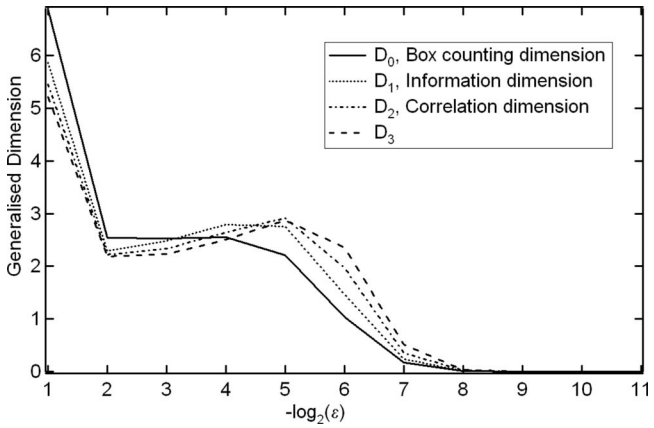


FIG. 14. Plot of the first four generalized dimensions from the original data.

Figures 14 and 15 present the results of our calculations of the first four generalized dimensions performed on the flow data embedded in five dimensions and the sectioned data embedded in four dimensions, respectively. Displayed are estimates of the first four generalized entropies for varying box sizes. Additional tests were also performed for the embedding dimensions of 3–6 and for the next four generalized entropies. The results indicated that for $p>q$, $D(p) \leqslant D(q)$, which agrees with the theory.

The estimates for fractal dimension were derived from where the slope of of $H_q(\varepsilon)$ vs $\ln(\varepsilon)$ showed the least deviation for successive values of ε ($\varepsilon=2^{-4}$ and 2^{-5}). Estimates ranged from 1.8 to 2.2 for all fractal dimensions calculated on the sectioned data and from 2.6 to 3.0 for all fractal dimensions calculated on the flow data when the embedding dimension was greater than or equal to 4. In both cases the estimate agrees with our choice of the embedding dimension and is also in rough agreement with the result from the Grassberger–Proccacia algorithm. However, in general, the dimensionality of the sectioned data is less than 1 plus the dimensionality of the flow data. This may be accounted for primarily by the small data set size for the sectioned data (approximately 10 000 points), which is known to cause a slight overestimation of fractal dimension, and the relatively high sample rate of the flow data, which skewed estimates downward.[29]

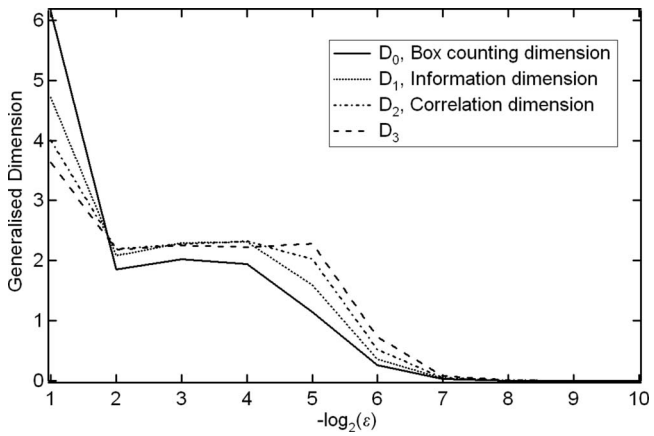


FIG. 15. Plot of the first four generalized dimensions from the sectioned data.

E. Lyapunov exponents

The Lyapunov exponents characterize how chaotic a system is. For a D dimensional dynamical system, consider the infinitesimally small D-sphere centered around a point on the attractor. As this evolves, the sphere will become a D-ellipsoid due to the deforming nature of the attractor flow. Then the ith Lyapunov exponent is defined in terms of the exponential growth rate of the ith principal axis of the ellipsoid,

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \ln \frac{p_i(t)}{p_i(0)}. \tag{10}$$

Thus the spectrum of Lyapunov exponents, $\{\lambda_1, \lambda_2, \ldots, \lambda_D\}$ describes the rate of growth of the distance between nearby trajectories in phase space. Chaos is often defined by the existence of a positive dominant Lyapunov exponent, which indicates that nearby trajectories will, over time, diverge exponentially away from each other.

The determination of Lyapunov exponents from noisy experimental data is a difficult task. Although many methods have been presented, there are also numerous examples of these methods breaking down when tested against real data, as well as questions concerning the validity of the methods. Thus the results of exponent determination were held to scrutiny. Criteria were established for the identification of Lyapunov exponents.

There should be an agreement between exponents as measured from different algorithms, and some measurement of error should be provided. Embedding parameters used in estimating exponents must be confirmed independently by other methods. The results should remain consistent under various parameter settings for exponent estimation. For flow data, a zero exponent should be clearly found. The estimation of exponents from both flow and sectioned data should be in agreement. The sum of all the exponents must be negative, and the sum of the positive exponents should be less than or equal to the metric entropy. In fact, for many cases they should be equal.[30] Under the proper conditions, the Lyapunov exponents should all, approximately, switch sign when measured from a time reversal of the data.[31]

The Lyapunov dimension may be defined as

$$D_\lambda = L + \frac{1}{|\lambda_{L+1}|} \sum_{j=1}^{L} \lambda_j, \tag{11}$$

where L is the maximum integer such that the sum of the L largest exponents is still non-negative. That is, $\Sigma_{j=1}^{L}\lambda_j \geqslant 0$ and $\Sigma_{j=1}^{L+1}\lambda_j < 0$. The Kaplan–Yorke conjecture[32] proposes that this is equal to the information dimension. Within error bounds, this seems to be true. Therefore, a final criterion is that the Lyapunov dimension estimates should agree with the information dimension estimates.

It is doubtful that all criteria can be satisfied unless one is dealing with a long noise-free time series of low dimensional simulated data. Noise, high dimensionality, and short time series length (few orbits or a small number of points or both) negatively affect all methods of analysis. Some criteria, such as the confirmation of embedding parameter choices, are a virtual necessity before any calculation is made. Others,

TABLE I. Results of the estimation of the Lyapunov exponent(s) using three different techniques. All calculations were performed with a four dimensional embedding of the peak value data.

Algorithm	λ_1	λ_2	λ_3	λ_4	Sum
Rosenstein *et al.*	0.403				
Wolf *et al.*	0.391				
Eckmann–Ruelle	0.380	0.054	−0.184	−0.484	−0.234

such as the agreement between the Lyapunov and information dimension, are very strong indicators that Lyapunov exponents have been reasonably determined. Still other criteria require calculations of quantities that are particularly difficult to compute from time series, such as metric entropy. Previous authors chose to reject the use of estimated Lyapunov exponents as discriminating statistics.[33]

Three methods of determining Lyapunov exponents were implemented: the method of Eckmann and Ruelle[34] for determining the Lyapunov spectra and the methods of Rosenstein *et al.*[35] and Wolf *et al.*[36] for determining the largest exponent. Since these methods are fundamentally different, one would not expect an agreement between the estimates to be simply due to them incorporating the same mistakes. The sectioned data were used for all estimates since these reduce dependence on the choice of delay time. Abarbanel's method was also applied,[37] although this is based on the same technique as Eckmann and Ruelle's method and the sectioned data provide very similar results.

Wolf *et al.*[36] defined the exponents in terms of the rates of divergence of volumes as they evolve around the attractor. The method of Wolf *et al.* involves following a trajectory in its path around the attractor. The rate of growth between points on this trajectory and a nearby trajectory is used to estimate the largest Lyapunov exponent. When the distance between these trajectories becomes too large, then a new nearby trajectory is found that is within a reasonable angular distance from the previous nearby trajectory. In order to minimize parameter dependence, we used a variation on the method of Wolf *et al.*[29] In previous work,[36,38,39] both maximum allowable displacement and maximum angular displacement were left as free parameters. We use one parameter: the number of near neighbors to consider after a given number of iteration steps.

The method of Rosenstein *et al.* involves looking at average divergence rates of nearest neighbors. It also finds the dominant exponent. The Eckmann and Ruelle method involves using a small neighborhood of points and iterating them forward to estimate the local Jacobian, and then determining the Lyapunov spectrum from the eigenvalues of the Jacobians around the attractor.

In Table I, results of exponent calculations are provided. All calculations were performed with a time delay of 1 and an embedding dimension of 4, as suggested by the FNN routine. The exponents are given in units of l/time, where the time scale is defined so that the time between samples is 1. Many more calculations were performed until a reasonable and stable parameter regime was found for all methods. In general, exponent calculations converged to within 5% of their final value when averaging local estimates of the dominant Lyapunov exponent over only 1000 points (although the entire sectioned data set was used to find near neighbors).

Several of our criteria are determined immediately upon inspection. The dominant exponent results from all three methods provide a rough agreement. One check on the algorithm of Wolf *et al.* was calculating the average angular displacement. This was typically less than 20%, well within reasonable bounds. For the Lyapunov spectrum, the second exponent is very close to zero and the sum of the exponents is negative.

The folding of the attractor brings diverging orbits back together. So any effects of nonlinearities will most likely serve to move all exponents closer to zero. Also increasing the number of near neighbors used may underestimate the value because this allows a larger distance between neighbors. Hence a slight underestimate of the positive exponents for the Eckmann–Ruelle algorithm (and for Abarbanel's technique[37]) was expected. For the algorithm of Wolf *et al.*, the angular displacement errors are not likely to accumulate, but each error may skew the largest positive exponent downward. These assumptions are confirmed by the slightly lower estimate of the dominant exponent for the Eckmann–Ruelle and Wolf algorithms as compared to the technique of Rosenstein *et al.* which is less susceptible to these errors.

However, it was not possible to confirm all criteria. The measurement of the metric entropy is still ongoing. Sectioning the data introduced additional noise. More importantly, the uncertainty in sample values tended to dominate over the divergence on a small time scale, thus introducing errors into the measurement of Lyapunov exponents from the original flow data. Thus it was not possible to get an agreement between exponent estimates from the section and from the flow, nor was it expected. For the measurement of the Lyapunov spectrum, uncertainty in the values of other exponents meant that it was not possible to get a reliable estimate of the Lyapunov dimension. Time reversal results were also inconclusive at best. However, simulated data with the addition of noise would not usually switch the signs of the exponents under time reversal either. So the sign change of exponents when the data are reversed may not be a suitable criterion for noisy data.

F. Unstable periodic orbits

In addition to the occurrence of a positive Lyapunov exponent, a chaotic system may also be characterized by having an infinite number of unstable periodic orbits (UPOs). The identification of UPOs plays a critical role in many chaos control algorithms.[40] Most standard chaos control algorithms attempt to control the system onto a UPO while operating within the chaotic regime. Small time-dependent perturbations applied to an accessible parameter may then be used to force the system onto the stable manifold and hence enforce stability and a periodic behavior. The drive frequency is the most preferable candidate to use as the varied parameter in a control scheme. This is because it is easily adjustable, and a small change in drive frequency often yields appropriate changes in the dynamics. It is useful in

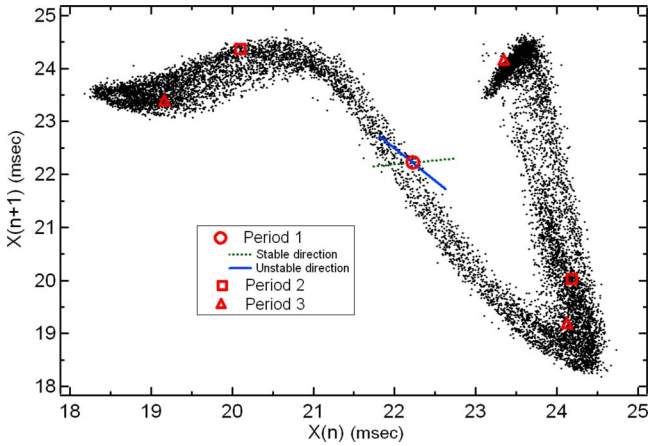


FIG. 16. (Color online) A delay coordinate embedding plot of the times between zero crossings. Periods 1–3 points are identified, along with the stable and unstable manifolds of the period 1 orbit.

occasional proportional feedback control schemes in tracking and targeting of trajectories, and in the identification of symbolic dynamics.[40,41] Thus we will also attempt to characterize the UPOs exhibited during the chaotic state.

By looking for when the dynamics approach the same region after a given number of iterates, periodic orbits can be found. UPOs of period p are found simply by establishing a threshold ε,

$$\|Y(n) - Y(n+p)\| < \varepsilon, \tag{12}$$

in which case a periodic orbit exists in the vicinity of $Y(n)$ and a least squares fit of all data in the region can be used to estimate its exact location. The exact value of ε may be varied depending on the size of the data set, the period p, and the number of UPOs that one wishes to find. False positive UPOs may be identified using the mean squared error of the least squares fit.

Figure 16 shows the delay coordinate embedding of the Poincaré section using times between zero crossings. In this figure, we identified period 1, period 2, and period 3 orbits. From the least squares fit estimate of the local dynamics, the eigenvalues and eigenvectors corresponding to the stable and unstable manifolds can be found. For the period 1 orbit, we have a fixed point located at 0.0222 s, corresponding to the drive frequency of the system, 45 Hz. Its eigenvalues are 0.155 and −1.571, with corresponding eigenvectors (1,0.155) and (−0.637,1). Similar results can be obtained for other identified periodic orbits, and these results can be used to implement a chaos control technique.

IV. FURTHER ANALYSIS AND CONCLUSIONS

An analysis was attempted on time series data from an experimental electrodynamic loudspeaker in order to characterize the embedding dimension, the fractal dimension, the Lyapunov exponents, and the UPOs. The obtained results indicate that the system is governed by low dimensional chaotic dynamics and thus is highly amenable to control, tracking, synchronization, noise reduction, and so forth. Particular care was made in verifying the presence of a positive Lyapunov exponent.

Various estimates of fractal dimension were performed, including the measurement of correlation dimension,[26,27] information dimension, and box counting dimension for different embedding dimensions. Although there was qualitative confirmation of low dimensional behavior, consistent results for quantitative values for the fractal dimension were not achieved.

However, the estimation of the dominant Lyapunov exponent, which is less reliant on a large data set size, provided consistent results regardless of the method of estimation.[10] We used several different techniques that provided a rough agreement in their estimates of the dominant exponent, and the results further agreed with the theory concerning the Lyapunov spectrum and its properties. They reliably showed evidence of a positive Lyapunov exponent, a strong indicator of chaos.

Finally, we attempted to estimate the eigenvalues and eigenvectors associated with detected UPOs. These may be easily identified. Control may be applied to allow the loudspeaker to operate as desired within the chaotic regime. Tracking and maintenance should also be possible since the appropriate dynamics have been found for the application of several well-known algorithms.

However, extraction of many empirical quantities, particularly fractal dimension, proved difficult due to a number of issues. This may be accounted for partly by the nonstationarity and short data set size. Although the original data set is over 200 000 points, it represents about 10 000 orbits. This is insufficient for an effective calculation of fractal dimension in the presence of noise and long term dynamics. Furthermore, although the dynamics are somewhat stable, there is still a gradual change in various statistical quantities when examined using a sliding window through the data. This tends to distort various measurements. A longer data set size could capture the long term dynamics.

However, we believe that the difficulty in estimating some measures of nonlinear behavior and dimensionality is primarily due to low sample rates and low resolution due to limitations in the data acquisition system. First, the data were sampled at 1024 samples/s, or approximately 23 samples per period. A higher sampling rate would yield more accurate Poincaré sections. Alternatively, if the sampling data could be sampled at exactly the drive frequency, 45 Hz, then this would produce a natural Poincaré section. The analyzed data had a 16 bit precision, but this was further limited by the 8 bit resolution of the A/D converter of the laser distance meter. This meant that there was a significant uncertainty in the location of nearby points. Since most analyses of chaotic time series rely on an analysis of near neighbors in a locally linear region, these resulted in inaccuracies in the estimation of fractal dimension and Lyapunov exponents.

Current work is focused on improvements to the data acquisition system. This would allow a more accurate analysis of the data, including measurements of how fractal and Lyapunov dimensions change with parameter settings, and use of chaotic time series prediction methods on the data. This could also be used for further direct comparison with a dynamical behavior from enhanced models of the loudspeaker. An accurate model of an electrodynamic loud-

speaker would represent a significant advance in the field, particularly since model parameters could then be modified to yield a loudspeaker design with optimal performance.

[1]P. Klipsch, "Loudspeaker distortion," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '76, Philadelphia, PA, pp. 543–546.
[2]D. Davis and E. Patronis, *Sound System Engineering*, 3rd ed. (Elsevier, New York, 2006).
[3]E. S. Olsen and K. Thorbog, "Diaphragm area and mass nonlinearities of cone loudspeakers," Proceedings of the 99th AES Convention, New York, 1995, Convention Paper No. 4082.
[4]M. Colloms, *High Performance Loudspeakers*, 6th ed (Wiley, New York, 2005).
[5]E. S. Olsen, "Nonlinear modeling of low frequency loudspeakers: A practical implementation," Proceedings of the 102nd AES Convention, Munich, Germany, 1997, Convention Paper No. 4469.
[6]P. Woafo, "Harmonic oscillations, stability and chaos control in a nonlinear electromechanical system," J. Sound Vib. **259**, 1253–1264 (2003).
[7]P. Woafo, "Stability and chaos control in electrostatic transducers," Phys. Scr. **62**, 255–260 (2000).
[8]A. Steindl and H. Troger, "Chaotic oscillations of a fluid-conveying viscoelastic tube," in *Nonlinearity and Chaos in Engineering Dynamics*, edited by J. M. T. Thompson and S. R. Bishop (Wiley, Chichester, 1994), pp. 231–240.
[9]J. Argyris, V. Belubekyan, N. Ovakimyan, and M. Minasyan, "Chaotic vibrations of a nonlinear viscoelastic beam," Chaos, Solitons Fractals **7**, 151–163 (1996).
[10]A. Petosic, I. Djurek, and D. Djurek, "Modeling of an electrodynamic loudspeaker using Runge-Kutta ODE solver," 122nd Convention of Audio Engineering Society, Vienna, Austria, 2007.
[11]T. Leishman and G. Dix, "Additional modeling details for the suspension of a moving-coil loudspeaker," J. Acoust. Soc. Am. **116**(4), 2619 (2004).
[12]R. A. Ribeiro, A. J. Serralheiro, and M. S. Piedade, "Application of Kalman and RLS adaptive algorithms to non-linear loudspeaker controller parameter estimation: a case study," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, 2005.
[13]M. D. Fränken and J. Wassmuth, "Passive parametric modelling of dynamic loudspeakers," IEEE Trans. Speech Audio Process. **9**, 885–891 (2001).
[14]R. J. Wei, Q. T. Tao, and W. S. Ni, "Bifurcation and chaos of direct radiation loudspeaker," Chin. Phys. Lett. **3**, 469–472 (1986).
[15]D. Djurek, I. Djurek, and A. Petosic, "Intrinsic membrane friction and onset of chaos in an electrodynamic loudspeaker," Proceedings of the 123rd Aes Convention, New York, NY, 2007.
[16]I. Djurek, A. Petosic, and D. Djurek, "Chaotic state in an electrodynamic loudspeaker controlled by gas pressure," Proceedings of the 153rd ASA Meeting, Salt Lake City, UT, 2007.
[17]D. Djurek, I. Djurek, and A. Petosic, "Chaotic state in an electrodynamic loudspeaker," Proceedings of the 122nd Convention of the AES, Vienna, Austria, 2007.
[18]H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series (Cambridge University Press, Cambridge, England, 2004).
[19]I. Djurek, D. Djurek, and A. Petosic, "Damping of an electrodynamic loudspeaker by air viscosity and turbulence," Proceedings of the 123rd Convention of Audio Engineering Society, New York, NY, 2007.
[20]T. Y. Li and J. A. Yorke, "Period three implies chaos," Am. Math. Monthly **82**, 985–992 (1975).
[21]M. J. Feigenbaum, "The universal metric properties of nonlinear transformations," J. Stat. Phys. **21**, 669–706 (1979).
[22]A. M. Fraser, "Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria," Physica D **34**, 391–404 (1989).
[23]A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**, 1134–1140 (1986).
[24]J. D. Reiss, N. Mitianoudis, and M. B. Sandler, "Computation of generalized mutual information from multichannel audio data," Proceedings of the Audio Engineering Society 110th Convention, Amsterdam, The Netherlands, 2001.
[25]M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," Phys. Rev. A **45**, 3403–3411 (1992).
[26]P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors.," Physica D **9**, 189–208 (1983).
[27]P. Grassberger and I. Procaccia, "On the characterization of strange attractors," Phys. Rev. Lett. **50**, 346–349 (1983).
[28]A. M. Fraser, "Information and entropy in strange attractors," IEEE Trans. Inf. Theory **35**, 245–262 (1989).
[29]J. D. Reiss, "The analysis of chaotic time series," Ph.D. thesis, School of Physics, Atlanta, Georgia Institute of Technology, 2001, pp. 250.
[30]J. P. Crutchfield and N. H. Packard, "Symbolic dynamics of noisy chaos," Physica D **7**, 201 (1983).
[31]U. Parlitz, "Identification of true and spurious Lyapunov exponents from time series," Int. J. Bifurcation Chaos Appl. Sci. Eng. **2**, 155–165 (1992).
[32]P. Constantin and C. Foias, "Global Lyapunov exponents, Kaplan-Yorke formulas and the dimension of attractors for 2D Navier-Stokes equations," Commun. Pure Appl. Math. **38**, 1–27 (1985).
[33]J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," Physica D **58**, 77–94 (1992).
[34]J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Liapunov exponents from time series," Phys. Rev. A **34**, 4971–4979 (1986).
[35]M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," Physica D **65**, 117–134 (1993).
[36]A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," Physica D **16**, 285–317 (1985).
[37]H. D. I. Abarbanel, "Tools for the analysis of chaotic data," in *Nonlinear Dynamics and Time Series, Fields Institute Communications*, edited by C. D. Cutler and D. T. Kaplan (American Math. Soc., Providence, RI, 1997), pp. 1–16.
[38]A. Wolf, "Quantifying chaos with Lyapunov exponents," in *Chaos*, edited by A. V. Holden (Manchester University Press, Manchester, 1986).
[39]A. Wolf and J. Vastano, "Intermediate length scale effects in Lyapunov exponent estimation," in *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress (Springer-Verlag, Berlin, 1986).
[40]E. Schöll and H. G. Schuster, *Handbook of Chaos Control*, 2nd ed. (Wiley, New York, 2007), pp. 849.
[41]K. Mischaikow, M. Mrozek, J. Reiss, and A. Szymczak, "Construction of symbolic dynamics from experimental time series," Phys. Rev. Lett. **82**, 1144–1147 (1999).

A quantitative analysis of signal reproduction from cylinder recordings measured via noncontact full surface mapping

Antony Nascè, Martyn Hill,[a] John W. McBride, and Peter J. Boltryk
Electromechanical Research Group, University of Southampton, Hampshire SO17 1BJ, United Kingdom

(Received 28 March 2008; revised 24 July 2008; accepted 26 July 2008)

Sound reproduction via a noncontact surface mapping technique has great potential for sound archives, aiming to digitize content from early sound recordings such as wax cylinders, which may otherwise be "unplayable" with a stylus. If the noncontact techniques are to be considered a viable solution for sound archivists, a method for quantifying the quality of the reproduced signal needs to be developed. In this study, a specially produced test cylinder recording, encoded with sinusoids, provides the basis for the first quantitative analysis of signal reproduction from the noncontact full surface mapping method. The sampling and resolution of the measurement system are considered with respect to the requirements for digital archiving of cylinder recordings. Two different methods of audio signal estimation from a discrete groove cross section are described and rated in terms of signal-to-noise ratio and total harmonic distortion. Noncontact and stylus methods of sound reproduction are then compared using the same test cylinder. It is shown that noncontact methods appear to have distinct advantages over stylus reproduction, in terms of reduced harmonic distortion and lower frequency modulation. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2973238]

PACS number(s): 43.38.Ne, 43.38.Md, 43.58.Ry, 43.60.Dh [AJZ] Pages: 2042–2052

I. INTRODUCTION

Sound reproduction from early mechanical recordings via noncontact optical methods has been the topic of increasing interest in recent years.[1–13] A noncontact method ensures that no further damage or pressure is applied to the often fragile grooves, wherein the sound was originally encoded. In general, two noncontact reproduction strategies have been proposed in literature: (i) signal reproduction, *in situ*;[1–4] and (ii) surface mapping via optical sensors, followed by postprocessing.[5–13]

In (i), accurate sound reproduction requires robust tracking of the grooves with a laser[1–3] or an interferometer system,[4] *in situ*. This strategy can be problematic in cases where the recording is damaged or deformed. Furthermore, only a single audio signal estimate can be obtained from each transfer attempt.

An alternative transfer strategy (ii) uses optical sensing to map the full surface of cylinder recordings. The philosophy behind this technique is aimed at preservation of the full recorded surface, not real-time playback. Most of these approaches have stored a two-dimensional (2D) image of the surface, but recent work[9–12] has measured the full three-dimensional (3D) surface topology. The advantage of mapping the full surface topology is that sound can be extracted from anywhere within the groove, any number of times, at any point in the future. Tracking of the grooves in postprocessing can therefore be more accurate than *in situ*, especially in cases where the surface is highly deformed or damaged. Furthermore, this technique has potential for recovering sound from broken artefacts, by scanning fragments of the recording, and then stitching data sets together.

In previous work[10] based on (ii), signal reproduction from cylinder recordings has been judged qualitatively to provide "faithful playback of the recorded information." In this study, we seek a method of quantifying the signal quality, in terms of metrics such as signal-to-noise ratio (SNR) and total harmonic distortion (THD). This is achieved through the use of a specially produced cylinder, encoded with sinusoids. In addition, the concept that multiple audio signal estimates can be derived from the same dataset is investigated, by comparing signal reproduction from two groove depth estimation methods. The noncontact signal reproduction described in this paper considers cylinder recordings only, but similar methods can also be applied to other early recordings such as phonograph tinfoil recordings and vertically cut disks.

This study is organized in the following way. Following a brief summary of cylinder recordings, the noncontact full surface mapping technique is described, with emphasis on sampling requirements for audio signal reproduction. In Sec. II, a review of previous work relating to signal reproduction and quality assessment from noncontact mapping methods is given. Section III gives details of the sound reproduction process and signal quality testing methods. The results are discussed in Sec. IV, and conclusions are given in Sec. V.

A. Cylinder recordings

Cylinder recordings represent the earliest mechanical means for archiving and playing back sound. Recording systems and cylinder technology developed over many years, but the basic recording principles remained the same (see Fig. 1). In recording mode, a cylinder "blank" is set to rotate with constant angular velocity, $\dot{\theta}$. A cutting tool (stylus),

[a]Author to whom correspondence should be addressed. Electronic mail: m.hill@soton.ac.uk

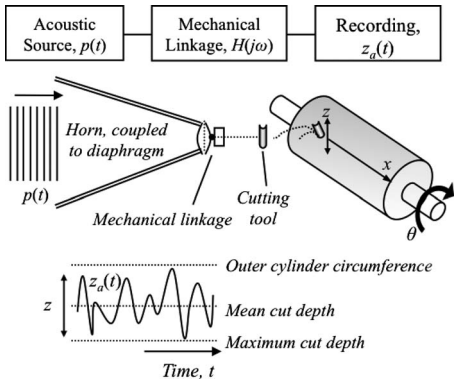


FIG. 1. Overview of the recording process for an acoustically recorded cylinder.

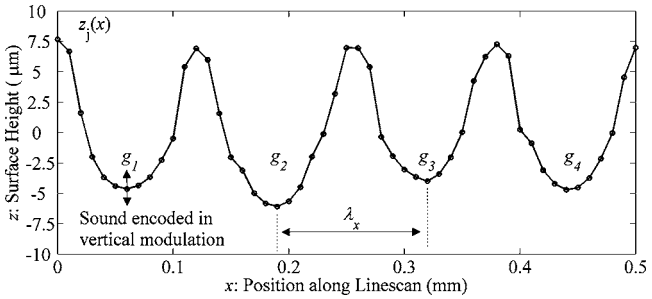


FIG. 3. A typical linescan of length $L=0.5$ mm, measured from a Blue Amberol cylinder, ca. 1912 ($\Delta x=10$ μm). This cylinder was recorded with 200 turns/in. and has a groove spacing λ_x of 127 μm. The *groove seed*, denoted by g_n is the nth groove cross section along the linescan.

which is coupled to a diaphragm-horn arrangement, advances along the cylinder axis (x-axis), via a lead screw with constant pitch, λ_x. Prior to recording, the cutting stylus is forced into the outer cylinder surface, producing an unmodulated cut of fixed radial depth. During recording, incident acoustic pressure variations are channeled via a horn from the surrounding air, toward a thin diaphragm, forcing the stylus to cut undulations that are normal to the surface (in the radial plane) about its mean cut depth. This radial modulation, $z_a(t)$, is often referred to as "hill-and-dale" or "vertical," as opposed to the majority of monophonic flat disk recordings, which are encoded with a "side-to-side" or "lateral" modulation. The encoded groove follows a helix of pitch λ_x, with cross section dependent on the stylus tip geometry. Cylinder recordings were typically recorded with a pitch of 100–200 turns/in. at a rotational speed of 100–160 rpm. For the interested reader, a history of the acoustic era of recorded sound is given by Millard[14] and Read and Welch.[15]

Modern electrical reproduction of cylinder recordings is achieved by tracing over the surface modulations with a contacting stylus. The stylus transducer converts the physical hill-and-dale modulations into an electrical signal. Sound reproduction can also be achieved by mapping the full cylinder surface via a noncontact optical sensor and then estimating $z_a(t)$ in postprocessing.

B. Cylinder surface measurement and definitions

Figure 2 shows a cylinder recording mounted onto a noncontact measurement system. The methodology, which is typical of the approaches described in literature[9–12] uses a single-point optical sensor to measure the full surface topology of the cylinder recording. A motion system is used to traverse the sensor linearly along the cylinder's axis, forming a "linescan." The cylinder is mounted on a rotatable mandrel, allowing the sensor to map the continuous cylinder surface $z(x,\theta)$ via a succession of linescan measurements. An additional linear stage provides a third axis of motion orthogonal to the linescan axis, which controls the sensor's standoff distance from the artefact to keep the surface within the sensor's gauge range.

The discrete grid over which the cylinder is mapped is defined by the two spatial increments, Δx (in microns) and $\Delta\theta$ (in degrees), which are the linescan and rotational increments of the motion system, respectively. We denote the discrete surface height matrix by $z(x_i,\theta_j)$. The jth linescan, denoted by $z_j(x)$, is the linear scan measured along the x-axis at rotational index j. A typical linescan and surface map from a Blue Amberol cylinder is shown in Figs. 3 and 4.

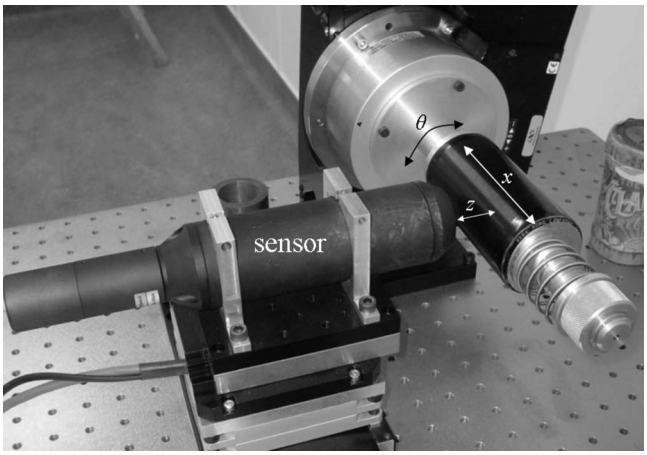


FIG. 2. Cylinder recording mounted on the measurement system, showing the three measurement axes.

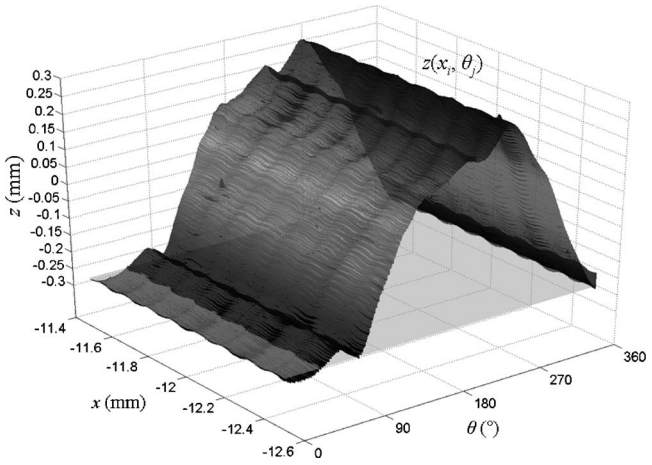


FIG. 4. Surface measurement of a Blue Amberol cylinder surface. The surface is built up through a succession of linescans. The large deviation in z is due to the lack of cylinder concentricity.

C. Resolution requirements for sound reproduction

To ensure that sound can be accurately recovered from the cylinder, sampling resolution of the measurement system must be considered to avoid aliasing and to optimize scanning time.

The choice of Δx determines the number of samples over which the groove cross section is resolved along the linescan. This is dependent on the cylinder pitch and complexity of the groove cross section. For a cylinder recorded with 200 turns/in., a linescan grid spacing of $\Delta x = 10\ \mu$m gives approximately 13 samples per groove cross section (see Fig. 3, for example). This sampling has been previously judged to provide sufficient resolution for sound reproduction.[10]

The choice of $\Delta\theta$ (in degrees) determines the playback sample rate f_s of the recovered audio signal. Assuming that the cylinder was recorded with a constant rotational speed of ω (in rpm), the estimated playback sample rate is given by

$$f_s = \frac{6\omega}{\Delta\theta}. \tag{1}$$

To avoid aliasing of the signal by the Nyquist criterion, we have

$$f_s > 2f_{\max}, \tag{2}$$

where $f_{\max}$ is the maximum frequency limit of the cylinder record. For cylinder recordings, this upper limit is around 5 kHz.[10,16] For 160 rpm cylinders, an angular sampling increment of $\Delta\theta = 0.1°$ gives a playback sample rate of 9.6 kHz, which should be considered as an absolute minimum for accurate sound reproduction. The technical committee of the International Association of Sound and Audiovisual Archives (IASA) recommends a minimum sample rate of 48 kHz at 24 bit word length for digitizing analog recordings.[17] In some cases, a sampling of 96 kHz at 24 bits has been adopted.[17] With the noncontact system, a playback sample rate of 96 kHz is achievable with an angular sampling of $\Delta\theta = 0.01°$ (for cylinders recorded at 160 rpm). The datasets discussed in this paper use a sampling increment of $\Delta\theta = 0.01°$, giving $f_s = 96$ kHz.

The third sampling dimension is the axial resolution of the sensor, Δz, which is the ability of the sensor to resolve displacements in the z-plane. It was shown by Boltryk *et al.*[2] that in order to resolve the smallest displacement amplitudes (high frequency and low amplitude) for typical cylinder recordings, an axial resolution in the order of 10 nm is required. For vertically modulated grooves, Δz determines the dynamic range of the digital audio signal. Current hardware permits a *measurement dynamic range* of ~15 bits (3.5×10^4 discrete z-values) over the sensor's full gauge range. However, in order to relate Δz to the dynamic range of the digital audio signal, the range over which the groove depth modulates in the z-axis must be known. A typical range value for the test cylinder described in this study is 10 μm (for a 1 kHz tone, measured from peak to trough). The current sensor has an axial resolution of $\Delta z = 10$ nm, which in this case gives an equivalent digital audio bit depth of just under 10 bits (1000 discrete z-values).

II. REVIEW OF RELATED TECHNIQUES

A. Sound recovery from full surface mapping of cylinder recordings

Sound reproduction from full surface mapping of cylinder recordings presents a relatively recent area of research. There are currently no commonly used or universally approved methods for audio signal estimation from discrete surface maps of cylinders in literature.

In a previous collaborative study,[10] between Lawrence Berkeley National Laboratory and the University of Southampton, a method was described for extracting sound from a discrete surface map of a Blue Amberol cylinder (ca. 1912), measured via single-point optical displacement sensor, as described in Sec. I B.

In Ref. 10, the radial displacement track encoded by the cutting tool $z_a(t)$ is found by initially estimating the local minimum height (minima) of each groove cross section for all linescans in the discrete surface, $z(x_i, \theta_j)$. Groove minima data are then reorganized into a time series, which follows a helical trajectory. A list of groove valleys (minima) is initially located via a neighborhood search along each linescan, and outlying candidates are removed. A set of points in the kth groove valley is then fitted to a quadratic function, $H_k(x) = A_k x^2 + B_k x + C_k$. The quadratic term ($A_k$) remains fixed and is chosen such that the curvature of the parabola is an approximation to that of the original cutting tool, which produced the recording. The depth (minimum ordinate) of the fitted quadratic curve is then used to form a radial displacement track estimate, $z_a(t)$, corresponding to the encoded sound. The concept of three stand-alone data streams were considered: (1) the groove bottom (valley), (2) the groove top (ridge), and (3) the groove bottom position with respect to the groove top position (top-bottom).

It was noted that the use of the ridge data stream provided audible albeit noisy sound content, but with interference from different times in the recording. This kind of temporal distortion was also noted for playback systems *in situ*[1,2] when the width of the laser beam spot was large compared with the groove width. The use of "top-bottom" subtraction[10] was shown to remove low frequency structures (below 150 Hz) due to surface form.

The audio signal reproduced by Fadeyev *et al.*[10] was compared with a stylus transfer of the same recording. Visual comparisons of the time histories and spectra showed that the optical method provided an "accurate audio transcription," compared with the stylus playback, but no quantitative analysis of the audio signal quality was presented. In this paper, we seek to evaluate signal reproduction of the noncontact system for cylinders using quantitative measures. This allows for different audio signal estimates to be rated and compared numerically without subjectivity.

B. Signal quality testing for noncontact sound reproduction

An audio signal quality analysis was developed by Stotzer[13] for their noncontact reproduction method. This analysis was carried out to determine the feasibility of photographic methods of sound reproduction for flat disk record-

ings and to make comparisons with stylus playback. The SNR and THD are determined through the use of test records encoded with single frequency tones. In this way, the recovered signal is of known type; thus, the SNR and THD are calculated by estimating the ratio of powers in the signal and noise bands.

A summary of discrete-time methods for estimating the signal-to-noise power ratio estimates is given by Jenq.[18] Two algorithms are described for estimating the SNR of a noisy sinusoid from discrete-time data. The first is based on estimating directly the four parameters of a sine wave (amplitude, frequency, phase, and dc offset), and the second is based on a spectrum averaging method via the discrete Fourier transform (DFT).

Estimates for SNR using spectrum averaging techniques are given by Stotzer[13] for various test disks recorded with a 300 Hz tone. The SNR is defined as the ratio between the signal power P_{sig} and noise power P_n for tracks containing a single frequency, f_0. The SNR is expressed in decibels and is given by

$$\text{SNR} = 10 \log_{10}\left(\frac{P_{\text{sig}}}{P_n}\right). \tag{3}$$

The signal and noise powers are derived from discrete power spectral estimates of the recovered tonal waveform, where P_{sig} is calculated over a frequency range of $\pm f_{\text{margin}}$ around the center frequency, f_0. (In Ref. 13, $f_{\text{margin}}=5$ Hz.) The noise power P_n is the power in the remaining frequency bins, outside of the range $f_0 \pm f_{\text{margin}}$, from 100 Hz up to 10 kHz (these limits were chosen arbitrarily).

Another method of measuring the quality of sinusoidal signals is to examine the THD. The spectrum of an ideally recovered sinusoid should contain a single peak at the fundamental frequency. In practice, geometric distortion of the sine wave can be introduced at time of recording or at time of measurement. Harmonic distortion can also be introduced by errors in the signal recovery method itself and this effect must be minimized. Stotzer[13] gave the following definition for THD:

$$\text{THD} = 10 \log_{10}\left(\frac{P_{\text{harm}}}{P_{\text{sig}}}\right), \tag{4}$$

where P_{sig} is defined as in Eq. (3) and P_{harm} is the power of the harmonics $(2f_0, 3f_0, \ldots)$ up to 10 kHz. Harmonic peaks are also considered on a fixed frequency width of $\pm f_{\text{margin}}$ around the peak frequency with $f_{\text{margin}}=5$ Hz.

III. METHODOLOGY

A. Proposed methods of audio signal recovery

In previous work,[10] the estimation of stylus trajectory and groove depth was not considered independently but as a single process. For damaged cylinder surfaces, it is useful to visually assess the estimated trajectory before extracting the sound to ensure that the trajectory is coherent (no skipping of grooves occurs) and complete (no regions containing sound are missed). For this reason, we consider sound recovery in two independent stages—trajectory estimation and audio signal estimation—as was previously discussed.[11]

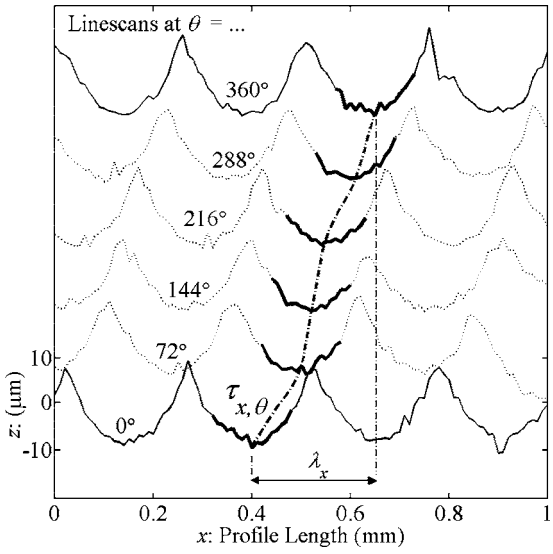


FIG. 5. Linescan measurements taken at six angular intervals from 0° to 360° around the cylinder circumference (linescans have been offset in the z-axis for visualization purposes). The path traced by $\boldsymbol{\tau}_{x,\theta}$ shows the shift along the x-axis of the highlighted groove valley. This shift vector remains approximately equal for all groove valleys and can be estimated by calculating the relative phase shift between adjacent linescans.

1. Trajectory estimation

A method of rapidly approximating the trajectory produced by the cutting tool was previously introduced,[11] by modeling the linescan as a periodic signal and by measuring the relative shift between linescans around the circumference. Reasons for adopting this strategy include the reduction in computation time and the ability to observe and account for any potential phase offsets between adjacent linescans caused by measurement error.

Through the recording process, the linescan has a periodic structure with adjacent grooves separated by a constant distance λ_x, corresponding to the groove pitch. In most cases, the groove spacing λ_x can be derived from historical records or by observation, but for completeness and records of unknown type, it can be found numerically by spectral analysis of the linescan.

The DFT of the jth linescan is given by the complex sequence $\tilde{Z}_j(k)$, where k denotes the kth spatial frequency bin. The fundamental spatial frequency of the linescan k_0 is related to λ_x via the reciprocal relationship of period and frequency and is determined by calculating the expectation for all linescans as follows:

$$k_0 = E[\max|\tilde{Z}_j(k_p)|], \quad j \in \theta{:}[0^\circ, 360^\circ], \tag{5}$$

where k_p is the range of possible frequency bins in which k_0 is valid, which for cylinder recordings of the era is between 3.9 and 7.9 mm^{-1} (for 200 and 100 turns/in., respectively).

Figure 5 shows that if we observe a single groove valley at linescan at $\theta=0°$ and track its shift along the x-axis around the cylinder circumference until linescan at $\theta=360°$, the x-position of the valley propagates through a distance of λ_x. For a cylinder recorded with constant pitch, this shift is ap-

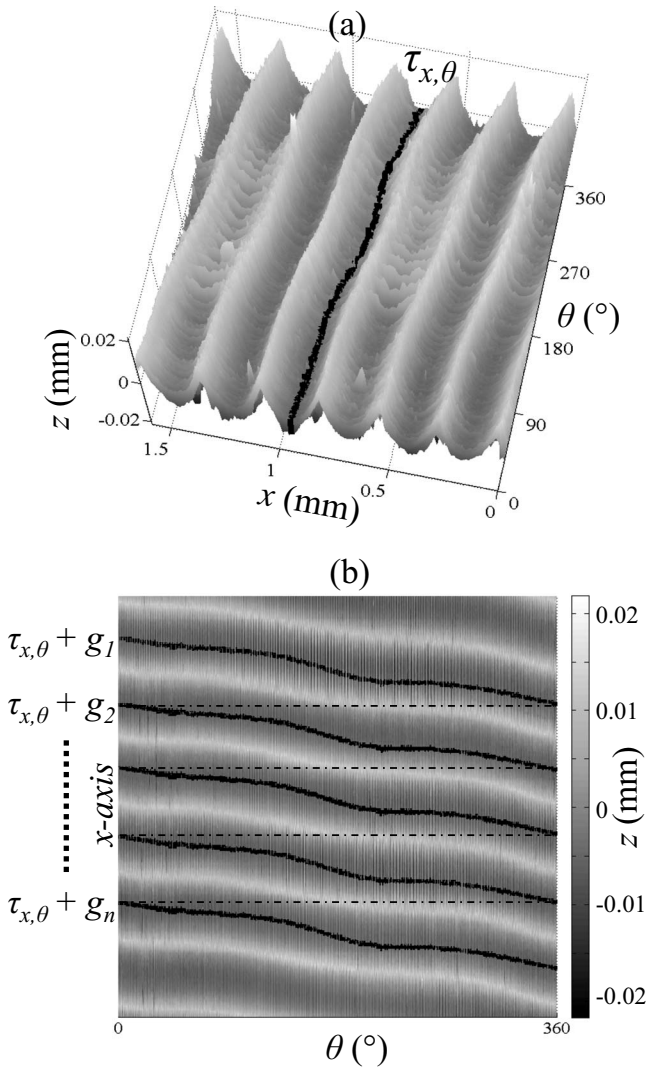


FIG. 6. (a) Scaled discretized shift estimation vector $\boldsymbol{\tau}_{x,\theta}$ overlaid onto a 3D surface plot of the test cylinder data. (b) Formation of stylus trajectory vector $A(x,\theta)$ using the phase shift estimate $\boldsymbol{\tau}_{x,\theta}$ and groove seed positions, $g_1, g_2, \ldots, g_n$.

proximately equal for all grooves. An estimate of this shift vector, $\boldsymbol{\tau}_{x,\theta}$, can therefore be used to form an approximate stylus trajectory.

The shift vector $\boldsymbol{\tau}_{x,\theta}$ can found by first estimating the relative phase (shift) between linescans from 0° to 360°. With k_0 determined from Eq. (6), the phase estimate of the jth linescan is given by

$$\varphi_j = \arctan\left[\frac{\mathrm{Im}\{\tilde{Z}_j(k_0)\}}{\mathrm{Re}\{\tilde{Z}_j(k_0)\}}\right], \quad j \in \theta{:}[0^\circ, 360^\circ]. \tag{6}$$

The result from Eq. (6) is given in radians; thus, in order to obtain $\boldsymbol{\tau}_{x,\theta}$, the phase angle vector φ_j must be scaled such that a phase shift of 2π rad is equal to the distance λ_x. It should be noted that an equivalent trajectory estimate may also be obtained using cross-correlation between linescans, but in this example, a Fourier-based method is described.

The resultant shift vector $\boldsymbol{\tau}_{x,\theta}$ is then rounded to the nearest grid position along the x-axis, so that the discrete surface data can be referenced. Figure 6(a) shows an example of $\boldsymbol{\tau}_{x,\theta}$ overlaid onto a discrete cylinder surface map.

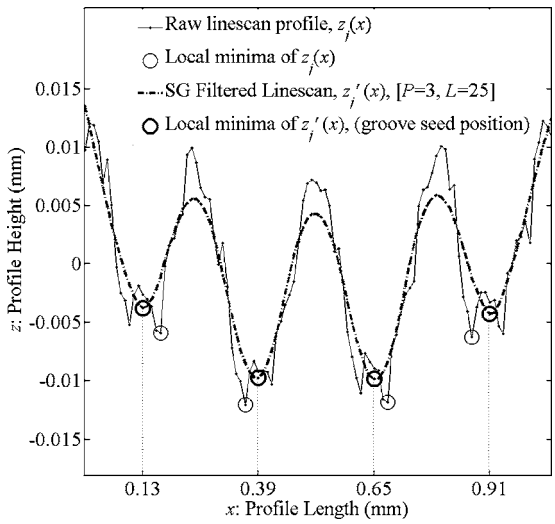


FIG. 7. The abscissa of local groove minima from the raw linescan profile, $z_j(x)$, varies from one groove to the next due to the roughness of the cutting tool. The abscissa of minima from the filtered linescan $z_j'(x)$ can be seen to be spaced equidistantly, by a distance of λ_x=0.254 mm, which is the expected spacing for a cylinder recorded at 100 turns/in. The local minima of $z_j'(x)$ form the groove seed positions, $g_1, g_2, \ldots, g_n$.

This shift vector $\boldsymbol{\tau}_{x,\theta}$ can then be used as a "template" for segmenting the surface into groove regions (groove seed). The nth groove seed, g_n (see Fig. 3), is found by locating the medial axis of each groove valley at 0°. For smooth symmetric groove cross sections, the medial axis is indexed by the local minimum point of the groove and can be found via a simple neighborhood search. In cases where the groove is rough and/or asymmetric (see Fig. 7, for example), the medial axis does not always lie at the local minimum point.

To ensure that each groove is indexed by its medial axis, an appropriately designed Savitzky–Golay (SG) polynomial smoothing filter is applied to the linescan $z_j(x)$ to form the filtered linescan $z_j'(x)$.

SG filters were initially developed to identify the relative widths and heights of spectral peaks in noisy spectrometric data.[19] In this work, the SG filter can be used in a similar way to obtain the position of groove valleys (to the authors' knowledge, this is the first time that SG filters have been applied to aid groove extraction). The SG filter works by a local polynomial regression (of degree P) on a distribution of length L (of at least $P+1$ equidistant samples) to determine the smoothed value for each point.

The frame length, L of the SG filter is chosen to match the groove pitch, λ_x. In this way, the SG filter operates locally on each groove valley, replacing the rough groove cross section with a smoothed version. Figure 7 shows a linescan filtered with a cubic SG filter (P=3) with frame length L =25 (the approximate number of samples across one groove cross section for 100 turns/in. cylinder/Δx=10 μm). The position of minima in the filtered linescan, $z_j'(x)$, are located closer to the middle (medial axis) of the groove valley than the minima found using the raw linescan, $z_j(x)$. The local minima of this SG filtered linescan are used to locate the groove seed positions.

A stylus trajectory vector, $\boldsymbol{A}=A(x,\theta)$, is then constructed [see Fig. 6(b)] by positioning $\boldsymbol{\tau}_{x,\theta}$ at the groove seed positions, $g_1, g_2, \ldots, g_n$ by

$$A(x,\theta)=\begin{bmatrix}[\tau_{x,\theta}+g_1]^T & [\theta]^T\\ \vdots & \vdots\\ [\tau_{x,\theta}+g_n]^T & [\theta]^T\end{bmatrix}, \tag{7}$$

where $\boldsymbol{A}$ is of size nJ rows by two columns, n is the number of complete grooves, and J is the number of linescans measured around the full cylinder circumference. The column vectors $[\tau_{x,\theta}+g_1]^T$ and $[\theta]^T$ are both of length J.

2. Audio signal estimation

Trajectory estimation provides an approximation to the helical path traced by a reproduction stylus. The radial displacement signal, $z_a(t)$, which corresponds to the encoded sound, is recovered by a discrete estimate of the groove depth along the trajectory vector $\boldsymbol{A}$, defined in Eq. (7).

Proposed schemes for deriving a discrete estimate for $z_a(t)$, include the following:

(i) single point data streams (using raw z-data, no averaging);
(ii) averaging samples across the groove cross section;[11]
(iii) stylus model, fitting based methods;[10] and
(iv) rule-based methods.

In previous work,[11] the notion of a "groove matrix" was introduced. The groove matrix, denoted by $\boldsymbol{G}_t$, is a matrix of z-values of size n by m, where n is the number of groove cross sections (equivalent to the number of samples in the time domain) and m is the number of samples across the groove cross section found along the trajectory vector $\boldsymbol{A}$.

Representing the groove depth by a single discrete data point (i) is not ideal for rough groove cross sections, as the selection of the sample is very sensitive to noise. Scheme (ii) was investigated previously by averaging data streams at the groove bottom and sidewall.[11] The cylinder in question was a molded Blue Amberol cylinder. In this case, the groove shape was symmetric and the abscissa of groove minima remained relatively unchanged from one groove to the next. A comparative examination of three different signal estimates in the time and frequency domains showed that the SNR increased as the estimate approached the bottom of the groove cross section.

The parabola fitting method[10] described in Sec. II A, based on scheme (iii), is well suited for cylinder recordings with a smooth symmetric groove cross section, as shown in Fig. 8(b), where the quadratic fit agrees well with the groove shape. In some cases, however [see Figs. 8(a) and 8(c), for example], the cylinder groove cross section is irregular, asymmetric, and does not resemble the quadratic fit. This irregularity is likely to have been caused due to the roughness of the cutting tool [as with the case of Fig. 8(c)] or as a result of wear due to repeated stylus playback [in the case of Fig. 8(a)].

The cylinder used for test purposes in this study has an irregular groove cross section, which is in contrast to the smooth groove cross section investigated previously.[10,11] The test cylinder, which contains signals of known type, allows for a quantitative assessment of signal quality and a comparison of the two different estimation methods for $z_a(t)$.

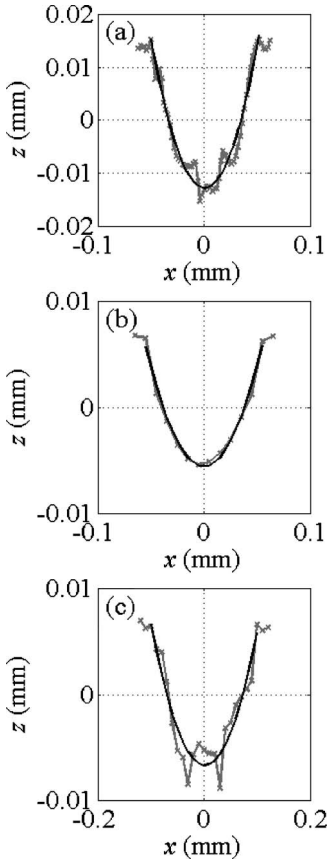


FIG. 8. Three groove valley examples. The smooth dark line in each plot shows a polynomial fit of order 2, which best fits the groove data, in a least squares sense. The Blue Amberol Cylinder (b) has strong agreement with the quadratic curve, whereas the graphophone (a) and test cylinder (c) example show a more complex groove shape, due to the cutting tool that produced the recording, and increased surface roughness or damage.

In this paper, we compare two methods of estimating the radial displacement track, $z_a(t)$, based on schemes (ii) and (iii). Both methods use the trajectory $\boldsymbol{A}=A(x,\theta)$, as defined by Eq. (7) as a guide for the approximate path followed by the cutting tool. These methods are now described.

a. Method A: Unweighted mean sample averaging across the groove matrix. A groove matrix $\boldsymbol{G}_t$ of size n by m is formed from the trajectory vector $\boldsymbol{A}=A(x,\theta)$, where n is the number of samples in the time domain and m is the number of samples across the groove cross section.

Radial displacement estimates $z_a(t)$ are then derived from $\boldsymbol{G}_t$ by an unweighted mean averaging process across the nth row of $\boldsymbol{G}_t$.

The mean average is taken from the central column of the $\boldsymbol{G}_t$, which corresponds to the medial axis of the groove cross section. The number of points included in the average varies from $N=1$ (central column of the $\boldsymbol{G}_t$ only) to $N=25$ points (± 12 samples either side of the central column of $\boldsymbol{G}_t$).

Method A provides the most simplistic estimate for the groove depth, based on averaging of samples across the groove cross section. By investigating different length sample averages, it is possible to observe the effects of mean averaging for the recovered signal in terms of SNR and THD.

TABLE I. Description of test cylinder recording and measurement details.

Parameter	Value
Length	100 mm
Diameter	55.56 mm
Recording speed	160 rpm
Recording pitch	100 turng/in.
Recording method	Electrical, direct cut
Recording medium	Wax cylinder
Sampling details	$\Delta x = 10\ \mu$m
	$\Delta\theta = 0.01°$
	$\Delta z = 10$ nm
Playback sample rate	96 kHz

b. Method B: Use of Savitzky–Golay filters of to locate unique groove minima. Raw surface data are filtered along each linescan via SG filter of polynomial order P and frame length L.

Along the trajectory vector $\boldsymbol{A}=A(x,\theta)$, the radial displacement estimate $z_a(t)$ is found by locating the local minimum point of the filtered surface.

Different SG filters of orders $P=\{1,3,\ldots,9\}$ and frame length $L=\{11,13,\ldots,25\}$ are used to obtain unique groove minima at different positions in the groove cross section.

In Method B, a SG filter is used to firstly smooth the groove cross section in order to account for the increased surface roughness caused by the cutting tool. The local minima of the smoothed cross section are then used as an estimate of the groove depth. Through the use of different combinations of filter polynomial order (P) and filter length (L), an optimal filter can be found by evaluating the SNR and THD of the recovered signal.

The best signal estimates derived by Methods A and B can then be compared in terms of SNR and THD.

B. Method for comparing audio signal quality

1. Details of test cylinder

A wax cylinder recording was produced for purposes of assessing the quality of the recovered signals. The cylinder was recorded at Poppy Records.[20] The recording stylus was not driven by acoustic pressure but was directly cut by an electrical transducer. Pure tones (sinusoids) in the range 200 Hz–5 kHz were recorded onto a cylinder at 160 rpm with a pitch of 100 turns/in. A geometrical description of the test cylinder is given in Table I.

Following the recording process, the groove structure remained untouched by stylus or cleaning. The surface map produced by the optical scanning system used a measurement grid of $[\Delta x=10\ \mu\text{m},\ \Delta\theta=0.01°]$. This sampling scheme gave approximately 25 data points per groove cross section and a playback sample rate $f_s=96$ kHz. The recovered signals exhibited some frequency modulation and harmonic distortion, which is in part due to the nature of the recording process. The degree of noise and distortion introduced at the time of recording is unknown. To this end, the experiment does not give an "absolute" measure of SNR/THD, but is useful for a comparative study of different audio signal estimation methods and comparisons with stylus replay. Recovered signals from the 1 kHz tone region provide the basis for comparison of audio signal estimates.

2. SNR analysis

The SNR is computed by locating the peak frequency and then calculating the total noise power in the remaining frequency bins of the DFT (considered to be noise). Definitions of the signal and noise powers, P_{sig} and P_n, as described in Eq. (3) are not immediately obvious for pseudosinusoidal signals, which are frequency modulated and contain harmonic distortion. Care must therefore be taken when deciding what constitutes the signal and noise. For example, it is questionable whether energy from harmonics in the spectrum should be considered as part of the signal or noise band. The choice of frequency band in which to search for P_{sig} is complicated by the fact that the recovered signals were frequency modulated. Other factors affecting the calculation include window length and frequency resolution of the DFT.

The short-time Fourier transform was used to carry out spectral analysis on windowed segments of the 1 kHz tone. Each frequency window was of duration 0.125 s, with a frequency bin resolution of 2 Hz. The signal power P_{sig} is found by locating the peak in the power spectrum (fundamental frequency) around 1 kHz.

Frequencies that contribute to the noise power P_n band were considered in two ranges: (1) below 1 kHz: 100–950 Hz and (2) above 1 kHz: 1050 Hz–10 kHz.

Harmonic frequencies above the fundamental were considered as undesirable and were also included in the noise band calculation. Frequencies below 100 Hz were not included in the noise band calculation, as these were attributed to the surface form and not to the audio signal.

3. Total harmonic distortion analysis

The THD is the ratio of the sum of the powers of all harmonics components above the fundamental frequency, to the power of the fundamental. Expressed as a percentage, the THD is given by

$$\text{THD}(\%) = 100 \times \frac{\sqrt{F_2^2 + F_3^2 + F_4^2 + \cdots + F_n^2}}{F_1}, \tag{8}$$

where F_n is the Fourier component of the nth harmonic; $n=1$ denotes the fundamental frequency. In the case of the 1 kHz tone, the first four harmonics (2, 3, 4, and 5 kHz) are included in this calculation, as harmonics above 5 kHz become lost in background noise. The lower the harmonic distortion, the more the recovered signal resembles a true sinusoid.

IV. RESULTS

A. Signal quality comparisons for Methods A and B

SNR and THD results for Method A are shown in Fig. 9.

Figure 9(a) shows that the SNR increases as the number of points in the average window, N, increases, but care must be taken when interpreting this trend. The best SNR for Method A comes from a 21 point mean average ($N=21$); however, this signal did not have the lowest THD [see Fig.

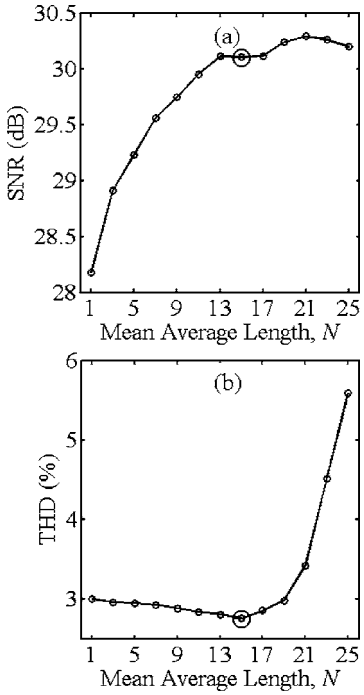


FIG. 9. SNR (a) and THD (b) results for Method A (mean averaging across the groove matrix). Larger circles indicate the optimal points on the curve.

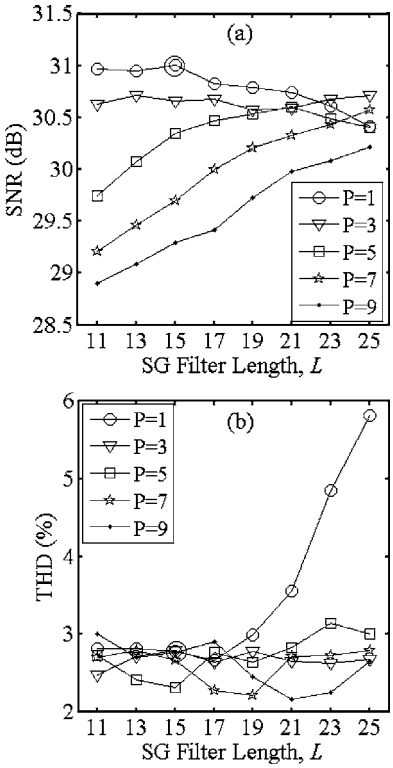


FIG. 10. SNR (a) and THD (b) results for Method B (SG filtering). Larger circles indicate the optimal points on the curve.

9(b)], and upon listening to the signal, a small echo effect was noticeable. This kind of temporal distortion is not accounted for in the spectrum averaging technique and is discussed in Sec. IV B. The signal that had the lowest THD (and hence the "purest" sinusoid) was from a mean average of $N=15$.

Mean averaging is a low pass filtering operation since its effect is to allow lower spatial frequencies to be retained, while suppressing higher frequency components. A larger mean average will remove more noise (high frequency) but can reduce the level of detail in the groove cross section. The result of this for the recovered sinusoid is a reduction in peak signal power and increased harmonic distortion when N is large (greater than 19 samples). In addition, by listening to signals where N is large, temporal distortion is also present in the form of an echo. A mean average of $N=15$ produced a signal with lowest THD, reasonably high SNR score, no noticeable temporal distortion, and was therefore considered to be the best signal produced by Method A.

SNR and THD results for Method B for the different SG filters are shown in Fig. 10.

In general, the SNR decreased as the filter order P increased. This is due to the fact that higher order polynomial filters retain the higher spatial frequency content (roughness) of the groove cross section (see Fig. 11), which results in increased high frequency noise for the audio signal. The highest SNR came from the filter $[P=1, L=15]$. A SG filter of order $P=1$ is essentially an unweighted moving average filter, similar to the mean averaging in Method A. This similarity can be observed by noticing that the THD curves for Methods A and B ($P=1$) have similar trends. An echo effect was also noticeable for $[P=1, L=19-25]$, as was found with Method A for large N.

With the exception of the first order filter ($P=1$), the THD scores remained relatively constant as L increased. The lowest THD came from the higher order filters ($[P=9, L=21]$, for example). This may be due to the fact that the unique groove minima produced by higher order SG filters tended toward the positions of raw groove minima, as opposed to being positioned higher up in the groove cross section, where raw displacement data do not exist (as can be seen in Fig. 11).

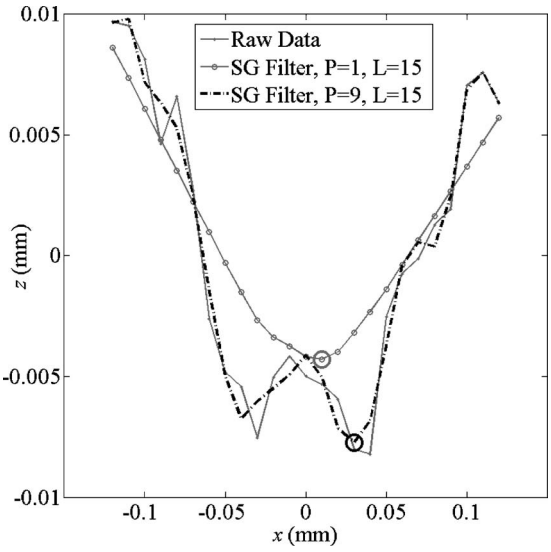


FIG. 11. First and ninth order SG filters ($L=15$) applied to the groove cross section of test cylinder data. The ninth order filter retains more of the higher spatial frequencies, whereas the first order filter produces a smoother cross section. Large circles indicate the position of unique groove minima from the filtered linescan used to form the displacement estimate, $z_a(t)$.

TABLE II. Summary of best SNR and THD results for Methods A and B.

	Method A (Mean averaging)	Method B (SG filtering)
Highest SNR (dB)	30.3	31.0
Lowest THD (%)	2.78	2.16

A summary of the best SNR and THD results is given in Table II. In both cases, an appropriately designed SG filter (Method B) gave better signal quality scores compared with simple mean averaging (Method A). This suggests that mean averaging is not optimal for noisy asymmetric groove cross sections, as seen with the test cylinder data.

B. Sources of noise and distortion

A time-frequency analysis shows that the recovered sinusoid signals exhibit the following types of noise and distortion.

1. Low frequency noise

The low frequency noise (below 200 Hz) is mainly due to the macroscopic surface form of the cylinder, which should be accounted for through an equalization stage. In order to match the velocity response of a magnetic cartridge (as is common with stylus reproduction), an appropriate equalization should be applied. In this paper we have only dealt with the raw displacement signals $z_a(t)$, which were not equalized or differentiated in any way.

2. Impulsive noise

Impulsive noise can be caused by surface debris, which was not cleaned from the cylinder prior to scanning. Any debris that lies above the imaging plane will therefore be measured by the sensor. This can result in transients for the recovered audio signal, as can be seen by the vertical bands in the spectrograms in Figs. 12 and 13. In order to minimize this effect, any loose debris should be removed from the surface prior to scanning wherever possible. Additionally, if the cylinder surface is measured in piecewise segments (as is required for heavily warped surfaces,[12] and the test cylinder in this case), impulsive noise can be introduced at segment joints if the segments are not appropriately aligned in software prior to sound extraction.

3. Temporal distortion

It is observed that displacement estimates that used groove cross section data toward the tops of the groove exhibited an echo. This phenomenon was first noted by Iwai *et al.*[1] for real-time optical playback of phonograph cylinders, when the width of their laser beam was wide compared with the groove width. Similarly, Fadeyev *et al.*[10] noted that groove ridge data provided audible albeit noisy sound content, but with interference from different times in the recording. The test cylinder data allow us to quantify the interference observed from the ridge data stream. Figure 12 shows a time-frequency analysis (spectrogram) of data streams derived from the top and bottom of the groove cross section. It

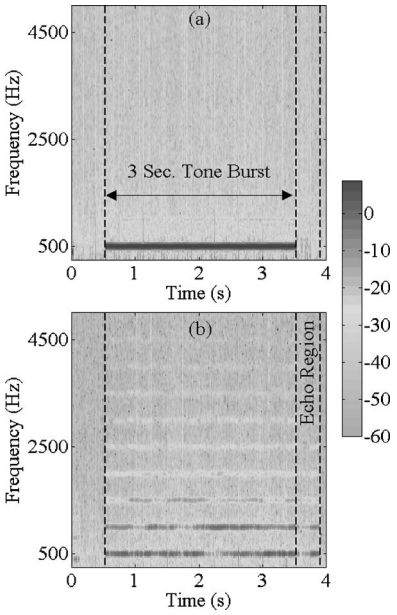


FIG. 12. Spectrogram analysis of signals recovered from the bottom (a) and top (b) of the groove cross section for the 500 Hz tone. An echo region is clearly visible in the ridge spectrogram (b) due to the superposition of adjacent groove displacement data at times t and $t+T$. In addition, the ridge data stream has more harmonic distortion.

appears that an echo is introduced due to the superposition of adjacent grooves at times t and $t+T$, where T is the time taken for one complete rotation (0.375 s in this case for 160 rpm). To avoid introducing temporal distortion, it is therefore suggested that the use of displacement data from the groove ridge should be avoided.

C. Comparison of stylus and noncontact reproduction

In addition to optical reproduction, the test cylinder audio was also transferred using conventional stylus techniques

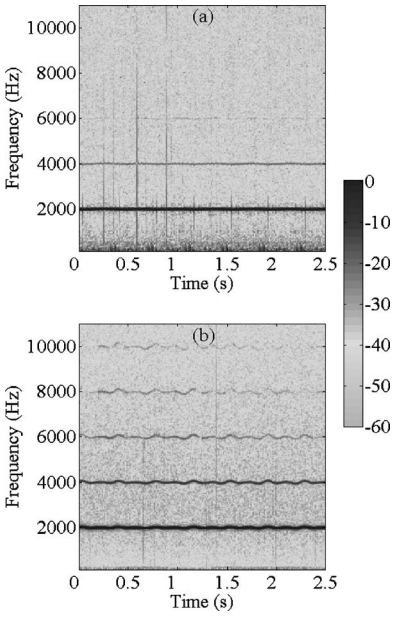


FIG. 13. Spectrogram comparisons for noncontact (a) and stylus reproduction (b) of 2 kHz tone. Harmonic distortion is greater with stylus reproduction. The noncontact method (a) shows less harmonic distortion than in (b) but more low frequency (below 500 Hz) and impulsive noise. In addition, frequency modulation is less apparent for the noncontact method.

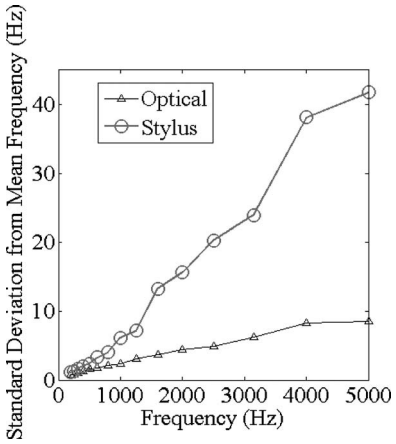


FIG. 14. Comparison of measured frequency modulation of tones recovered from the test cylinder via noncontact and stylus methods.

at the British Library Sound Archive. Due to difficulties in matching the equalization of the stylus and optical reproduction, direct signal quality comparisons are not presented in this paper. Subjective listening suggests that stylus reproduction has a higher SNR, in terms of a lower noise floor power and less impulsive noise (the increased impulsive noise found with optical reproduction is likely due to measurement of remaining debris above the imaging plane, as described in Sec. IV B 2). However, the accuracy of sine wave reproduction in terms of harmonic distortion and frequency modulation appears to be greater with the noncontact method. Examples of the recovered signals from optical and stylus methods can be found here.[21]

1. Harmonic distortion

The spectrogram in Fig. 13 compares the reproduction of a 2 kHz tone from both stylus and optical methods. Stylus reproduction appears to show increased harmonic distortion compared with the optical signal. This is likely due to the stylus-groove relationship, known in literature,[22,23] as "tracing distortion," which arises from the fact that the reproducing stylus tip is of finite size. When the stylus tip traces over a sinusoidal groove, a mismatch in curvature between the tip radius and the groove contour means that the reproduced signal is not sinusoidal.[22] This geometric effect results in harmonic distortion. The noncontact method of sound reproduction is not affected by tracing distortion.

2. Frequency modulation

The degree of frequency modulation was also compared by locating the position of the peak frequency in the DFT for windowed segments of recovered tones. The standard deviation in measured frequency from peaks in the DFT was used as an indication of frequency modulation. Figure 14 shows frequency modulation for all tones reproduced from the test cylinder by optical and stylus methods. At 5 kHz, the frequency modulation with stylus reproduction is approximately ± 40 Hz, and for optical reproduction is less than ± 10 Hz. The increased frequency modulation with stylus reproduction is likely due to the poor compliance of the cylinder bore with the rotating mandrel (due to lack of cylinder concentricity). If mounted incorrectly, the surface speed of the cylinder will vary, meaning that for a constant playback sample rate a modulation in frequency occurs. In this particular case, it appears that the stylus reproduction is more heavily affected by speed fluctuations, than with the noncontact method.

V. CONCLUSIONS

The noncontact full surface mapping system, which has been developed, allows for accurate sound reproduction from cylinder recordings. The advantage of mapping the full surface topology (as opposed to optical playback *in situ*) is that numerous estimates for the audio signal can be derived from a single data set in postprocessing. To the authors' knowledge, this study represents the first quantitative analysis of test signal reproduction using the noncontact surface mapping method.

A stylus trajectory estimation method based on the shift in relative phase between consecutive linescans has been described. Such a "global" method of trajectory estimation has advantages in terms of reduced processing time, when compared with "local" tracking of individual grooves, and has been found to be robust for the test cylinder recording. For cylinders that may be damaged or broken, local tracking of groove features is still required.

The test cylinder encoded with sinusoids provides a basis for quantitative assessment of sound reproduction quality. Two different methods of estimating the radial displacement track from a discrete surface have been shown to produce variable audio signal quality, in terms of SNR and THD. The use of appropriately designed SG filters (Method B) to produce an estimate for the groove depth provided a better audio signal estimate signal in terms of SNR and THD compared with unweighted mean averaging across the groove matrix (Method A).

Audio signal estimation from discrete surface maps is not restricted to the two methods described in this paper. With stylus reproduction, the use of different styli are chosen, based on the groove geometry and surface condition of the recording. The same principle applies to the noncontact method, in that the "optimal" method of signal reproduction is likely to be dependent on the cylinder under test. To improve further the quality of the recovered signal, other rule- or model-based navigation methods may be used to avoid the introduction of outliers (caused by debris, etc.).

Care must be taken when obtaining an estimate for the groove depth. For example, the use of groove ridge data was shown to introduce temporal distortion (echo). Averaging across the full groove cross section to obtain an estimate for $z_a(t)$ is therefore not advisable as the inclusion of ridge data from adjacent grooves can result in echo and increased harmonic distortion.

The audio signal recovered via stylus reproduction exhibited more frequency modulation and harmonic distortion compared with the optical method. Although this is a feature of this particular transfer (and not necessarily inherent for all stylus playback), it highlights some advantages of noncontact reproduction. For example, geometric effects of the

stylus-groove interaction, such as tracing distortion, do not affect optical reproduction. Additionally, in this particular case, the noncontact method appears to be less affected by speed irregularities, found with the stylus reproduction.

Difficulties encountered from very early wax cylinders include poor surface condition, such as cracking, mold growth, and wear caused to the groove structure through repeated stylus playback or poor storage. In cases where a cylinder is cracked, it is possible to scan the broken segments and then reconstruct the surface. In cases where the groove structure is damaged, it is possible to track more closely in postprocessing than with playback *in situ*.

Topics for further investigation in this field include the development of signal reproduction methods for cylinders in poor surface condition, including broken or shallow grooved recordings. In addition, a reduction in measurement time remains a priority for the data acquisition process. The use of other specially produced test cylinders encoded with known waveforms may also provide further scientific analysis of signal reproduction from the noncontact full surface mapping method.

ACKNOWLEDGMENTS

The authors would like to thank The British Library Sound Archive for assisting with the stylus transfer and The Engineering and Physical Sciences Research Council for funding this research.

[1]T. Iwai, T. Asakura, T. Ifukube, and T. Kawashima, "Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method," Appl. Opt. **25**, 597–604 (1986).
[2]T. Ifukube, T. Kawashima, and T. Asakura., "New methods of sound reproduction from old wax phonograph cylinders," J. Acoust. Soc. Am. **85**, 1759–1766 (1989).
[3]T. Asakura and J. Uozumi, "Optical methods for reproducing sounds from old phonograph records," *International Trends in Optics and Photonics*, ICO Vol. **IV** (Springer-Verlag, Berlin, Germany, 1999), pp. 65–81.
[4]S. M. Shanoylo, I. V. Kosyak, V. Petrov, and A. A. Kryuchin, "Reading and processing of audio information reproduced from Edison phonograph cylinders by method of laser interferometry," Proc. SPIE **4402**, 194–201 (2001).
[5]S. S. Cavaglieri, O. Johnsen, and F. Bapst, "Optical retrieval and storage of analog sound recordings," in AES 20th International Conference: "Archiving, Restoration, and New Methods of Recording," 2001.
[6]S. Stotzer, O. Johnsen, F. Bapst, C. Milan, and R. Ingold, "Phonographic sound extraction using photography and signal processing," Digit. Signal Process. **17**, 433–450 (2007).
[7]S. Stotzer, O. Johnsen, F. Bapst, and C. Sudan, "Phonographic sound extraction using image and signal processing," *Acoustics, Speech, and Signal Processing*, in Proceedings ICASSP'04, 2004, Vol. **4**, pp. 289–292.
[8]V. Fadeyev and C. Haber, "Reconstruction of mechanically recorded sound by image processing," J. Audio Eng. Soc. **51**, 1172–1185 (2003).
[9]V. Fadeyev, C. Haber, Z. Radding, C. Maul, J. W. McBride, and M. Golden, "Reconstruction of mechanically recorded sound by image processing," J. Acoust. Soc. Am. **115**, 2494 (2004).
[10]V. Fadeyev, C. Haber, C. Maul, J. W. McBride, and M. Golden, "Reconstruction of recorded sound from an Edison cylinder using three-dimensional non-contact optical surface metrology," J. Audio Eng. Soc. **53**, 485–508 (2005).
[11]A. J. Nascè, J. W. McBride, M. Hill, and P. J. Boltryk, "Signal processing methods for the recovery of audio from early acoustic cylinder recordings, measured via non-contact optical sensor," in AES 31st International Conference: New Directions in High Resolution Audio, Gainesville, Florida, 2007.
[12]P. J. Boltryk, M. Hill, J. W. McBride, and A. J. Nascè, "A comparison of precision optical displacement sensors for the 3D measurement of complex surface profiles," Sens. Actuators, A **142**, 2–11 (2008).
[13]S. Stotzer, "Phonographic record sound extraction by image processing," Ph.D. thesis, University of Fribourg, Switzerland.
[14]A. Millard, *America on Record: A History of Recorded Sound* (Cambridge University Press, Cambridge, 1995).
[15]O. Read and W. Welch, *From Tin Foil to Stereo: Evolution of the Phonograph* (Howard W. Sams & Co., Inc., Indianapolis, IN, 1977).
[16]H. Meulengracht-Madsen, "On the transcription of old phonograph wax records," J. Audio Eng. Soc. **24**, 27–32 (1976).
[17]IASA TC-03, "The safeguarding of the audio heritage: Ethics, principles and preservation strategy," 2005.
[18]Y.-C. Jenq, "Discrete-time method for signal-to-noise power ratio measurement," IEEE Trans. Instrum. Meas. **45**, 431–434 (1996).
[19]A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Anal. Chem. **36**, 1629–1639 (1964).
[20]Poppy Records, http://www.poppyrecords.co.uk/ (Last viewed 8/23/2008).
[21]The Sound Archive Project Website, "Examples of signal reproduction from test cylinder recordings," http://www.sesnet.soton.ac.uk/archivesound/audio/testcylinder/ Last viewed 8/23/2008.
[22]J. A. Pierce and F. V. Hunt, "On distortion in sound reproduction from phonograph records," J. Acoust. Soc. Am. **10**, 14–28 (1938).
[23]W. D. Lewis and F. V. Hunt, "A theory of tracing distortion in sound reproduction from phonograph records," J. Acoust. Soc. Am. **12**, 348–365 (1941).

Elastic medium equivalent to Fresnel's double-refraction crystal

José M. Carcione[a)]
Instituto Nazionale di Oceanografia e di Geofisica Sperimentale, Borgo Grotta Gigante 42c, 34016 Trieste, Italy

Klaus Helbig[b)]
Kiebitzrain 84, D 30657 Hannover, Germany

(Received 13 November 2007; revised 1 June 2008; accepted 18 July 2008)

In 1821, Fresnel obtained the wave surface of an optically biaxial crystal, assuming that light waves are vibrations of the ether in which longitudinal vibrations (P waves) do not propagate. An anisotropic elastic medium mathematically analogous to Fresnel's crystal exists. The medium has four elastic constants: a P-wave modulus, associated with a spherical P wave surface, and three elastic constants, c_{44}, c_{55}, and c_{66}, associated with the shear waves, which are mathematically equivalent to the three dielectric permittivity constants ϵ_{11}, ϵ_{22}, and ϵ_{33} as follows: $\mu_0\epsilon_{11} \Leftrightarrow \rho/c_{44}$, $\mu_0\epsilon_{22} \Leftrightarrow \rho/c_{55}$, $\mu_0\epsilon_{33} \Leftrightarrow \rho/c_{66}$, where μ_0 is the magnetic permeability of vacuum and ρ is the mass density. These relations also represent the equivalence between the elastic and electromagnetic wave velocities along the principal axes of the medium. A complete mathematical equivalence can be obtained by setting the P-wave modulus equal to zero, but this yields an unstable elastic medium (the hypothetical ether). To obtain stability the P-wave velocity has to be assumed infinite (incompressibility). Another equivalent Fresnel's wave surface corresponds to a medium with anomalous polarization. This medium is physically unstable even for a nonzero P-wave modulus.
 [DOI: 10.1121/1.2968705]

PACS number(s): 43.40.At, 43.20.Wd, 43.20.Bi, 43.20.Dk [MCH] Pages: 2053–2060

I. INTRODUCTION

Scientists of the 19th century have made extensive use of the analogy between light and elastic waves to study the behavior of light in matter. In particular, they have exploited the mathematical equivalence between the electromagnetic constitutive equations and the elastic rheological equations. In many cases, this practice led to important discoveries. For instance, Fresnel's formulas and Maxwell's equations were obtained from mathematical analogies—and physical analogies to a lesser degree—with shear wave propagation and Hooke's law, respectively.

As early as the 17th century it was known that light waves and elastic waves are of a similar nature. Hooke believed light to be a vibratory displacement of a medium (the ether), through which it propagates at finite speed. Later, in the 19th century, Maxwell and Lord Kelvin made extensive use of physical and mathematical analogies to study wave phenomena in elastodynamics and electromagnetism. In many cases, this formal analogy becomes a complete mathematical equivalence such that the problems in both fields can be solved by using the same analytical (or numerical) methodology.

Green[1] made the analogy between elastic waves in an incompressible solid (the ether) and light waves. One of the most remarkable analogies in science is the equivalence between electric and elastic displacements used by Maxwell to obtain his famous electromagnetic equations. Fresnel showed that if light were a transverse wave, then it would be possible to develop a theory accommodating the polarization of light. Therefore, the study of acoustic wave propagation and light propagation are intimately related, and this fact is reflected in the course of scientific research. With the advent of the theory of relativity, the concept of the ether was abandoned. However, the fact that electromagnetic waves are transverse waves is very useful.

Carcione and Cavallini[2] showed that the two-dimensional Maxwell equations describing propagation of the transverse-magnetic mode in anisotropic media are mathematically equivalent to the *SH* wave equation in an anisotropic-viscoelastic solid where attenuation is described with the Maxwell model. The problem of energy definition in the time domain, particularly for lossy media, has been discussed by Carcione[3] using mathematical analogies. Later, Carcione and Robinson[4] established the analogy for the reflection-transmission problem, showing that contrasts in compressibility yield the reflection coefficient for light polarized perpendicular to the plane of incidence (Fresnel's sine law—the electric vector perpendicular to the plane of incidence), and density contrasts yield the reflection coefficient for light polarized in the plane of incidence (Fresnel's tangent law).

Fresnel[5] read a summary of a memoir to the Academy of Sciences in Paris on 26 November 1821. He presented the wave surface of an optically biaxial purely dielectric medium—a crystal such as calcite or iceland spar. Referring to his equation he writes: "If in the construction that Huygens made to determine the direction of the refracted rays by Iceland spar, and which can be applied to any waveform, one substitutes the sphere and the ellipsoid of revolution by the surface composed of the two terms represented by this last

a)Electronic mail: jcarcione@inogs.it
b)Electronic mail: helbig.klaus@t-online.de

equation, and operating elsewhere in the same way, one will have two tangent planes at the points of contact, joined at the center of the wave, which will give the direction of the ordinary ray and that of the extraordinary ray." It constitutes, therefore, a generalization of the isotropic equations to the anisotropic case, where double refraction occurs. In this work, we obtain Fresnel's double-refraction equivalent anisotropic elastic medium, starting from the Newton–Euler equations and the stress–strain relations of an orthorhombic elastic medium. The examples illustrate the different cases where the Fresnel wave surface can be obtained.

II. ELECTROMAGNETISM

A. Maxwell's equations

Maxwell's equations, for a purely dielectric (lossless) medium, in the absence of electric and magnetic sources, are[6]

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B},$$

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D}, \tag{1}$$

where **E**, **H**, **D**, and **B** are the electric vector, the magnetic vector, the electric displacement, and the magnetic induction, respectively, the multiplication sign denotes vector product and ∂_t represents the time derivative. Constitutive equations are needed to relate **D** and **B** to the field vectors. For a dielectrically anisotropic medium we have

$$\mathbf{D} = \boldsymbol{\epsilon} \cdot \mathbf{E},$$

$$\mathbf{B} = \mu_0 \mathbf{H}, \tag{2}$$

where $\boldsymbol{\epsilon}$ is the dielectric-permittivity matrix, μ_0 is the magnetic permeability of vacuum, and the dot indicates matrix product.

For optically biaxial media, the dielectric-permittivity matrix is given by[2]

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{pmatrix}. \tag{3}$$

B. Kelvin–Christoffel matrix and slowness surface

Assume harmonic plane waves with a phase factor

$$\exp[i\omega(t - \mathbf{s} \cdot \mathbf{x})], \tag{4}$$

where $i = \sqrt{-1}$, ω is the angular frequency, **s** is the slowness vector, and **x** is the position vector. We use the following correspondence between space–time and slowness–frequency domains:

$$\nabla \times \; \rightarrow -i\omega \mathbf{s} \times, \quad \partial_t \rightarrow i\omega. \tag{5}$$

Substituting the plane wave (4) into Maxwell's equations (1) and using Eqs. (2) and (5) gives

$$\mathbf{s} \times \mathbf{E} = \mathbf{B} = \mu_0 \mathbf{H}, \tag{6a}$$

$$\mathbf{s} \times \mathbf{H} = -\mathbf{D} = -\boldsymbol{\epsilon} \cdot \mathbf{E}. \tag{6b}$$

The vector product of Eq. (6b) with **s** and use of Eq. (6a) yields

$$\mathbf{s} \times [(\boldsymbol{\epsilon})^{-1} \cdot \mathbf{s} \times \mathbf{H}] + \mu_0 \mathbf{H} = 0. \tag{7}$$

In terms of components we have

$$(\epsilon_{ijk} s_j (\epsilon_{kl})^{-1} \epsilon_{lpq} s_p + \mu_0 \delta_{iq}) H_q = 0, \quad i = 1, \ldots, 3, \tag{8}$$

where ϵ_{ijk} are the components of the Levi-Civita tensor. Consider the case given by Eq. (3). Then, the equivalent of the elastic Kelvin–Christoffel equation (e.g., Refs. 7 and 8) for the magnetic vector is

$$\boldsymbol{\Gamma} \cdot \mathbf{H} = \mathbf{0}, \tag{9}$$

where the Kelvin–Christoffel matrix is

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 - \left(\frac{s_2^2}{\eta_3} + \frac{s_3^2}{\eta_2}\right) & \frac{s_1 s_2}{\eta_3} & \frac{s_1 s_3}{\eta_2} \\ \frac{s_1 s_2}{\eta_3} & 1 - \left(\frac{s_1^2}{\eta_3} + \frac{s_3^2}{\eta_1}\right) & \frac{s_2 s_3}{\eta_1} \\ \frac{s_1 s_3}{\eta_2} & \frac{s_2 s_3}{\eta_1} & 1 - \left(\frac{s_1^2}{\eta_2} + \frac{s_2^2}{\eta_1}\right) \end{pmatrix}, \tag{10}$$

where

$$\eta_i = \mu_0 \epsilon_{(i)i}. \tag{11}$$

Equation (9) has two solutions, which represent light waves transversally polarized.[6,7,9]

The dispersion relation (i.e., the vanishing of the determinant of the Kelvin–Christoffel matrix) is the slowness surface:

$$(\eta_1 s_1^2 + \eta_2 s_2^2 + \eta_3 s_3^2)(s_1^2 + s_2^2 + s_3^2) - (\eta_1 \zeta_1 s_1^2 + \eta_2 \zeta_2 s_2^2 + \eta_3 \zeta_3 s_3^2) + \eta_1 \eta_2 \eta_3 = 0, \tag{12}$$

where

$$\zeta_i = \eta_j + \eta_k, \quad j \neq k \neq i, \tag{13}$$

There are only quartic and quadratic terms of the slowness components in the dispersion relation of an anisotropic medium.

C. Fresnel's wave surface

Now, we use the following property:

$$\mathbf{s} \cdot \mathbf{v} = 1, \tag{14}$$

where **v** is the group or energy velocity.[7,8] Equation (14) states that the slowness and wave-velocity surfaces are polar reciprocal, i.e.,

$$x s_1 + y s_2 + z s_3 = 1, \tag{15}$$

where we have assumed $\mathbf{x} = (x, y, z) = \mathbf{v}t$, with $t = 1$; x, y, and z define a point in the wave surface. The wave surface can be obtained by standard methods, from the group and energy velocities,[10,7] or using the duality principle.[6] Using these two different methods, it is shown in the Appendix that the wave surface is given by

$$\left(\frac{x^2}{\eta_1}+\frac{y^2}{\eta_2}+\frac{z^2}{\eta_3}\right)(x^2+y^2+z^2) - \left[\frac{x^2}{\eta_1}\left(\frac{1}{\eta_2}+\frac{1}{\eta_3}\right)+\frac{y^2}{\eta_2}\left(\frac{1}{\eta_1}+\frac{1}{\eta_3}\right) + \frac{z^2}{\eta_3}\left(\frac{1}{\eta_1}+\frac{1}{\eta_2}\right)\right]+\frac{1}{\eta_1\eta_2\eta_3}=0, \tag{16}$$

This is Fresnel's wave surface,[5]

$$(a^2x^2+b^2y^2+c^2z^2)(x^2+y^2+z^2)-[a^2x^2(b^2+c^2) + b^2y^2(a^2+c^2)+c^2z^2(a^2+b^2)]+a^2b^2c^2=0, \tag{17}$$

where

$$a=\frac{1}{\sqrt{\eta_1}}, \quad b=\frac{1}{\sqrt{\eta_2}}, \quad c=\frac{1}{\sqrt{\eta_3}} \tag{18}$$

are wave (light) velocities along the principal axes of the crystal ("axes optiques," according to Fresnel). Quoting Fresnel: "The three semi-axes a,b,c, represent here the propagation velocities of the parallel vibrations."

It can easily be shown that at the coordinate planes the wave surface is factorable into two factors, and each factor is an ellipse. Also, the intersection of the slowness surface with the Cartesian planes are ellipses, since the polar reciprocal of an ellipse is another ellipse.

The extension of the theory to the lossy case is given in Born and Wolf,[6] and for instance, in Carcione and Schoenberg,[9] and Carcione.[7]

III. ELASTODYNAMICS

In the following, we obtain the elastic medium equivalent to Fresnel's double-refraction crystal. Fresnel chose his medium to be orthorhombic, thus it is reasonable to look for the mechanical equivalent in a similar medium. Moreover, the electromagnetic field is governed by a tensor of rank two, thus the slowness and wave surfaces are closed surfaces of order 2, i.e., they are ellipsoids (and thus have three mutually perpendicular symmetry planes). Elastic wave fields are governed by tensors of rank 4, thus slowness surfaces are of order 6 and wave surfaces of order n with n an even number with $6 \leqslant n \leqslant 150$. Thus, geometrically similar slowness and wave surfaces in an elastic medium can exist only if the medium has at least orthorhombic symmetry (to guarantee three symmetry planes), and if given conditions are satisfied to force the surfaces to be (three-axial) ellipsoids [see Eq. (32)].

A. Newton–Euler equation

In the absence of body forces, the Newton–Euler equation governing the dynamic of continuum media is

$$\bar{\nabla}\cdot\boldsymbol{\sigma}=\rho\partial_{tt}^2\mathbf{u}, \tag{19}$$

where ρ is the density,

$$\boldsymbol{\sigma}=(\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_5,\sigma_6)^T \tag{20}$$

is the stress vector,

$$\mathbf{u}=(u_1,u_2,u_3)^T \tag{21}$$

is the displacement vector, and

$$\bar{\nabla}=\begin{pmatrix}\partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2\\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1\\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0\end{pmatrix} \tag{22}$$

is the gradient operator.[10,7]

The strain–displacement relation can be written as

$$\mathbf{e}=\bar{\nabla}^T\cdot\mathbf{u}, \tag{23}$$

where

$$\mathbf{e}=(e_1,e_2,e_3,e_4,e_5,e_6)^T \tag{24}$$

is the strain vector

The stress–strain relation reads

$$\boldsymbol{\sigma}=\rho\mathbf{C}\cdot\mathbf{e}, \tag{25}$$

with the elasticity matrix for orthorhombic media—normalized by the density—given by

$$\mathbf{C}=\begin{pmatrix}C_{11} & C_{12} & C_{13} & 0 & 0 & 0\\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0\\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0\\ 0 & 0 & 0 & C_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & C_{55} & 0\\ 0 & 0 & 0 & 0 & 0 & C_{66}\end{pmatrix}, \tag{26}$$

where $C_{IJ}=c_{IJ}/\rho$ and c_{IJ} are the elastic constants.

B. Kelvin–Christoffel matrix and slowness surface

Combining Eqs. (19), (23), and (25)yields

$$\bar{\nabla}\cdot\mathbf{C}\cdot\bar{\nabla}^T\cdot\mathbf{u}=\partial_{tt}^2\mathbf{u}. \tag{27}$$

Substituting the plane wave (4) into Eq. (27) and using $\partial_t\rightarrow i\omega$ and

$$\bar{\nabla}\rightarrow -i\omega\begin{pmatrix}s_1 & 0 & 0 & 0 & s_3 & s_2\\ 0 & s_2 & 0 & s_3 & 0 & s_1\\ 0 & 0 & s_3 & s_2 & s_1 & 0\end{pmatrix}\equiv -i\omega\mathbf{S}, \tag{28}$$

we obtain

$$\boldsymbol{\Gamma}\cdot\mathbf{u}=\mathbf{0} \tag{29}$$

where

$$\boldsymbol{\Gamma}=\mathbf{I}-\mathbf{S}\cdot\mathbf{C}\cdot\mathbf{S}^T, \tag{30}$$

is a symmetric Kelvin–Christoffel matrix, and $\mathbf{I}$ is the 3×3 identity matrix.

Equation (29) is analogous to Eq. (9), with

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1-(C_{11}s_1^2 + C_{66}s_2^2 + C_{55}s_3^2) & -(C_{12}+C_{66})s_1s_2 & -(C_{13}+C_{55})s_1s_3 \\ -(C_{12}+C_{66})s_1s_2 & 1-(C_{66}s_1^2 + C_{22}s_2^2 + C_{44}s_3^2) & -(C_{23}+C_{44})s_2s_3 \\ -(C_{13}+C_{55})s_1s_3 & -(C_{23}+C_{44})s_2s_3 & 1-(C_{55}s_1^2 + C_{44}s_2^2 + C_{33}s_3^2) \end{pmatrix}. \tag{31}$$

The vanishing of the determinant of the Kelvin–Christoffel matrix $\boldsymbol{\Gamma}$ yields the slowness surface. The three solutions of Eq. (29) in normally polarized media[7] correspond to the qP wave and two qS waves.

C. Elastic medium equivalent to an optically biaxial crystal

We know from the acoustic-electromagnetic analogy[2] that light waves are mathematically analogous to shear waves. Then, in order to find the elastic medium mathematically equivalent to an optically biaxial medium corresponding to Fresnel's equation (17), we have to decouple the P wave from the S waves. These relations are necessary and sufficient:[11]

$$C_{11} = C_{22} = C_{33} = C_{12} + 2C_{66} = C_{13} + 2C_{55} = C_{23} + 2C_{44} \equiv E, \tag{32}$$

where E is a P wave modulus (normalized by the density). Substituting these relations into Eq. (31) gives

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1-(C_{66}s_2^2 + C_{55}s_3^2) & C_{66}s_1s_2 & C_{55}s_1s_3 \\ C_{66}s_1s_2 & 1-(C_{66}s_1^2 + C_{44}s_3^2) & C_{44}s_2s_3 \\ C_{55}s_1s_3 & C_{44}s_2s_3 & 1-(C_{55}s_1^2 + C_{44}s_2^2) \end{pmatrix} - E\begin{pmatrix} s_1^2 & s_1s_2 & s_1s_3 \\ s_1s_2 & s_2^2 & s_2s_3 \\ s_1s_3 & s_2s_3 & s_3^2 \end{pmatrix}. \tag{33}$$

It can easily be shown that the dispersion relation obtained from Eq. (33) gives two coupled S waves and a decoupled isotropic compressional wave, with a spherical wave surface and velocity $\sqrt{E}$. Anisotropic elastic materials having spherical wave surfaces are discussed by Ting.[12] Other anisotropic media propagating pure longitudinal waves are studied by Rychlewski[13] and Ostrosablin.[14] Note that for $E=0$, S waves propagate only. In isotropic media $\rho E = \lambda + 2\mu$, where λ and μ are the Lamé constants. If $E=0$, the compression modulus $k = \lambda + (2/3)\mu = -(4/3)\mu < 0$. Such a medium would be unstable (see the following).

The mathematical analogies between elastic and light waves were known to scientists of the 19th century. In fact, Fresnel, Green, MacCullagh, and Cauchy, among others, obtained expressions of wave surfaces and reflection coefficients in crystals by using the elastodynamic equations, assuming that light waves are the vibrations of a medium called ether. Fresnel[5] states "I consider a medium endowed with the double refraction having different elasticities along the various directions."

MacCullagh,[15] (see also Ref. 16, p. 141), in a conference held at the Royal Irish Academy in 1839, presented an isotropic medium, whose potential energy is only based on rotation of the volume elements, thus ignoring pure dilatations from the beginning. The result is a rotationally elastic ether and the wave equation for shear waves. The corresponding reflection and refraction coefficients coincide with Fresnel's formulae.[6] Green[1] assumed the P-wave velocity to be infinite and dismissed a zero P-wave velocity on the basis that the medium would be unstable since the compression modulus is negative (from the previous discussion, please note that the potential energy must be positive). Cauchy,[17] (see also Ref. 16) neglecting this fact considered that P waves have zero velocity ($E=0$), and obtained the sine law and tangent law of Fresnel. He assumed the shear modulus to be the same for both media. Cauchy's ether is known as the *contractile* or *labile aether*. It corresponds to an elastic medium of negative compressibility, as shown earlier.

Fresnel[5] states "The light vibrations perform only by following the directions parallel to the wave surface. In the note already cited I have presented this hypothesis with some development, I showed that it is enough to admit that the ether has a resistance so high to compression to conceive the absence of longitudinal vibrations."

Rudzki[11] investigated the condition for the separate propagation of "dilatational" (rot u=0) and "torsional" (div u=0) waves. The second condition leads immediately to

$$s_1u_1 + s_2u_2 + s_3u_3 = 0. \tag{34}$$

In this way, he obtains the dispersion relation for the shear waves. In fact, Eq. (29) results in

$$[1-(Es_1^2 + C_{66}s_2^2 + C_{55}s_3^2)]u_1 + (C_{66}-E)s_1s_2u_2 + (C_{55}-E)s_1s_3u_3 = 0,$$

$$(C_{66}-E)s_1s_2u_1 + [1-(C_{66}s_1^2 + Es_2^2 + C_{44}s_3^2)]u_2 + (C_{44}-E)s_2s_3u_3 = 0,$$

$$(C_{55}-E)s_1s_3u_1+(C_{44}-E)s_2s_3u_2+[1-(C_{55}s_1^2+C_{44}s_2^2+Es_3^2)]u_3=0, \quad (35)$$

or, in view of Eq. (34),

$$[1-(C_{66}s_2^2+C_{55}s_3^2)]u_1+C_{66}s_1s_2u_2+C_{55}s_1s_3u_3=0,$$

$$C_{66}s_1s_2u_1+[1-(C_{66}s_1^2+C_{44}s_3^2)]u_2+C_{44}s_2s_3u_3=0,$$

$$C_{55}s_1s_3u_1+C_{44}s_2s_3u_2+[1-(C_{55}s_1^2+C_{44}s_2^2)]u_3=0, \quad (36)$$

which correspond to the first matrix of the right-hand side in Eq. (33), describing the behavior of the coupled S waves only.

D. Equivalence and Fresnel's slowness and wave surfaces

Like Cauchy, we set $E=0$ in Eq. (33) and use the following mathematical analogies:

$$C_{44}=\frac{c_{44}}{\rho}\Leftrightarrow\frac{1}{\eta_1}=\frac{1}{\mu_0\epsilon_{11}},$$

$$C_{55}=\frac{c_{55}}{\rho}\Leftrightarrow\frac{1}{\eta_2}=\frac{1}{\mu_0\epsilon_{22}},$$

$$C_{66}=\frac{c_{66}}{\rho}\Leftrightarrow\frac{1}{\eta_3}=\frac{1}{\mu_0\epsilon_{33}}. \quad (37)$$

Substituting these relations into Eq. (33), we obtain Fresnel's slowness surface (12), associated with Fresnel's wave surface (17), where $\sqrt{C_{44}}$, $\sqrt{C_{55}}$, and $\sqrt{C_{66}}$ are the wave velocities of the shear waves along the principal axes of the medium. Strictly, it is not necessary to set $E=0$, leading to an unstable medium, to obtain Fresnel's wave surface, because the P wave is decoupled from the S waves. However, the mathematical equivalence is not complete. Physically, the hypothetical medium (the ether), supporting only transverse vibrations, is an elastically unstable medium.

Another way to get rid of the P wave is to assume incompressibility. This problem was of the utmost importance to the proponents of the elastic ether. Green,[1] for instance, chose infinite velocity for the outer wave front. With infinite stiffness, there is an immediate signal (infinite velocity) with vanishing displacement amplitude (infinite impedance), while with vanishing stiffness the signal arrives after infinite time with vanishing pressure amplitude (zero impedance).

E. Anomalously polarized medium

There are media with the same slowness and wave surfaces but drastically different polarization behavior. Such media are kinematically identical but dynamically different. Examples of anomalous polarization have been discussed for transverse isotropy by Helbig and Schoenberg,[18] and for orthorhombic symmetry by Carcione and Helbig.[19]

We do not repeat the details of the theory here,[7] but give a specific example. Let us consider the following Kelvin–Christoffel matrix:

$$\mathbf{\Gamma}=\begin{pmatrix} 1-(C_{11}s_1^2+C_{66}s_2^2+C_{55}s_3^2) & (C_{12}^*+C_{66})s_1s_2 & -(C_{13}+C_{55})s_1s_3 \\ (C_{12}^*+C_{66})s_1s_2 & 1-(C_{66}s_1^2+C_{22}s_2^2+C_{44}s_3^2) & (C_{23}^*+C_{44})s_2s_3 \\ -(C_{13}+C_{55})s_1s_3 & (C_{23}^*+C_{44})s_2s_3 & 1-(C_{55}s_1^2+C_{44}s_2^2+C_{33}s_3^2) \end{pmatrix}. \quad (38)$$

It is clear that the dispersion equation associated with this matrix is the same as that corresponding to Eq. (31), where C_{12}^* and C_{23}^* are the elastic constants of the anomalously polarized medium, which differ from those of the normally polarized medium. Comparing the two matrices we obtain the following relations, instead of Eq. (32):

$$C_{11}=C_{22}=C_{33}=-C_{12}^*=C_{13}+2C_{55}=-C_{23}^*\equiv E. \quad (39)$$

Then, we have

$$\mathbf{\Gamma}=\begin{pmatrix} 1-(C_{66}s_2^2+C_{55}s_3^2) & -C_{66}s_1s_2 & C_{55}s_1s_3 \\ -C_{66}s_1s_2 & 1-(C_{66}s_1^2+C_{44}s_3^2) & -C_{44}s_2s_3 \\ C_{55}s_1s_3 & -C_{44}s_2s_3 & 1-(C_{55}s_1^2+C_{44}s_2^2) \end{pmatrix}-E\begin{pmatrix} s_1^2 & -s_1s_2 & s_1s_3 \\ -s_1s_2 & s_2^2 & -s_2s_3 \\ s_1s_3 & -s_2s_3 & s_3^2 \end{pmatrix}, \quad (40)$$

instead of Eq. (33). Such a medium is anomalously polarized, because the polarization vectors are not perpendicular and tangent to the *P*- and *S*-wave slowness surfaces, respectively, as it is for a normally polarized medium corresponding to the elastic constants given in Eq. (32). An example is shown in Sec. IV.

Although the *S* waves of this medium are described by the Fresnel wave surface, the medium is physically unstable since the stability conditions[8] are not satisfied; for instance, $C_{11}C_{22}-C_{12}^2>0$ implies $0>0$, etc.

IV. EXAMPLES

Here, the theoretical results are verified by means of full-wave three-dimensional numerical simulations. The modeling code is based on the Fourier pseudospectral method for computing the spatial derivatives and a Chebyshev expansion of the evolution operator as time-integration technique. This algorithm possesses spectral accuracy for band-limited signals and is not affected by temporal or spatial numerical dispersion. The details can be found in Carcione *et al.*[20] The source is an additional term on the right-hand side of Eq. (19), i.e., $\rho\partial_{tt}^2\mathbf{u}+\mathbf{f}$, where $\mathbf{f}=(f_1,f_2,f_3)$, for directional forces, $\mathbf{f}=\boldsymbol{\nabla}\phi$, where $\phi=\phi(x,y,z)$ for a pressure source, and $\mathbf{f}=\boldsymbol{\nabla}\times\mathbf{A}$, where $\mathbf{A}=(A_1,A_2,A_3)$, $A_i=a_i\phi(x,y,z)$, for a shear source; ϕ is a Gaussian function.

We consider a medium with $E=10$, $C_{44}=3$, $C_{55}=2$, and $C_{66}=1$ ($C_{12}=8$, $C_{13}=6$ and $C_{23}=4$), where these elastic constants are normalized by $\rho\times$MPa, with ρ given in kg/m^3. Figure 1 shows sections of the slowness and wave (energy or group-velocity) surfaces on the symmetry planes. The polarization is indicated in the velocity sections. The outer curve (dashed line) represents the *P*-wave spherical front, and the inner curves (solid lines) correspond to the shear waves, equivalent to Fresnel's electromagnetic waves. The conical point is located in the plane defined by the maximum and minimum principal axes, i.e., the (x,z) plane.

The size of the numerical mesh to perform the simulations is 165×165 grid points, and the motion is initiated by a directional force making an angle $\pi/4$ with the principal axes. A Ricker pulse is used.[7] Figure 2 shows the snapshots of the components of the displacement vector on the three symmetry planes of the medium.

The anomalously polarized medium has the same elastic constants, except $C_{12}^*=C_{23}^*=-10$. The wave surface and snapshots are shown in Figs. 3 and 4, respectively. The shape of the slowness and wave surfaces are the same as those of Fig. 1. Figure 5 represents the wave field in terms of the polarization vector $\mathbf{u}$. As can be seen, the polarization presents anomalous behavior, changing from perpendicular to tangent to the wave front in the same mode. The medium has anomalous polarization in the (x,y) and (y,z) planes, while the polarizations in the (x,z) plane are unaltered.

ACKNOWLEDGMENT

Thank you to F. Cavallini for the valuable collaboration.

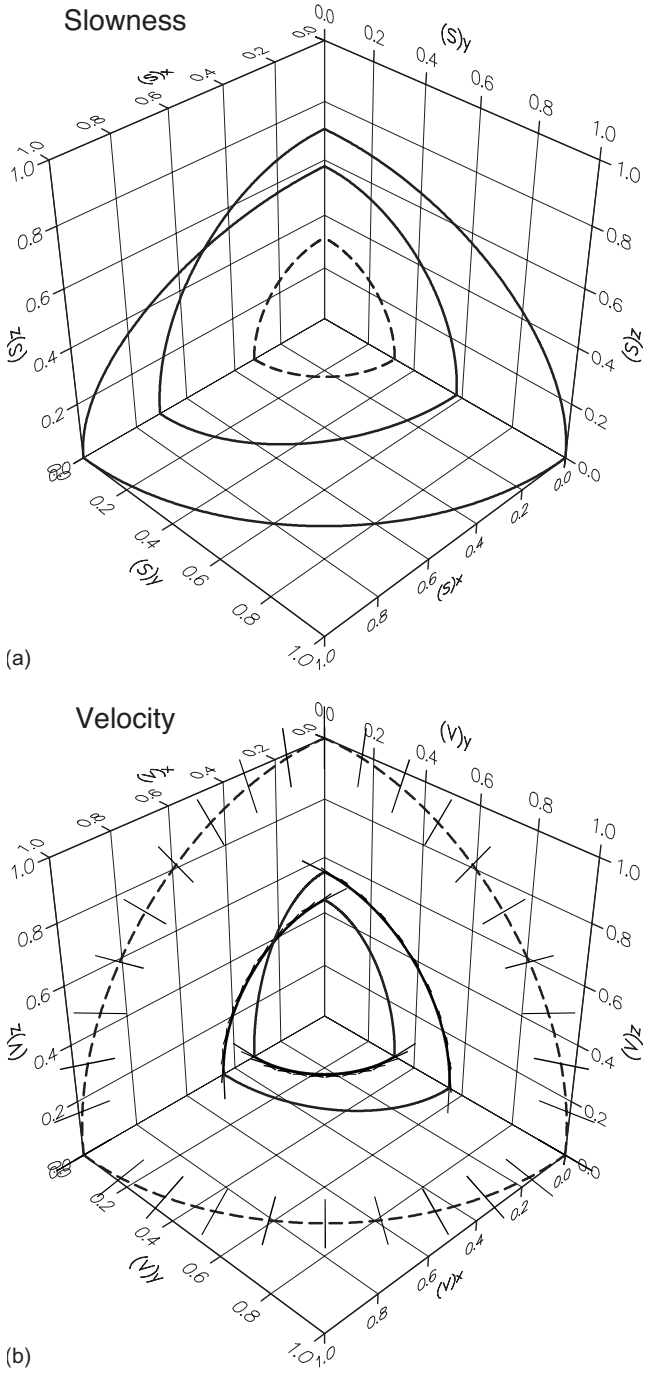


FIG. 1. Normalized sections of the slowness (a) and wave (velocity) (b) surfaces of an elastic anisotropic medium equivalent to an optically biaxial dielectric medium. One octant is shown because of symmetries. The spherical sections (dashed line) correspond to the *P* wave. The polarization is indicated for each wave mode.

APPENDIX: CALCULATION OF THE WAVE SURFACE

We obtain Fresnel's wave surface by using two different methods.

(1) Let us consider the s_1 and x components, and Eq. (12) for $s_2=s_3=0$. Then,

$$s_1^4-(\eta_2+\eta_3)s_1^2+\eta_2\eta_3=0. \tag{A1}$$

Using Eq. (15), we obtain

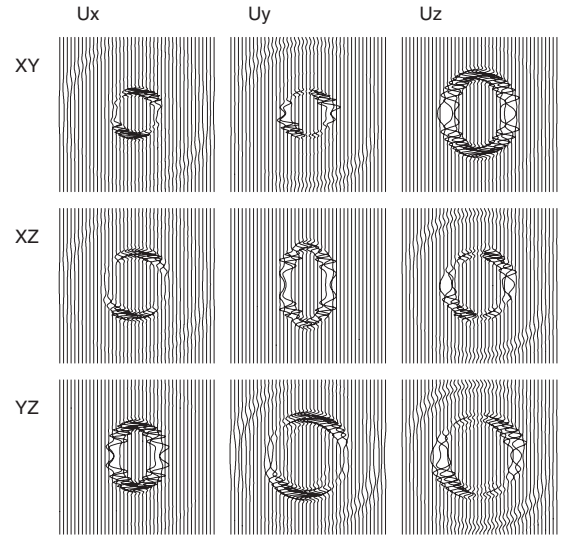


FIG. 2. Snapshots of the displacement vector corresponding to the case shown in Fig. 1. The displacement fields u_x, u_y, and u_z are shown for the three symmetry planes of the medium.

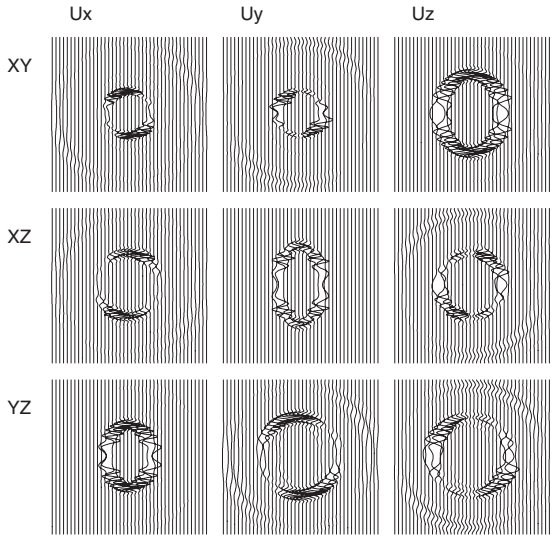


FIG. 4. Anomalous polarization. Snapshots of the displacement vector corresponding to the case shown in Fig. 3. The displacement fields u_x, u_y, and u_z are shown for the three symmetry planes of the medium.

$$x^4 - \left(\frac{1}{\eta_2} + \frac{1}{\eta_3}\right)x^2 + \frac{1}{\eta_2\eta_3} = 0. \tag{A2}$$

Dividing this equation by η_1 and performing a similar procedure for the other components, we have

$$\frac{x^4}{\eta_1} - \left[\frac{1}{\eta_1}\left(\frac{1}{\eta_2} + \frac{1}{\eta_3}\right)\right]x^2 + \frac{1}{\eta_1\eta_2\eta_3} = 0,$$

$$\frac{y^4}{\eta_2} - \left[\frac{1}{\eta_2}\left(\frac{1}{\eta_1} + \frac{1}{\eta_3}\right)\right]y^2 + \frac{1}{\eta_1\eta_2\eta_3} = 0,$$

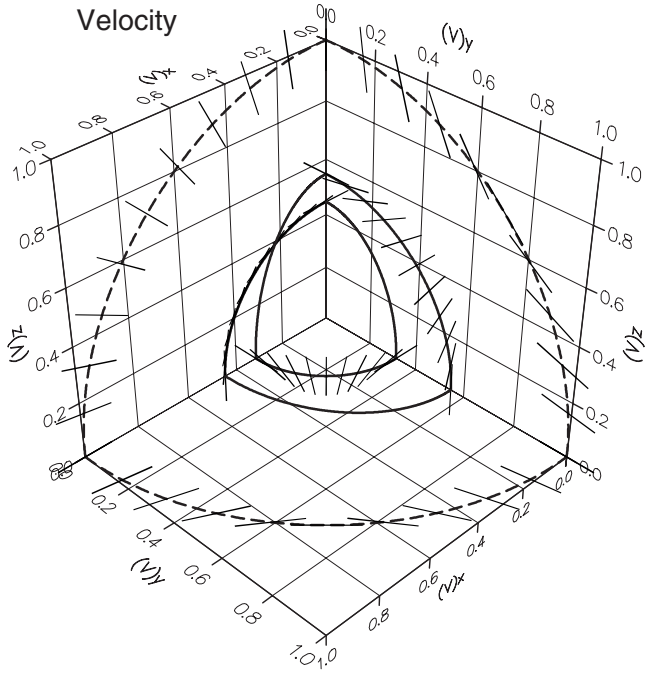


FIG. 3. Wave (velocity) surface corresponding to the anomalously polarized medium.

$$\frac{z^4}{\eta_3} - \left[\frac{1}{\eta_3}\left(\frac{1}{\eta_1} + \frac{1}{\eta_2}\right)\right]z^2 + \frac{1}{\eta_1\eta_2\eta_3} = 0. \tag{A3}$$

Since the slowness surface is a quadratic surface[21] and the slowness and velocity vectors are reciprocal, this implies that the wave surface has the same functional form as the slowness surface (12). Then

$$(a_1x^2 + a_2y^2 + a_3z^2)(x^2 + y^2 + z^2) - (b_1x^2 + b_2y^2 + b_3z^2) + d = 0. \tag{A4}$$

Comparing Eqs. (A3) and (A4), we have $a_1 = \eta_1^{-1}$, $a_2 = \eta_2^{-1}$, $a_3 = \eta_3^{-1}$, $b_1 = \eta_1^{-1}(\eta_2^{-1} + \eta_3^{-1})$, $b_2 = \eta_2^{-1}(\eta_1^{-1} + \eta_3^{-1})$, $b_3 = \eta_3^{-1}(\eta_1^{-1} + \eta_2^{-1})$ and $d = (\eta_1\eta_2\eta_3)^{-1}$. Hence, Eq. (16) is obtained.

(2) A more rigorous method is the so-called principle of duality.[6] It is known that in lossless—pure dielectrics—anisotropic media, **E** and **H** are tangent to the slowness surface and that the group-velocity vector **v** is perpendicular to that surface. Therefore

$$\mathbf{v}\cdot\mathbf{E} = 0, \quad \mathbf{v}\cdot\mathbf{H} = 0.$$

Let us consider the property $\mathbf{v}\times(\mathbf{s}\times\mathbf{A}) = -(\mathbf{v}\cdot\mathbf{s})\mathbf{A} + (\mathbf{v}\cdot\mathbf{A})\mathbf{s} = -\mathbf{A}$, where **A** is **E** or **H** and Eq. (14) has been used. Taking the vector product of Eq. (6) with **v** and using that property gives

$$\mathbf{v}\times\mathbf{D} = \mu_0^{-1}\mathbf{B},$$

$$\mathbf{v}\times\mathbf{B} = -\boldsymbol{\epsilon}^{-1}\cdot\mathbf{D}. \tag{A5}$$

Equations (6) and (A5) are dual equations. This means that the dispersion relation for **v** (the wave surface) can be obtained from the slowness surface by the following substitutions:

$$\mathbf{s} \to \mathbf{v}, \quad \mu_0 \to \mu_0^{-1}, \quad \boldsymbol{\epsilon} \to \boldsymbol{\epsilon}^{-1}.$$

The duality property also implies

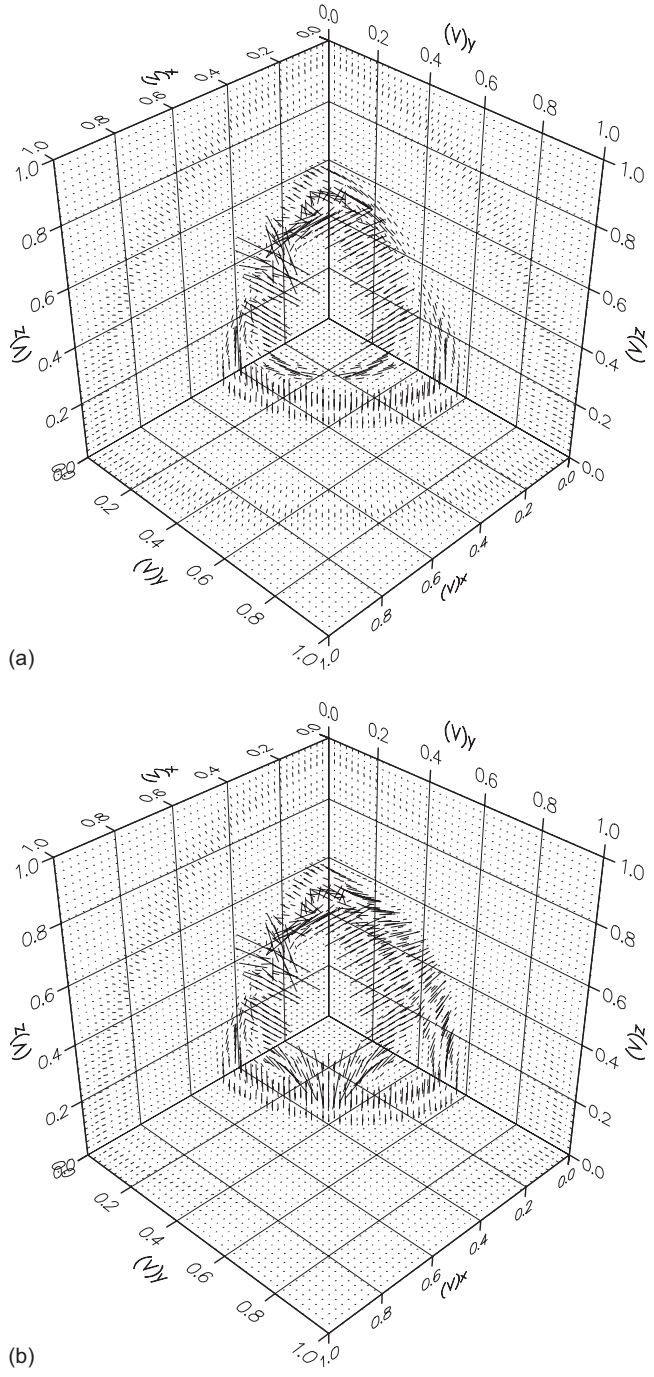


FIG. 5. Snapshots of the polarization vector for the normally (a) and the anomalously (b) polarized media.

$$\mathbf{s}\cdot\mathbf{E}=0,\quad \mathbf{s}\cdot\mathbf{H}=0,$$

and that **D** and **B** are tangent to the wave surface.

[1] G. Green, "On the laws of the reflection and refraction of light at the common surface of two non-crystallized media," J. Chem. Phys. **7**, 1–24 (1842).

[2] J. M. Carcione and F. Cavallini, "On the acoustic-electromagnetic analogy," Wave Motion **21**, 149–162 (1995).

[3] J. M. Carcione, "On energy definition in electromagnetism: An analogy with viscoelasticity," J. Acoust. Soc. Am. **105**, 626–632 (1999).

[4] J. M. Carcione and E. Robinson, "On the acoustic-electromagnetic analogy for the reflection-refraction problem," Stud. Geophys. Geod. **46**, 321–345 (2002).

[5] A. Fresnel, "Extrait d'un Mémoire sur la Double Réfraction," *Oeuvres complétes d'Augustin Fresnel* (Imprimerie Impérial, Paris, 1868), pp. 261–269. English translation: A. Fresnel, "Excerpt of a Memory on Double Refraction," *Complete works of Augustin Fresnel*, (Imperial Printing Office, 1868), pp. 261–309.

[6] M. Born and M. E. Wolf, *Principle of Optics* (Cambridge University Press, Cambridge, 1999).

[7] J. M. Carcione, *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic, Porous and Electromagnetic Media*, 2nd ed. (Elsevier, Amsterdam, 2007).

[8] K. Helbig, *Foundations of Anisotropy for Exploration Seismics* (Pergamon, Oxford, 1994).

[9] J. M. Carcione and M. Schoenberg, "3-D ground-penetrating radar simulation and plane wave theory," Geophysics **65**, 1527–1541 (2000).

[10] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. **I**.

[11] M. P. Rudzki, "Parametric representation of the elastic wave in anisotropic media," Anzeiger Akad. Wiss. Krakau 503–536 (1911).

[12] T. C. T. Ting, "Anisotropic elastic materials that have one or two sheets of spherical slowness surface," Proc. R. Soc. London, Ser. A **462**, 3133–3149 (2006).

[13] J. Rychlewski, "Elastic waves under unusual anisotropy," J. Mech. Phys. Solids **49**, 2651–2666 (2001).

[14] N. I. Ostrosablin, "Elastic anisotropic material with purely longitudinal and transverse waves," J. Appl. Mech. Tech. Phys. **44**, 271–278 (2003).

[15] J. MacCullagh, Trans. R. Irish Acad. **21** (1848).

[16] E. T. Whittaker, *The History of Modern Physics* (Tomash, New York, 1987), Vol. **7**.

[17] A. L. Cauchy, Compt. Rend. **9**, 676 (1839); **9**, 726 (1839).

[18] K. Helbig and M. Schoenberg, "Anomalous polarization of elastic waves in transversely isotropic media," J. Acoust. Soc. Am. **81**, 1235–1245 (1987).

[19] J. M. Carcione and K. Helbig, "Wave polarization in transversely isotropic and orthorhombic media," in *Proceedings of the Seventh International Workshop on Seismic Anisotropy*, Miami, FL, 1996, pp. 289–232.

[20] J. M. Carcione, D. Kosloff, A. Behle, and G. Seriani, "A spectral scheme for wave propagation simulation in 3-D elastic-anisotropic media," Geophysics **57**, 1593–1607 (1992).

[21] G. Salmon, *A Treatise on the Analytical Geometry of Three Dimensions* (Longmans, Green, London, 1915).

Surface wave conversion analysis on a lengthwise soldered circular cylindrical shell

André Baillard, Jeremiah Chiumia, Dominique Décultot, and Gérard Maze[a)]
Laboratoire d'Acoustique Ultrasonore et d'Electronique UMR CNRS 6068, Université du Havre, place Robert Schuman, 76610 Le Havre, France

Aleksander Klauson and Jaan Metsaveer
Department of Mechanics, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia

(Received 26 October 2007; revised 7 July 2008; accepted 24 July 2008)

This paper deals with wave conversion phenomena through a study of the acoustic scattering from a stiffened cylindrical shell at normal incidence. The analysis presented follows the experimental study which explored the acoustic wave propagation and scattering processes in the case of air-filled submerged cylindrical shells having internal axial solder [J. Chiumia, N. Touraine, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, J. Acoust. Soc. Am. **105**, 183–193 (1999)]. The significant observed phenomena were generation, reflection, and conversion of circumnavigating waves at the solder. The present work confirms the presence of the three phenomena through a theoretical approach based on the elasticity theory. In particular, resonances with new feature whose frequencies are very close to those of the S_0 wave are highlighted here. The origin of this new resonance feature is identified and can be associated to wave type conversions between A_0 and S_0 Lamb waves, occurring when these propagating waves encounter the solder. The air-filled stainless steel studied tube submerged in water has an internal lengthwise solder, which is, in theoretical computations, considered as an internal axial mass layer. The reduced frequency range (k_1a) of the study is between 25 and 90, (k_1: wave number in water; a: external radius of the shell).
 [DOI: 10.1121/1.2973233]

PACS number(s): 43.40.Ey, 43.20.Fn, 43.20.Tb, 43.40.Fz [EGW] Pages: 2061–2067

I. INTRODUCTION

Acoustic scattering from objects as plates, cylinders, or cylindrical shells has been widely analyzed both theoretically and experimentally.[1–4] Several studies of surface wave propagation on plates have shown that discontinuities in the geometry do not cause wave reflections only but also wave type conversions.[5–9] In particular, authors of Ref. 5 have examined the interaction of the A wave with a tenon or a mortise and have shown the conversion of this wave into A_0, S_0, and A_1 Lamb waves. In our case, the discontinuity is caused by an internal solder situated along a generating line of cylindrical shells and can be represented as an internal added mass. Acoustic scattering from such cylindrical shells, with a lengthwise solder, is analyzed in the perspective of further studies on more complex objects, in which solders are used to attach internal structures onto shells.[10–14]

In a previous paper,[15] the authors have studied the influence of the solder on the scattered sound field, in time and frequency domains. In the time domain, the observed echo series has been explained considering the solder line as a site of excitation, scattering, reflection, and conversion of waves. Different combinations of these four phenomena in the determination of echo arrival times, in addition to classical generation and remission of waves, have been taken into account. In the frequency domain, the presence of the solder modifies angular patterns on the backscattering spectra. The authors have highlighted the enhancement of resonance amplitudes on the resonance spectra, due to the presence of the solder, which depends on the incident wave direction. They have then shown that S_0 wave resonances are observable in the frequency band $45 < k_1a < 60$, whereas they are not observable in that range on the unsoldered shell of the same characteristics.

The present paper complements this experimental analysis with a theoretical approach using elasticity theory.[12] Considering the frequency range of the study, where wavelengths are close to the shell thickness, the use of this model is fully justified and can be applied for any thickness, contrary to thin shell theories.

Section II outlines the theoretical basis used to describe the acoustic scattering from a fluid-loaded stiffened cylindrical shell. More accurately, cylindrical shell displacements and stresses expressed from the elasticity theory are coupled to those of the stiffener at the junction. In the present case, the internal mass layer influence is expressed by inertial reactions. Further, wave conversion phenomena are theoretically considered in this model.

In Sec. III, spectral analysis reveals new resonances, very close to those of the S_0 wave. It is shown that these resonances are due to the wave conversion phenomenon between A_0 and S_0 Lamb waves at the stiffener. This phenomenon is clearly explained in Sec. IV in the time domain analysis, and more particularly in the frequency-wave number representation.

[a)]Author to whom correspondence should be addressed. Electronic mail: gerard.maze@univ-lehavre.fr

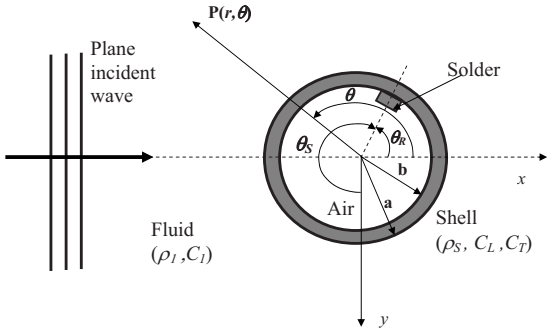


FIG. 1. Geometry of the problem.

II. THEORETICAL STUDY

An elastic cylindrical shell stiffened by an internal lengthwise mass layer is considered. The shell, made of a homogenous and isotropic material, is immersed in an ideal fluid medium and fluid-loading inside the object is assumed to be negligible. The thickness of the shell is represented by its radius ratio b/a (a: outer radius, b: inner radius). The incident plane wave front is parallel to the axis of the infinite circular cylindrical shell. The geometry of the given two-dimensional problem is illustrated in Fig. 1, where polar coordinates (r,θ) are used. θ_R is the angular position of the solder in relation to the incidence direction, θ_S is the angular position of the solder defined in Ref. 15.

The scattering pressure P_{scat} from this object can be decomposed into angular harmonics in θ for each reduced frequency $k_1 r$, and written as

$$P_{\text{scat}} = P_0 \sum_{n=-\infty}^{n=+\infty} i^n a_{1n} H_n^{(1)}(k_1 r) e^{in\theta}, \tag{1}$$

where $H_n^{(1)}$ is the nth order Hankel function of the first kind, k_1 is the wave number in water, and P_0 is a constant. As the problem is linear, the time function $e^{-i\omega t}$ has been omitted. In the case of the unstiffened cylindrical shell, the unknown coefficients a_{1n} represent contributions of each mode n.

The solution to this problem is described in Ref. 12. It is obtained by considering displacements and stresses in both the fluid and the shell, which are expressed in the theory of elasticity by scalar and vector potentials. The influence of the stiffener is taken into account with the normal and tangential stress expressions, respectively, σ_{rr} and $\sigma_{r\theta}$ at the inner surface of the shell $(r=b)$;

$$\sigma_{rr} = -\delta(\theta-\theta_R) F^N u_r(b,\theta_R), \tag{2}$$

$$\sigma_{r\theta} = -\delta(\theta-\theta_R) F^T u_\theta(b,\theta_R), \tag{3}$$

where δ is the Dirac delta function; F^N and F^T are the normal and tangential reactions of the stiffener for unitary displacements; and $u_r(b,\theta_R)$ and $u_\theta(b,\theta_R)$ are the normal and tangential displacements of the shell at the stiffener position, respectively.

Contrary to the study in Ref. 12, where the cylindrical shell was stiffened by an elastic internal plate, in this case, it is stiffened by a simple mass layer, and the reactions to unitary displacements correspond to inertial forces only

TABLE I. Material parameters.

Parameter	Stainless steel	Water
Density (kg/m^3)	7900	1000
Longitudinal velocity (m/s)	5790	1470
Transversal velocity (m/s)	3100	

$$F^N = F^T = m\omega^2, \tag{4}$$

where ω is the angular frequency and m the mass per unit length of the stiffener.

The unknown modal coefficients $\mathbf{a}_n = (a_{1n} a_{2n} a_{3n} a_{4n} a_{5n})^t$ of the scattered pressure and of the vector and scalar potentials are solutions of the equation

$$\mathbf{D}_n \mathbf{a}_n = \mathbf{b}_n, \tag{5}$$

where $\mathbf{D}_n$ is a 5×5 matrix obtained from the boundary conditions. The characters in bold denote vectors and matrix. Expressions of $\mathbf{D}_n$ are given in Ref. 12 and $\mathbf{b}_n$ in the Appendix, where the reactions are those given by Eq. (4).

To solve Eq. (5), $\mathbf{b}_n$ have to be calculated; it depends on the normal and tangential displacements of the shell at the stiffener position as it is shown in Eqs. (2) and (3). So, the first step of the calculation is the determination of these displacements, as described in detail in Ref. 12. The displacements are then calculated from the following systems:

$$u_r(b,\theta_R) = u_r^0(b,\theta_R) + \alpha S^N F^N u_r(b,\theta_R), \tag{6}$$

$$u_\theta(b,\theta_R) = u_\theta^0(b,\theta_R) + \alpha S^T F^T u_\theta(b,\theta_R), \tag{7}$$

where $u_r^0(b,\theta_R)$ and $u_\theta^0(b,\theta_R)$ are the displacements of the unstiffened shell. These terms as well as α and the modal sums S^N and S^T are also developed in Ref. 12, where they are denoted as S_1^N and S_1^T.

The coefficients of $\mathbf{b}_n$ are now determined from Eqs. (6) and (7), and modal coefficients a_{1n} can then be calculated. But contrary to the case of the unstiffened shell, where a_{1n} depends only on the considered mode n, this coefficient now depends on the shell displacements which are functions of all the vibration modes. In the case of a stiffened shell, all modes are coupled to each other and this shows the possibility of wave conversions phenomena in this model.

III. SPECTRAL ANALYSIS

The studied object is a stainless steel cylindrical shell. It is characterized by its radius ratio $b/a=0.97$ and its outer radius $a=15.3$ mm. The internal lengthwise stiffener is made of the same material whose characteristics are given in Table I.

The thickness and the width of the mass layer are identical to the thickness of the shell.

Figure 2 compares the experimental backscattering spectrum (a) and the computed backscattering spectrum (b) calculated from the theoretical approach described in Sec. II.

Comparison between calculated and experimental results (Fig. 2) allows us to make some remarks.

(i) The influence of the A wave (Scholte–Stoneley wave) is observed for frequencies below $k_1 a=45$, as in the

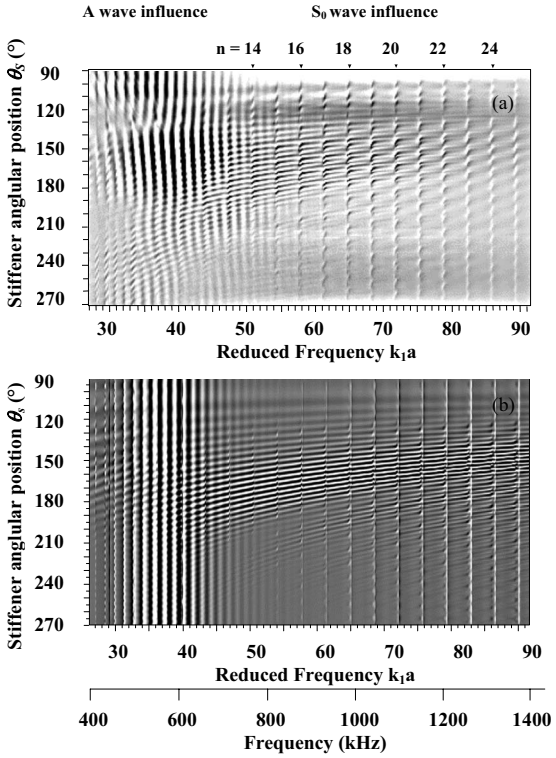


FIG. 2. (a) Experimental backscattering spectra (Ref. 15). (b) Calculated backscattering spectra. (Amplitude increases from black to white in linear scale.) n is the mode number of S_0 wave.

case of an unstiffened shell. Resonances obtained from calculations have a relatively constant amplitude on the whole angular window [Fig. 2(b)], contrary to experimental results, where resonance amplitudes are lower in the shadow zone [Fig. 2(a)], where the stiffener is not directly insonified ($180° < \theta_S < 270°$). For higher frequencies ($k_1a > 40$), no resonance of the A wave is detected, as its energy is strongly dissipated into the surrounding fluid.

(ii) Transitions due to S_0 resonances are detected on both calculated and experimental spectra. They appear in Fig. 2 as regularly spaced vertical lines, situated at the same frequencies and with the same mode numbers as for the unstiffened shell. Figure 3 presents experimental and calculated backscattering spectra at $\theta_S = 170°$ (Fig. 2).

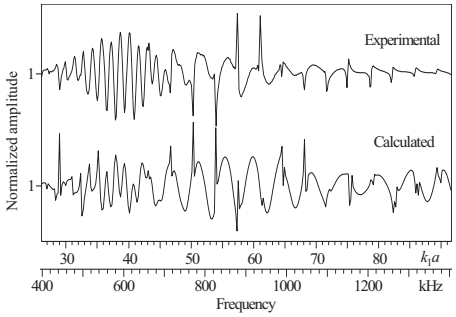


FIG. 3. Examples of experimental and calculated backscattering spectra $\theta_S = 170°$.

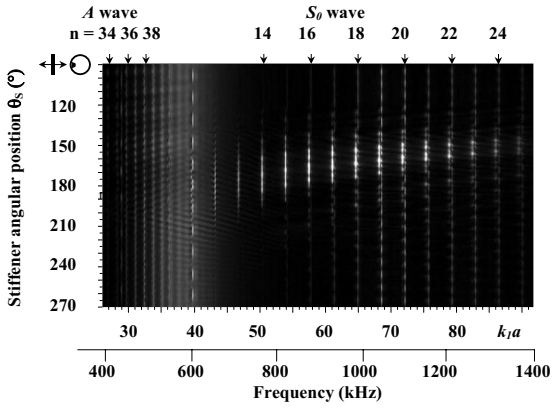


FIG. 4. Calculated resonance spectra of the stiffened shell. (Amplitude increases from black to white in linear scale.) n is the mode number of A and S_0 waves.

A modulation is also observed on these curves. It is explained by the interference phenomenon between two echoes in the time domain responses: the specular echo and the reflection on the solder line.[13,15,16] It has been shown that small variations of the incidence angle cause significant modifications of the frequency position on this modulation.[13,15] The determination of the experimental incident angle is not accurate enough to achieve perfect correspondence between experimental and calculated results of this modulation.

In order to avoid this modulation effect, resonance spectra are calculated by fast Fourier transform from time domain signals in which the first direct reflection echo (the specular echo) is omitted. The corresponding results are presented in Fig. 4.

Resonances previously observed are still clearly detected, in a good agreement with experimental results.[15] At incidence $\theta_S = 90°$, the stiffener is situated at the point of first contact of the incident wave with the shell. In this case, the spectral response is the same as that obtained with the unstiffened shell (Fig. 5).

In particular, it is noted in both cases that the peaks due to S_0 wave resonances are too narrow to be observable in the frequency zone around $k_1a = 45$. In this frequency zone, the coupling between the fluid and the shell is too weak for the

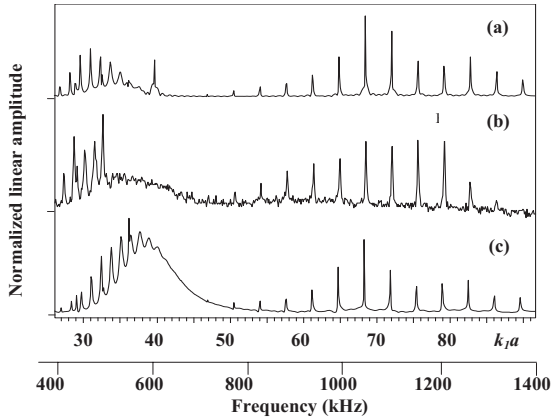


FIG. 5. (a) Calculated resonance spectrum of the unstiffened shell; (b) experimental and (c) calculated resonance spectra of stiffened shell for solder at the position $\theta_S = 90°$.

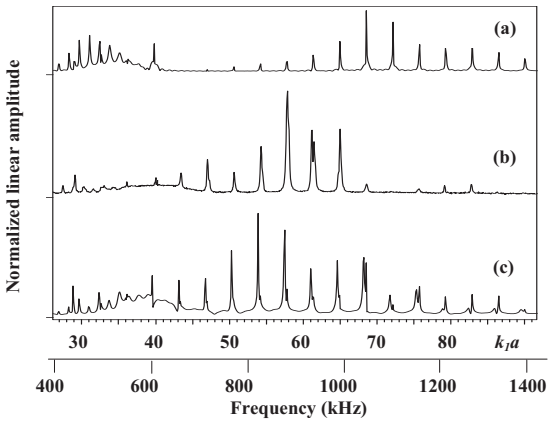


FIG. 6. (a) Calculated resonance spectrum of the unstiffened shell; (b) experimental and (c) calculated resonance spectra of stiffened shell for solder at the position $\theta_S=170°$.

S_0 wave such that it is difficult to generate this wave on the shell. A dominant feature in Fig. 4 is the presence of significant peaks observed only in the angular zone $210° > \theta_S > 135°$. These peaks are also detected in the frequency zone mentioned previously, as shown in both experimental and calculated spectra presented in Fig. 6 ($\theta_S=170°$). The authors of Ref. 15 have considered these resonances as the enhancement of S_0 resonance amplitudes, due to the wave conversion phenomenon. However, these peaks do not correspond to usual S_0 resonances.

Figure 7 presents the enlarged calculated spectra [Fig. 6(c)] in which it is possible to observe the new features of resonances: the smaller peaks correspond to S_0 wave resonances (narrow arrows in Fig. 7) on the unstiffened shell (identical frequencies), whereas the larger peaks are detected as resonances related to wave conversions. The latter peaks are shifted slightly in comparison to the former peaks. This shift is not easily detectable in experimental responses. This feature is very sensitive to the mass of the layer.[13] The gaps between two successive resonances of the new resonance feature are identical to those between S_0 resonances, suggesting that the phenomenon at the origin of these resonances is associated to the S_0 wave propagation. More precisely, as will be shown in the next section, it can be associated to the

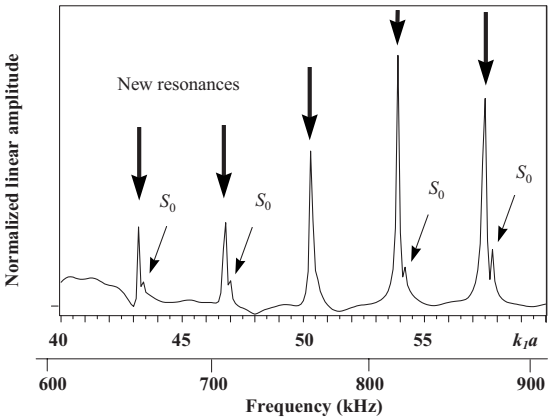


FIG. 7. Details of Fig. 6(c) (calculated resonance spectrum of stiffened shell for solder at the position $\theta_S=170°$). Small arrows indicate S_0 waves resonances. Big arrows indicate resonances related to the wave conversion.

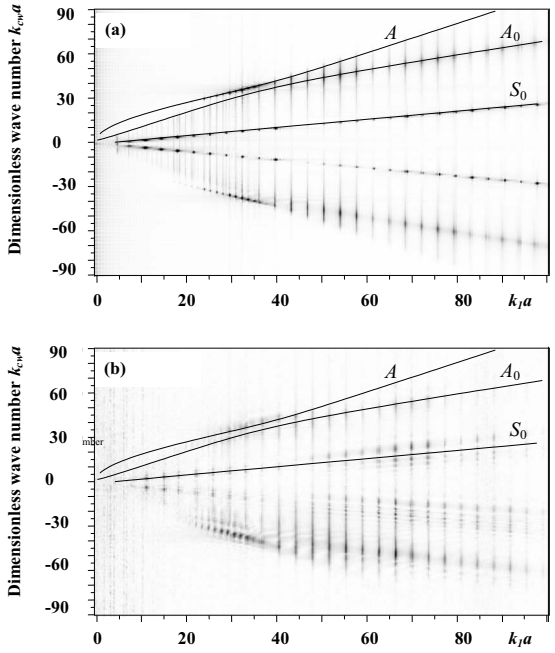


FIG. 8. (a) Calculated and (b) experimental frequency-wave number representations. Amplitude increases from white to black in linear scale. Continuous lines correspond to theoretical wave numbers calculated in the case of the unstiffened shell.

wave conversion phenomenon between A_0 and S_0 waves. In fact, the A_0 wave is not observable in spectral responses because its energy is highly reradiated into the fluid, nevertheless it is present and propagates in the shell.[17,18]

IV. WAVE CONVERSIONS

The authors of Ref. 15 have considered the large peaks mentioned earlier as the enhancement of the S_0 wave resonances. In particular, they have explained that the A wave propagating on the shell can be converted at the solder into S_0 wave. This phenomenon has been put in evidence in the time domain, where echoes observed in the responses are identified with the prediction of echo arrival times. In search of possible wave paths, Chiumia *et al.*[15] have taken into account only the A wave and the S_0 wave because these are the only two waves experimentally detected in the studied frequency range. In fact, these are the only resonances observed in their spectral responses. However the A_0 wave should also be considered if the structure has impedance discontinuity as it is shown in the study presented in Ref. 17. Furthermore, in Ref. 18, the propagation of the A_0 wave on a stiffened shell has clearly been observed.

A. Frequency-wave number domain analysis

The "surface wave analysis method" allows us to identify waves related to the observed resonances in the frequency-wave number plane.[19,20] These presentations obtained by spatial Fourier transform of resonance spectra calculated in bistatic setup (the emitter is fixed and the receiver turns around the object recording responses at different angle θ) and are presented in Fig. 8 for both cases, calculated [8(a)] and experimental [8(b)].

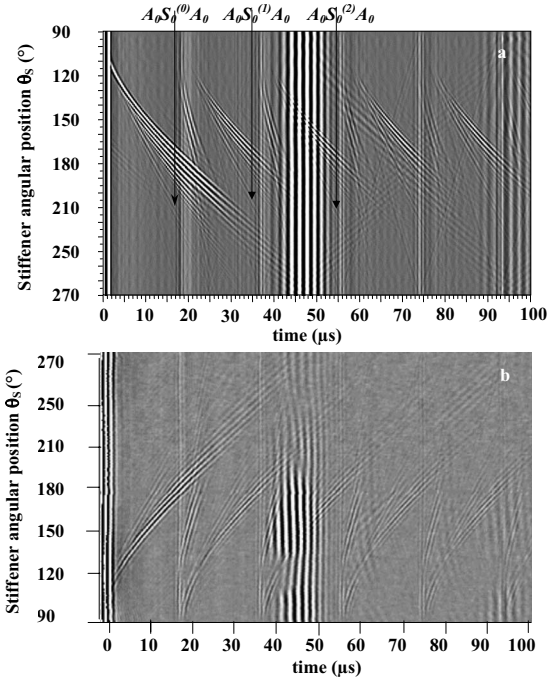


FIG. 9. (a) Calculated and (b) experimental (Ref. 15) time signals, (amplitude increases from black to white in linear scale).

In this study, the fixed position of the stiffener is at $\theta_S = 180°$. In order to identify the propagations involved, theoretical curves calculated for the unstiffened shell are also plotted in Fig. 8 (continuous lines). Resonances of the A wave are detected, in both calculated and experimental representations, in the frequency window mentioned earlier ($20 < k_1a < 40$). The S_0 wave resonances are observed over the whole frequency range, except around $k_1a = 45$, where the peaks are very narrow and thus hardly observable. The new resonances which appear in the spectral responses, in this frequency domain and particularly in higher frequencies, can then be associated to the A_0 wave propagation; this is shown in calculated [Fig. 8(a)] as well as in experimental representations [Fig. 8(b)], with peaks corresponding to the new resonances situated on the A_0 wave curve. The frequency-wave number representation clearly illustrates the wave conversion phenomenon between the S_0 and A_0 waves, where resonances of the former are detected and identified with wave numbers of the latter. We note that the peaks are wide (mode conversions) but well-centered on the curve of the A_0 wave numbers.

B. Time domain analysis

The phenomenon at the origin of these resonances is associated to the propagation of S_0 and A_0 waves, through wave conversions, and it can be determined through time domain analysis.

Figure 9(a) presents the backscattered time echoes calculated from the theoretical approach developed in Sec. II. The time signals are complex, nevertheless the observed echo series are in good agreement with experimental echo series presented in Fig. 9(b). All echoes experimentally identified by Chiumia *et al.*[15] are detected in the calculated responses. Now taking into account the A_0 wave, new echoes

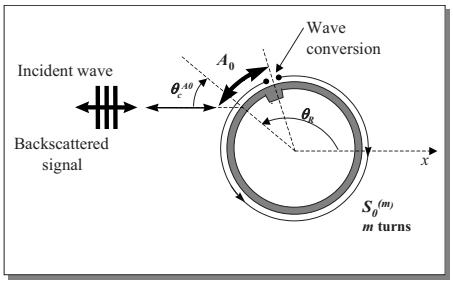


FIG. 10. Wave travel path description of echoes denoted $A_0S_0^{(m)}A_0$.

can be identified and interpreted in both experimental and calculated time responses. Let us consider the wave propagation scheme given in Fig. 10.

The A_0 wave is generated on the shell at its critical angle θ_c^{A0} (coupling angle). It propagates to the stiffener where it is converted into the S_0 wave which circumnavigates the shell, with weak dissipation. At each turn, as this wave encounters the stiffener, a part of its energy is reconverted into A_0 wave, which is strongly reemitted into the fluid. The echo arrival time of this wave propagation, noted $A_0S_0^{(m)}A_0$ is determined from the wave group velocities on the shell and is therefore given by

$$t = 2(a(1 - \cos\theta_c^{A0})/C_1 + a(\pi - \theta_c^{A0} - \theta_R)/C_g^{A0}) + 2m\pi a/C_g^{S0}, \quad (8)$$

with $m = 1, 2, \ldots$ and $0 \leqslant \theta_R \leqslant \pi - \theta_c^{A0}$, where $t = 0$ corresponds to the specular reflection arrival time and is taken as a reference time. Angles are expressed in radians. The number of turns traveled by the S_0 wave is denoted m, and the group velocity of the S_0 wave is $C_g^{S0} = 5200$ m/s. Concerning the A_0 wave, the group velocity C_g^{A0} is between 2000 and 3100 m/s (higher than sound velocity in water[16]) and the critical angle θ_c^{A0}, which depends on the phase velocity, decreases from 90° to 45°. In the lower frequencies, these values are close to those of the A wave, so, in order to make distinction between the two waves, a high value of group velocity is used to evaluate t: $C_g^{A0} = 3100$ m/s and $\theta_c^{A0} = 45°$.

The evolution of this calculated arrival time is presented in Fig. 11 by continuous lines. In addition the evolution of

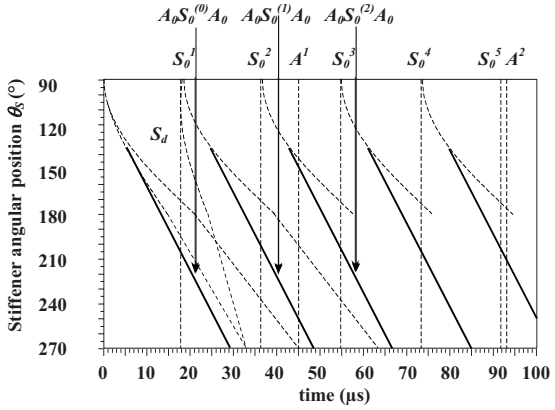


FIG. 11. Evolution of calculated echo arrival times. Bold lines correspond to $A_0S_0^{(m)}A_0$ echoes. Dashed lines correspond to analysis given in Ref. 15.

arrival times made in previous paper[15] is plotted in dashed lines.

The echo due to direct reflection of the incident wave on the stiffener is noted S_d, and the vertical lines correspond to echoes associated to A and S_0 waves generated and reemitted at respective critical angles (echoes independent of the stiffener). The first echo $A_0S_0^{(0)}A_0$ reappears several times on the curve and the time interval between two successive echoes of the series $A_0S_0^{(m)}A_0$ is equal to the time the S_0 wave takes to cover a circumference of the shell. These repetitions in the time domain correspond, in the frequency domain, to the interval between two successive resonances equal to that of the S_0 resonances (Fig. 2). Note in Fig. 11, that echoes coming from such wave path are not present when θ_S is lower than 135° because in this angular zone the A_0 wave is generated on the shell after the incident plane wave encounters the stiffener. This condition explains the upper limit of the angular zone where new resonances are observed (Fig. 4). The lower limit of this angular zone depends on the dissipation in the fluid of the two waves, especially that of the A_0 wave: The closer θ_S tends to 270° (the longer the path gets) the higher is the wave energy dissipation. This remark also applies to the time domain signals (Fig. 9). Furthermore, the shape of these echoes is characteristic of the A_0 wave propagation, considering the dispersion curve of the wave group velocity: The wave is faster at higher frequencies than at lower ones.[18]

In other words, the amplitudes of the A_0 wave peaks in the frequency-wave number representations (Fig. 8) are higher when those of the S_0 wave are lower. In fact, in this particular frequency zone, resonances of S_0 wave are difficult to detect by the usual reemission of its energy into the fluid (as in the case of the unstiffened shell). However, the S_0 wave can be generated through the conversion of the A_0 wave, at the stiffener, and then propagates around the shell.

The echoes noted $A_0S_0^{(m)}A_0$ are the most important manifestation of the conversion between S_0 and A_0 waves. This study complements the previous analysis presented in Ref. 15 by taking into account the A_0 wave as the origin of wave conversion.

V. CONCLUSION

The analysis of the acoustic scattering from stiffened circular cylindrical shells is performed in this paper through a theoretical approach using elasticity theory. In all the cases examined, backscattering and resonance spectra, time domain responses and frequency-wave number representations, the calculated results are in a good agreement with experimental results presented in a previous paper.[15]

In spectral responses, a new feature of resonances is observed for higher frequencies ($k_1a>45$) and for incident angles around $\theta_S=180°$. A large peak of resonance is observed near the S_0 wave resonances of unstiffened shell. These peaks present a shift in frequency which depends on the mass of the stiffener. These resonances are clearly associated to S_0 wave resonances, but not from the usual generation of this wave on the shell: First the A_0 wave is generated on the shell at its critical angle, then it is converted, at the stiffener, into S_0 wave which circumnavigates the shell. At each encounter with the stiffener, a part of its energy is reconverted into A_0 wave which is reemitted into the fluid. Further, the new feature of resonances depends on the characteristics of the A_0 wave. Resonances are observed only when the A_0 wave is generated for incidences just before it encounters the stiffener, however, not too far (due to its strong dissipation). This explains the particular angular zone where these new resonances are detected.

The wave conversions between S_0 and A_0 waves are clearly illustrated in frequency-wave number representations, with peaks close to S_0 wave resonance frequencies but situated near the wave numbers corresponding to the A_0 wave.

This study explains some wave propagation phenomena observable on stiffened structures and confirms the importance of the influence of soldering in complex assembled geometric objects.

APPENDIX: COEFFICIENTS OF b_n

The coefficients $\mathbf{b}_n$ corresponding to the left side term of Eq. (5) are decomposed in three parts

$$\mathbf{b}_n = \mathbf{b}_n^0 + \mathbf{b}_n^N + \mathbf{b}_n^T,$$

with

$$b_{1n}^0 = -x_T^2 J_n(x_1),$$

$$b_{2n}^0 = x_1 \frac{\rho S}{\rho_1} J_n'(x_1),$$

$$b_{3n}^0 = b_{4n}^0 = b_{5n}^0 = 0,$$

$$b_{1n}^N = b_{2n}^N = b_{3n}^N = b_{5n}^N = 0,$$

$$b_{4n}^N = i^{-n} f_1^N e^{-in\theta_R},$$

$$b_{1n}^T = b_{2n}^T = b_{3n}^T = b_{4n}^T = 0,$$

$$b_{5n}^T = i^{-n} i f_1^T e^{-in\theta_R},$$

where the following notations are used:

$$x_j = k_j a,$$

$$j = 1, L, T;$$

$$y_j = k_j b,$$

$$j = 1, L, T;$$

$$f_1^N = \frac{y_T^2}{2\pi b} F_l^N u_r(b, \theta_R),$$

$$f_1^T = \frac{y_T^2}{2\pi b} F_1^T u_\theta(b, \theta_R).$$

[1]H. Lamb, "On waves in an elastic plate," Proc. R. Soc. London, Ser. A **93**, 114–128 (1917).
[2]J. J. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust.

Soc. Am. **23**, 405–418 (1951).
[3]M. C. Junger, "Sound scattering by thin elastic shells," J. Acoust. Soc. Am. **24**, 366–373 (1952).
[4]J.-L. Izbicki, G. Maze, and J. Ripoche, "Influence of the free modes of vibration on the acoustic scattering of circular shells," J. Acoust. Soc. Am. **80**, 1215–1219 (1986).
[5]A. Tinel, H. Duflo, and J. Duclos, "Onde de type Scholte sur une plaque fine. Interaction avec un défaut de surface," (Scholte wave on a thin plate. Interaction with a surface defect) 4ème Congrès Français d'Acoustique (4th French Acoustic Congress), Marseille, France, 1997, Conference Proceedings, 889–892 (in French).
[6]R. Briers, O. Leroy, and G. N. Shkerdin, "Conversion of a Stoneley wave at the extremity of a fluid loaded plate," J. Acoust. Soc. Am. **101**, 1347–1357 (1997).
[7]A. Tinel, "Diffraction de l'onde de Scholte par un dièdre et par un réseau de stries," (Diffraction of Scholte wave by a dihedron and a groove network) Ph.D. University of Le Havre, 1991 (in French).
[8]El-Kettani M. Ech-Cherif, Pareige, P., Luppé, F., and Ripoche, J., "Experimental Study of the Conversion of Lamb Waves at the End of an Immersed Plate," Acust. Acta Acust. **82**, 251–259 (1996).
[9]J. J. Ditri, "Some results on the scattering of guided elastic SH waves from material and geometric waveguide discontinuities," J. Acoust. Soc. Am. **100**, 3078–3087 (1996).
[10]A. Klauson, J. Metsaveer, D. Décultot, G. Maze, and J. Ripoche, "Identification of the resonances of the cylindrical shell stiffened by an internal lengthwise rib," J. Acoust. Soc. Am. **100**, 3135–3143 (1996).
[11]A. Baillard, J.-M. Conoir, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, "High frequencies experimental study on time domain of backscattered signal from stiffened cylindrical shell," Acoust. Lett. **22**, 61–65 (1998).
[12]A. Baillard, J.-M. Conoir, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, "Acoustic scattering from fluid-loaded stiffened cylindrical shell: Analysis using elasticity theory," J. Acoust. Soc. Am. **107**, 3208–3216 (2000).
[13]A. Baillard, "Influence de raidisseurs longitudinaux sur la diffusion acoustique par un tube: Théorie et expérience," (Influence of lengthwise ribs on the acoustic scattering from a tube: Theory and Experiment), Ph.D., University of Le Havre, 2000 (in French).
[14]Y. P. Guo, "Sound scattering from cylindrical shells with internal elastic plates," J. Acoust. Soc. Am. **93**, 1936–1946 (1993).
[15]J. Chiumia, N. Touraine, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, "Experimental analysis of acoustic scattering from lengthwise soldered cylindrical shells," J. Acoust. Soc. Am. **105**, 183–193 (1999).
[16]G. Maze, N. Touraine, A. Baillard, D. Décultot, V. Latard, L. Derbesse, P. Pernod, and A. Merlen, "A_0-wave and *A*-wave in cylindrical shell immersed in water: Influence on the acoustic scattering," 17th ASME Design Engineering Technical Conference, Las Vegas, Nevada, 1999, Conference Proceedings on CD-ROM DETC99/VIB-8090.
[17]S. F. Morse, P. L. Marston, and G. Kaduchak, "High frequency backscattering enhancements by thick finite cylindrical shell in water at oblique incidence: Experiments, interpretation, and calculations," J. Acoust. Soc. Am. **103**, 785–794 (1999).
[18]A. Baillard, J.-M. Conoir, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, "A_0 wave in a submerged stiffened cylindrical shell," 17th ASME Design Engineering Technical Conference, Las Vegas, Nevada, 1999, Conference Proceedings on CD-ROM DECT99/VIB8092.
[19]L. Martinez, J. Duclos, and A. Tinel, "New tools in acoustics: Transient Surface Wave Analysis Method. Study of Lamb waves with positive or negative group velocity in an immersed shell," EAA/EEAA Symposium Transport Noise and Vibration, Tallinn, Estonia, 1998, Symposium proceedings, pp. 281–284.
[20]L. Martinez, J. Duclos, and A. Tinel, "Aspects of cylindrical shell resonances in the Fourier diamond spaces. New Surface Wave Analysis Methods (S.W.A.M.)," 16th International Congress on Acoustics and 135th Meeting of Acoust. Soc. Am., Seattle, Washington, 1998; J. Acoust. Soc. Am. **103**, 2901 (1998).

Vibration of fluid loaded conical shells

Mauro Caresta and Nicole J. Kessissoglou
School of Mechanical and Manufacturing Engineering, The University of New South WalesT, Sydney NSW 2052, Australia

(Received 3 April 2008; revised 25 July 2008; accepted 25 July 2008)

An analytical model is presented to describe the vibration of a truncated conical shell with fluid loading in the low frequency range. The solution for the dynamic response of the shell is presented in the form of a power series. Fluid loading is taken into account by dividing the shell into narrow strips which are considered to be locally cylindrical. Analytical results are presented for different boundary conditions and have been compared with the computational results from a boundary element model. Limitations of the model to the low frequency range are discussed.
 [DOI: 10.1121/1.2973237]

PACS number(s): 43.40.Ey, 43.20.Tb [EGW] Pages: 2068–2077

I. INTRODUCTION

Conical shells are used in engineering applications such as spacecraft fuselages and submarine hulls. For example, a truncated cone is used to support the aft shaft bearing for a submarine propeller. Interaction with an acoustic medium such as water can significantly alter the dynamic behavior of the shell. The dynamic characteristics and responses of conical shells and in particular, of fluid loaded conical shells, have not been as widely reported in the literature as in the case of cylindrical shells. This is due to the increased mathematical complexity associated with the effect of the variation of the radius along the length of the cone on the elastic waves. The free vibration of conical shells *in vacuo* has been studied by many researchers, resulting in various modifications to the equations of motion due to simplifications associated with thin shell or shallow shell theory. Much of the earlier work on the equations of motion for conical shells and forms of solution with different boundary conditions has been summarized by Leissa.[1] The approximate location of the natural frequencies has been found using the Rayleigh-Ritz method by several authors.[1–5] A transfer matrix approach was used by Irie *et al.*[6] to solve the free vibration of conical shells. Tong[7] presented a procedure for the free vibration analysis of isotropic and orthotropic conical shells in the form of a power series. Guo[8] studied the propagation and radiation properties of elastic waves in an *in vacuo* cone.

There has been very little work on conical shells under fluid loading. Guo[9] investigated fluid loading effects of waves on conical shells in which he applied a perturbation technique using the small vertex angle as a parameter. He showed that the amplitude and phase of the shear, compressional and flexural waves changed along the length of the cone. Guo[9] also showed that fluid loading has a significant effect on the flexural and compressional waves. Radiation damping is introduced to the shear and compressional waves due to their supersonic phase speeds. Norris and Rebinsky[10] examined thin shells under fluid loading, but the analysis was restricted to membrane waves.

In the present paper, the authors present a new approach to describe the vibration of a truncated conical shell under fluid loading in the low frequency range. Following the procedure delineated by Tong[7] for the *in vacuo* case, the solution for the dynamic response of a conical shell is presented in the form of a power series. Fluid loading is then taken into account by dividing the shell into narrow strips which are considered to be locally cylindrical. A local cylindrical approximation of the conical shell to model the effect of fluid loading has also been used by previous researchers,[9,10] but to the authors' knowledge, it has never been applied to the power series solution method. The natural frequencies of the fluid loaded cone are predicted using the numerical model and presented for different boundary conditions. Analytical results of the frequency response for the fluid loaded truncated cone are compared with computational results obtained from a boundary element model. The method presented in this paper provides an approachable solution to low frequency fluid loaded problems of conical shells.

II. MODELING OF A CONICAL SHELL

The equations of motion to describe the dynamic response of a conical shell under fluid loading can be written as

$$L_{11}u + L_{12}v + L_{13}w - \rho h\ddot{u} = 0, \tag{1}$$

$$L_{21}u + L_{22}v + L_{23}w - \rho h\ddot{v} = 0, \tag{2}$$

$$L_{31}u + L_{32}v + L_{33}w + \rho h\ddot{w} + p_a\frac{Eh}{(1-\nu^2)} = 0, \tag{3}$$

where u and v are, respectively, the displacements of the shell's middle surface along the x and θ directions. w is the displacement normal to the surface along the z direction. The geometrical parameters and coordinate system of the truncated conical shell are shown in Fig. 1. The partial differential operators L_{ij} are the same as those in Tong[7] for the *in vacuo* case. p_a is the external loading due to the fluid acting normally to the surface, as modeled by Junger and Feit[11] for fluid loaded cylindrical shells. E, ρ, ν, and h are, respectively, the Young's modulus, density, Poisson's ratio,

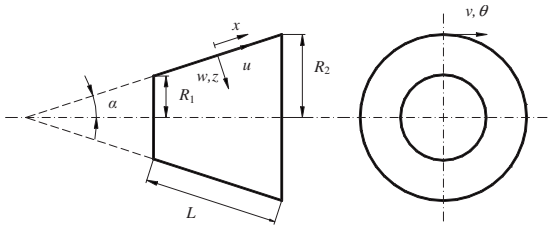


FIG. 1. Conical shell and coordinate system.

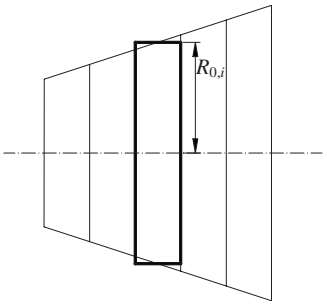


FIG. 2. Local cylindrical approximation of the conical shell.

and thickness of the shell. General solutions to the equations of motion can be expressed as

$$u(x,\theta,t) = u(x)\cos(n\theta)e^{-j\omega t}, \tag{4}$$

$$v(x,\theta,t) = v(x)\sin(n\theta)e^{-j\omega t}, \tag{5}$$

$$w(x,\theta,t) = w(x)\cos(n\theta)e^{-j\omega t}, \tag{6}$$

where n is the circumferential mode number. The x-dependent component of the displacement can be expressed in terms of a power series by

$$u(x) = \sum_{m=0}^{\infty} a_m x^m, \tag{7}$$

$$v(x) = \sum_{m=0}^{\infty} b_m x^m, \tag{8}$$

$$w(x) = \sum_{m=0}^{\infty} c_m x^m. \tag{9}$$

Substituting Eqs. (4)–(9) into Eqs. (1)–(3) results in the following recurrence relations[7] for $m=0,1,2,\ldots$:

$$a_{m+2} = G_{1,1}a_{m+1} + G_{1,2}a_m + G_{1,3}a_{m-1} + G_{1,4}a_{m-2} + G_{1,5}b_{m+1} + G_{1,6}b_m + G_{1,7}c_{m+1} + G_{1,8}c_m, \tag{10}$$

$$b_{m+2} = G_{2,1}a_{m+1} + G_{2,2}a_m + G_{2,3}b_{m+1} + G_{2,4}b_m + G_{2,5}b_{m-1} + G_{2,6}b_{m-2} + G_{2,7}c_m, \tag{11}$$

$$c_{m+4} = G_{3,1}a_{m+1} + G_{3,2}a_m + G_{3,3}a_{m-1} + G_{3,4}a_{m-2} + G_{3,5}b_m + G_{3,6}b_{m-1} + G_{3,7}b_{m-2} + G_{3,8}c_{m+3} + G_{3,9}c_{m+2} + G_{3,10}c_{m+1} + G_{3,11}c_m + G_{3,12}c_{m-1} + G_{3,13}c_{m-2} + G_{3,14}c_{m-3} + G_{3,15}c_{m-4}. \tag{12}$$

Tong[7] showed that the x-dependent part of the displacements can be expressed in terms of eight unknown coefficients $a_0,a_1,b_0,b_1,c_0,c_1,c_2,c_3$, which can be determined from the boundary conditions at both ends on the conical shell. In terms of the unknown coefficients, Eqs. (7)–(9) can be rewritten as follows

$$u(x) = \mathbf{u}\cdot\mathbf{x}, \quad v(x) = \mathbf{v}\cdot\mathbf{x}, \quad w(x) = \mathbf{w}\cdot\mathbf{x}, \tag{13}$$

where

$$\mathbf{u}[u_1(x) \ \cdots \ u_8(x)], \tag{14}$$

$$\mathbf{v} = [v_1(x) \ \cdots \ v_8(x)], \tag{15}$$

$$\mathbf{w} = [w_1(x) \ \cdots \ w_8(x)] \tag{16}$$

$$x = [a_0 \ a_1 \ b_0 \ b_1 \ c_0 \ c_1 \ c_2 \ c_3]^T. \tag{17}$$

In Eqs. (14)–(16), $u_i(x)$, $v_i(x)$, and $w_i(x)$ are the base functions of $u(x)$, $v(x)$ and $w(x)$ respectively.[7] The coefficients $G_{i,j}$ in Eqs. (10)–(12) are given in Appendix A.

III. FLUID LOADING MODEL

The fluid loading effect on the conical shell was taken into account using an approximate method which is described in what follows. The shell is divided into N narrow segments as shown in Fig. 2. The N conical parts are connected together by applying eight continuity equations at each interface. Expressions for continuity of the displacements and radial slope are given by

$$u_i = u_{i+1}, \quad i = 1,2,\ldots,N-1 \tag{18}$$

$$v_i = v_{i+1}, \quad i = 1,2,\ldots,N-1 \tag{19}$$

$$w_i = w_{i+1}, \quad i = 1,2,\ldots,N-1 \tag{20}$$

$$\frac{\partial w_i}{\partial x} = \frac{\partial w_{i+1}}{\partial x}, \quad i = 1,2,\ldots,N-1. \tag{21}$$

Equilibrium of the forces and moments results in

$$N_{x,i} = N_{x,i+1}, \quad i = 1,2,\ldots,N-1 \tag{22}$$

$$N_{x\theta,i} = N_{x\theta,i+1}, \quad i = 1,2,\ldots,N-1, \tag{23}$$

$$M_{x,i} = M_{x,i+1}, \quad i = 1,2,\ldots,N-1 \tag{24}$$

$$V_{x,i} = V_{x,i+1}, \quad i = 1,2,\ldots,N-1. \tag{25}$$

Forces and moments are given by[1]

$$N_x = \frac{Eh}{1-\nu^2}\left[\frac{\partial u}{\partial x} + \nu\left(\frac{\sin\alpha}{R(x)}u + \frac{1}{R(x)}\frac{\partial v}{\partial\theta} - \frac{\cos\alpha}{R(x)}w\right)\right], \tag{26}$$

$$N_\theta = \frac{Eh}{1-\nu^2}\left[\nu\frac{\partial u}{\partial x} + \left(\frac{\sin\alpha}{R(x)}u + \frac{1}{R(x)}\frac{\partial v}{\partial\theta} - \frac{\cos\alpha}{R(x)}w\right)\right], \tag{27}$$

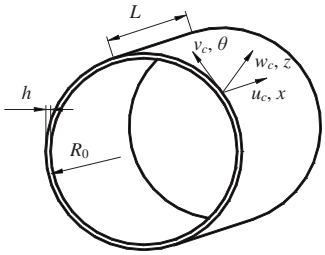


FIG. 3. Thin cylindrical shell and coordinate system.

$$N_{x\theta} = \frac{Eh}{2(1+\upsilon)}\left(\frac{1}{R(x)}\frac{\partial u}{\partial \theta} + \frac{\partial v}{\partial x} - \frac{\sin\alpha}{R(x)}v\right), \tag{28}$$

$$M_x = -\frac{Eh^3}{12(1-\upsilon^2)}\left[\frac{\partial^2 w}{\partial x^2} + \upsilon\left(\frac{\sin\alpha}{R(x)}\frac{\partial w}{\partial x} + \frac{1}{R^2(x)}\frac{\partial^2 w}{\partial \theta^2}\right)\right], \tag{29}$$

$$M_\theta = -\frac{Eh^3}{12(1-\upsilon^2)}\left(\upsilon\frac{\partial^2 w}{\partial x^2} + \frac{\sin\alpha}{R(x)}\frac{\partial w}{\partial x} + \frac{1}{R^2(x)}\frac{\partial^2 w}{\partial \theta^2}\right), \tag{30}$$

$$M_{x\theta} = -\frac{Eh^3}{12(1+\upsilon)}\frac{\partial}{\partial x}\left(\frac{1}{R(x)}\frac{\partial w}{\partial \theta}\right), \tag{31}$$

$$Q_x = \frac{1}{R(x)}\frac{\partial(R(x)M_x)}{\partial x} - \frac{M_\theta \sin\alpha}{R(x)} + \frac{1}{R(x)}\frac{\partial M_{x\theta}}{\partial \theta}, \tag{32}$$

$$V_x = Q_x + \frac{1}{R(x)}\frac{\partial M_{x\theta}}{\partial \theta}, \tag{33}$$

where N_x, N_θ, and $N_{x\theta}$ are the membrane forces, M_x, M_θ, and $M_{x\theta}$ are the bending moments, Q_x is the transverse shearing force and V_x is the Kelvin-Kirchhoff shear force.[1]

The segments are narrow enough to be considered as cylindrical in order to account for the fluid loading; that is, the fluid loading on the conical strip is considered the same acting on an equivalent cylindrical shell with the same width and radius $R_{0,i}$ as shown in Fig. 2. This approximation is only applicable to the calculation of the fluid loading acting on a shell segment. To solve for conical shell displacements, the equations of motion of a conical shell given by Eqs. (1)–(3) and hence the solutions given by Eqs. (4)–(9) are used.

A. Equations of motion of a fluid loaded cylindrical shell

To model the cylindrical shell segments and thus to calculate the fluid loading, the Flügge equations of motion were used since they are more accurate in the low frequency range with respect to other shell theories.[1] The equations of motion for a cylindrical shell are given in terms of axial u_c, circumferential v_c, and radial motion w_c. u_c, v_c, and w_c are, respectively, orthogonal components of the cylinder displacement in the x, θ, and z directions, as shown in Fig. 3. The equations of motion are:[1]

$$\frac{\partial^2 u_c}{\partial x^2} + \frac{(1-\upsilon)}{2R_0^2}(1+\beta^2)\frac{\partial^2 u_c}{\partial \theta^2} + \frac{(1+\upsilon)}{2R_0}\frac{\partial^2 v_c}{\partial x \partial \theta} + \frac{\upsilon}{R_0}\frac{\partial w_c}{\partial x} - \beta^2\left[R_0\frac{\partial^3 w_c}{\partial x^3} + \frac{(1-\upsilon)}{2R_0}\frac{\partial^3 w_c}{\partial x \partial \theta^2}\right] - \frac{\ddot{u}_c}{c_L^2} = 0, \tag{34}$$

$$\frac{(1+\upsilon)}{2R_0}\frac{\partial^2 u_c}{\partial x \partial \theta} + \frac{(1-\upsilon)}{2}\frac{\partial^2 v_c}{\partial x^2} + \frac{1}{R_0^2}\frac{\partial^2 v_c}{\partial \theta^2} + \frac{1}{R_0^2}\frac{\partial w_c}{\partial \theta} + \beta^2\left[\frac{3(1-\upsilon)}{2}\frac{\partial^2 v_c}{\partial x^2} - \frac{(3-\upsilon)}{2}\frac{\partial^3 w_c}{\partial x^2 \partial \theta}\right] - \frac{\ddot{v}_c}{c_L^2} = 0, \tag{35}$$

$$\beta^2\left(R_0^2\frac{\partial^4 w_c}{\partial x^4} + 2\frac{\partial^4 w_c}{\partial x^2 \partial \theta^2} + \frac{1}{R_0^2}\frac{\partial^4 w_c}{\partial \theta^4} - R_0\frac{\partial^3 u_c}{\partial x^3} + \frac{1-\upsilon}{2R_0}\frac{\partial^3 u_c}{\partial x \partial \theta^2} - \frac{3-\upsilon}{2}\frac{\partial^3 v_c}{\partial x^2 \partial \theta} + \frac{2}{R_0^2}\frac{\partial^2 w_c}{\partial \theta^2}\right) + \frac{\upsilon}{R_0}\frac{\partial u_c}{\partial x} + \frac{1}{R_0^2}\left[\frac{\partial v_c}{\partial \theta} + w_c(1+\beta^2)\right] + \frac{\ddot{w}_c}{c_L^2} - \frac{p_a(1-\upsilon^2)}{Eh} = 0, \tag{36}$$

where $\beta = h/(12R_0)^{1/2}$ is the thickness parameter and $c_L = [E/\rho(1-\upsilon^2)]^{1/2}$ is the longitudinal wave speed. General solutions to the equations of motion for a cylindrical shell given by Eqs. (34)–(36) can be assumed as[1]

$$u_c(x,\theta,t) = Ue^{jk_n x}\cos(n\theta)e^{-j\omega t}, \tag{37}$$

$$v_c(x,\theta,t) = Ve^{jk_n x}\sin(n\theta)e^{-j\omega t}, \tag{38}$$

$$w_c(x,\theta,t) = We^{jk_n x}\cos(n\theta)e^{-j\omega t}. \tag{39}$$

k_n is the axial wave number. The external pressure loading due to the fluid acting normally to the surface can be expressed in terms of a fluid loading parameter F_L. Expressions for the external pressure loading and the fluid loading parameter are given by[12]

$$p_a = \frac{\rho h c_L^2}{R_0^2}F_L w_c, \tag{40}$$

$$F_L = -\Omega^2\frac{R_0}{h}\frac{\rho_f}{\rho}\frac{H_n(k_{nr}R_0)}{H_n'(k_{nr}R_0)(k_{nr}R_0)}, \tag{41}$$

where $\Omega = \omega R_0/c_L$ is the nondimensional ring frequency and ρ_f is the density of the fluid. H_n is the Hankel function of order n and H_n' is its derivative with respect to the argument. The nondimensional radial wave number $k_n R_0$ is given by[12] $k_{nr}R_0 = [\Omega^2(c_L/c_f)^2 - (k_n R_0)^2]^{1/2}$. c_f is the speed of sound in the medium. The characteristic equation of the cylinder results from the vanishing of the determinant of the matrix $\mathbf{L}'$ to zero, where the elements of the matrix are

$$L_{11}' = \Omega^2 - (k_n R_0)^2 - \frac{(1-\upsilon)}{2}n^2(1+\beta^2), \tag{42}$$

$$L_{12}' = \frac{(1+\upsilon)}{2}jnk_n R_0, \tag{43}$$

$$L'_{13} = j\upsilon k_n R_0 + j\beta^2\left[(k_n R_0)^3 - \frac{(1-\upsilon)}{2}n^2(k_n R_0)\right], \tag{44}$$

$$L'_{21} = -L'_{12}, \tag{45}$$

$$L'_{22} = \Omega^2 - \frac{(1-\upsilon)}{2}(k_n R_0)^2(1+3\beta^2) - n^2, \tag{46}$$

$$L'_{23} = -n - \beta^2\frac{(3-\upsilon)}{2}n(k_n R_0)^2, \tag{47}$$

$$L'_{31} = L'_{13}, \tag{48}$$

$$L'_{32} = -L'_{23}, \tag{49}$$

$$L'_{33} = 1 + \beta^2\{[(k_n R_0)^2 + n^2]^2 + (1-2n^2)\} - (\Omega^2 + F_L). \tag{50}$$

B. Solution of the characteristic equation

As previously reported by Fuller and Fahy,[13] the dispersion equation of the fluid loaded cylinder leads to purely real, imaginary, or complex roots. Due to the Hankel functions, the characteristic equation is nonlinear and so a numerical method must be used to solve it. The location of the first two classes of structural wave numbers is obtained by observing a change in sign in the characteristic equation. With regards to the complex roots, the wave number can be written as $k_n = k_{re} + jk_{im}$. Similarly, the characteristic equation can be written as $C = C_{re} + jC_{im}$. Using the notation $X = k_{re}$, $Y = k_{im}$, $Z_1 = C_{re}$ and $Z_2 = C_{im}$, the complex solution of C can be found by solving the following system of equations.

$$\begin{cases} Z_1(X,Y) = 0 \\ Z_2(X,Y) = 0. \end{cases} \tag{51}$$

Substituting Eqs. (6) and (9) in Eq. (39), the external pressure in Eq. (3) can be expressed as

$$p_a = \frac{\rho h c_L^2}{R_0^2} F'_L \sum_{m=0}^{\infty} c_m x^m \cos(n\theta), \tag{52}$$

where the fluid loading applied to the conical shell has been modified by $F'_L = F_L/\cos\alpha$. This is to take into account that the surface of the conical shell is greater than that of the cylindrical shell by a factor of $1/\cos\alpha$. Hence a higher fluid loading effect is expected on the conical shell. The fluid loading used in Eq. (52) is calculated using the real solution of the characteristic equation that represents the travelling wave on the shell. For circumferential modes with $n > 1$, the traveling wave has a cut-on at higher frequencies. Below this cut-on frequency, there are no real solutions and the fluid loading is calculated using the complex root.

A qualitative analysis is given here in order to determine how many parts the conical shell should be divided into. For each value of Ω, the dispersion equation provides up to three real values of k_n. The three branches are associated with waves of the first, second, and third class.[14] At low frequencies, the wave numbers of interest for the fluid loading case

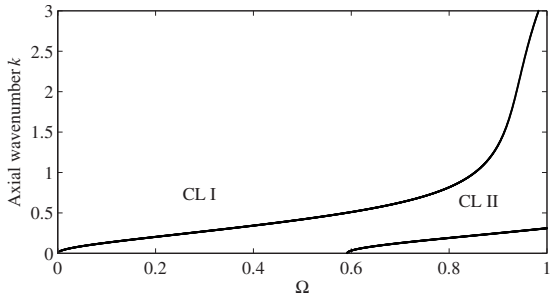


FIG. 4. Real solutions of the characteristic equation, $n=1$.

are the solutions of the class I (CL I) branch, as shown in Fig. 4 for the $n=1$ modes. In the low frequency range, the variation of the real wave number that is used to calculate the fluid loading parameter is almost linear as a function of the nondimensional frequency Ω and thus is also almost linear as a function of R_0. This is due to the fact that at low frequencies, the nature of this wave is predominantly extensional for the $n=0$ breathing modes, and neither flexural nor torsional for $n \geqslant 1$. Since the cylindrical approximation assumes that the wave number does not change along the length of the cone, a tolerance of $T\%$ is used for the variation of the wave number for each segment. Hence, the same percentage is fixed with respect to the mean radius $R_{0,i}$ for each segment of the cone. Thus, from the geometry of the cone, the length of each segment of the cone is $L_i \leqslant 2R_{0,i}T\%/(\sin\alpha)$. For example, let $T=5\%$ for a cone with $\alpha=18°$, $R_1=0.50$ m, and $R_2=3.25$ m. The corresponding division pattern shown in Fig. 5 is obtained. The length of the last cone can be adjusted to get the desired value of the final radius. It is evident that the length of the cone segments increases towards the end of larger radius.

C. Modification of the recurrence terms

The external pressure in Eq. (3) is now determined by Eq. (52). Substituting the general solutions given by Eqs. (4)–(9) into Eqs. (1)–(3), the modified recurrence terms $G'_{i,j}$ in the presence of fluid loading are found as

$$G'_{3,11} = G_{3,11} + R_0^4(j\omega Z)/(D_{11}R_0^4 m_{41}), \tag{53}$$

$$G'_{3,12} = G_{3,12} + 4R_0^3(j\omega Z)\sin\alpha/(D_{11}R_0^4 m_{41}), \tag{54}$$

$$G'_{3,13} = G_{3,13} + 6R_0^2(j\omega Z)\sin^2\alpha/(D_{11}R_0^4 m_{41}), \tag{55}$$

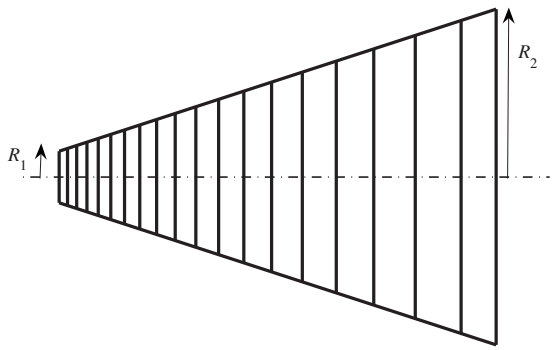


FIG. 5. Cone division pattern.

$$G'_{3,14} = G_{3,14} + 4R_0(j\omega Z)\sin^3\alpha/(D_{11}R_0^3 m_{41}), \tag{56}$$

$$G'_{3,15} = G_{3,15} + (j\omega Z)\sin^4\alpha/(D_{11}R_0^4 m_{41}). \tag{57}$$

Z is the acoustic impedance and is given by

$$Z = \frac{-j}{\omega}\frac{\rho h c_L^2}{R_0^2}F'_L. \tag{58}$$

Only the $G_{i,j}$ coefficients given by Eqs. (53)–(57) are modified due to the fluid loading. The remaining coefficients are the same as those for the *in vacuo* case given in Appendix A.

IV. CONICAL SHELL RESPONSE

A. Boundary conditions

The boundaries conditions at both ends of the conical shell and the continuity relationships between the conical segments can be written and assembled in matrix form $\mathbf{Ax}=\mathbf{0}$, where $\mathbf{x}=[\mathbf{x}_1^T\cdots\mathbf{x}_N^T]^T$ is the $8N\times 1$ unknown coefficient vector and

$$\mathbf{A} = \begin{bmatrix}
[\mathbf{B}_1] & & & & & \\
[\mathbf{F}_1(L_1/2)] & [\mathbf{F}_2(-L_2/2)] & & & & \\
[\mathbf{D}_1(L_1/2)] & [\mathbf{D}_2(-L_2/2)] & & & & \\
 & [\mathbf{F}_2(L_2/2)] & [\mathbf{F}_3(-L_3/2)] & & & \\
 & [\mathbf{D}_2(L_2/2)] & [\mathbf{D}_3(-L_3/2)] & & & \\
 & & \cdots & \cdots & & \\
 & & \cdots & \cdots & & \\
 & & & [\mathbf{F}_{N-1}(L_{N-1}/2)] & [\mathbf{F}_N(-L_N/2)] & \\
 & & & [\mathbf{D}_{N-1}(L_{N-1}/2)] & [\mathbf{D}_N(-L_N/2)] & \\
 & & & & [\mathbf{B}_N] &
\end{bmatrix}. \tag{59}$$

For the ith cone, the 4×8 matrix blocks of the displacements and forces used in Eq. (59) are

$$[\mathbf{D}_i(x)]_{4\times 8} = \begin{bmatrix} u_{i,1}\cdots u_{i,8} \\ v_{i,1}\cdots v_{i,8} \\ w_{i,1}\cdots w_{i,8} \\ \partial w_{i,1}/\partial x\cdots \partial w_{i,8}/\partial x \end{bmatrix}, \quad i=1,2,\ldots N, \tag{60}$$

$$[\mathbf{F}_i(x)]_{4\times 8} = \begin{bmatrix} N_{x,i,1} & \cdots & N_{x,i,8} \\ N_{x\theta,i,1} & \cdots & N_{x\theta,i,8} \\ V_{x,i,1} & \cdots & V_{x,i,8} \\ M_{x,i,1} & \cdots & M_{x,i,8} \end{bmatrix}, \quad i=1,2,\ldots,N. \tag{61}$$

The initial and final blocks $\mathbf{B}_1$ and $\mathbf{B}_N$ are expressed in terms of displacement and/or forces, depending on the boundary conditions at each end of the conical shell. Free, clamped, and shear-diaphragm boundary conditions are considered and are given by

Free end:

$$[\mathbf{B}_i]_{4\times 8} = \begin{bmatrix} N_{x,i,1} & \cdots & N_{x,i,8} \\ N_{x\theta,i,1} & \cdots & N_{x\theta,i,8} \\ V_{x,i,1} & \cdots & V_{x,i,8} \\ M_{x,i,1} & \cdots & M_{x,i,8} \end{bmatrix}, \quad i=1,N. \tag{62}$$

Clamped end:

$$[\mathbf{B}_i]_{4\times 8} = \begin{bmatrix} u_{i,1} & \cdots & u_{i,8} \\ v_{i,1} & \cdots & v_{i,8} \\ w_{i,1} & \cdots & w_{i,8} \\ \partial w_{i,1}/dx & \cdots & \partial w_{i,8}/dx \end{bmatrix}, \quad i=1,N. \tag{63}$$

Shear-Diaphragm (SD):

$$[\mathbf{B}_i]_{4\times 8} = \begin{bmatrix} N_{x,i,1} & \cdots & N_{x,i,8} \\ v_{i,1} & \cdots & v_{i,8} \\ w_{i,1} & \cdots & w_{i,8} \\ M_{x,i,1} & \cdots & M_{x,i,8} \end{bmatrix}, \quad i=1,N. \tag{64}$$

The natural frequencies for the cone are calculated from the vanishing of the determinant of matrix $\mathbf{A}$ for each circumferential wave number n.

B. Forced response

For the free/clamped and shear diaphragm (SD/SD) case the shell was driven at the small end with a point harmonic force of unity amplitude ($F_0=1$ N/m), as shown in Fig. 6. A point force located at (x_0,θ_0) can be described in terms of a Dirac delta function by

$$F(x,\theta,t) = F_0\delta(x-x_0)\delta(\theta-\theta_0)e^{-j\omega t}. \tag{65}$$

For the free end, the membrane force equation for N_x in the matrix block (62) becomes

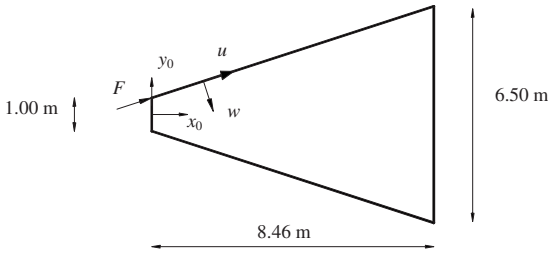


FIG. 6. Conical shell under point force excitation.

$$\frac{Eh}{1-\nu^2}\left[\frac{\partial u}{\partial x}+\nu\left(\frac{\sin\alpha}{R(x)}u+\frac{1}{R(x)}\frac{\partial v}{\partial\theta}-\frac{w\cos\alpha}{R(x)}\right)\right] = F_0\delta(x-x_0)\delta(\theta-\theta_0)e^{-j\omega t} \tag{66}$$

and can be rewritten as

$$\frac{Eh}{1-\nu^2}\left[\frac{\partial u(x)}{\partial x}+\nu\left(\frac{\sin\alpha}{R(x)}u(x)+\frac{nv(x)}{R(x)}-\frac{w(x)\cos\alpha}{R(x)}\right)\right]\cos(n\theta) = F_0\delta(x-x_0)\delta(\theta-\theta_0)e^{-j\omega t}. \tag{67}$$

Excluding the time harmonic dependency, multiplying the above equation by $\cos(n\theta)$ and then taking the integral from $-\pi$ to π, Eq. (67) becomes

$$\frac{Eh}{1-\nu^2}\left[\left.\frac{\partial u(x)}{\partial x}\right|_{x_0}+\nu\left(\left.\frac{\sin(\alpha)}{R(x)}u(x)\right|_{x_0}+\left.\frac{nv(x)}{R(x)}\right|_{x_0}-\left.\frac{w(x)\cos(\alpha)}{R(x)}\right|_{x_0}\right)\right] = \varepsilon F_0\cos(n\theta_0), \tag{68}$$

where $\varepsilon=1/2\pi R_1$ if $n=0$ and $\varepsilon=1/\pi R_1$ if $n\neq 0$. The boundary equations together with the equilibrium of the forces under point force excitation at the free end can be arranged in the matrix form $\mathbf{Ax}=\mathbf{F}$, where $\mathbf{F}$ is the 8×1 force vector with only one nonzero term, $\varepsilon F_0\cos(n\theta_0)$. Results are presented in terms of a frequency response function (FRF). The FRF was calculated solving $\mathbf{Ax}=\mathbf{F}$ for each frequency and circumferential mode number. The first ten circumferential mode numbers ($n=10$) were considered in the analysis. The cone's middle surface displacements are given by

$$u(x,\theta)=\varepsilon\sum_{n=0}^{10}\mathbf{u}\cdot\mathbf{x}\cos(n\theta), \tag{69}$$

$$v(x,\theta)=\varepsilon\sum_{n=0}^{10}\mathbf{v}\cdot\mathbf{x}\sin(n\theta), \tag{70}$$

$$w(x,\theta)=\varepsilon\sum_{n=0}^{10}\mathbf{w}\cdot\mathbf{x}\cos(n\theta). \tag{71}$$

The FRF was calculated by examining the component of the displacement along the axis x_0 or y_0, which are given by

$$u_{x_0}=u\cos\alpha+w\sin\alpha, \tag{72}$$

$$u_{y_0}=u\sin\alpha-w\cos\alpha. \tag{73}$$

V. RESULTS

The method presented here has been applied to a large conical shell typically used as a tail cone for submarine hulls. For the truncated conical shell shown in Fig. 1, the following data have been used: $L=8.90$ m, $R_1=0.50$ m, $R_2=3.25$ m, $h=0.014$ m, $\nu=0.3$, $E=2.1\times10^{11}$ N/m^2, $\rho=7800$ kg/m^3. For the fluid medium, $c_f=1500$ m/s, and $\rho_f=1000$ kg/m^3.

In a typical case, the principal low frequency excitation forces are at multiples of the blade passing frequency (bpf) and can be transmitted both through the propeller bearings and the external fluid. Typically, the frequency range up to 80 Hz will contain several multiples of bpf.

A. Natural frequencies

The natural frequencies for the cone can be calculated from the vanishing of the determinant of matrix $\mathbf{A}$ for each circumferential wave number n. The results are presented in the frequency range up to 80 Hz for three different boundary conditions and are given in terms of the dimensionless frequency parameter $\Omega_c=\sqrt{\rho h/A_{11}}\omega R_2$.

1. Free/Clamped boundary conditions

The boundary conditions used are free for the small end (of radius R_1) and clamped for the large one (of radius R_2). Table I presents the first five undamped frequency parameters.

2. SD/SD boundary conditions

Shear diaphragm (SD) boundary conditions are used for both ends of the conical shell. Table II presents the first five undamped frequency parameters.

3. Clamped/clamped boundary conditions

Clamped boundary conditions are used for both ends of the conical shell. Table III presents the first five undamped frequency parameters. In Fig. 7, the lowest natural frequency parameter Ω_c is plotted as a function of the circumferential mode number. It can be seen that the higher order frequencies are almost independent of the boundary conditions. For the free-clamped and clamped-clamped cases, there is no natural frequency associated with $n=0$ in the frequency range analyzed.

Figure 8 shows the lowest natural frequency parameters for the free/clamped and SD/SD boundary conditions. Results are presented for both the *in vacuo* case and with fluid loading. In general, under fluid loading the frequencies are shifted by more than half of their *in vacuo* values. For both *in vacuo* case and with fluid loading, the lowest frequency parameter for the SD/SD case is given by $n=0$.

B. Computational model

A computational boundary element model (BEM) was developed to compare results with those obtained from the analytical model presented in this paper. The structural un-

TABLE I. Frequency parameters for the free-clamped cone.

Circumferential mode number n	Frequency parameter Ω_c				
	1	2	3	4	5
0					
1	0.131 774	0.231 636			
2	0.058 566	0.127 269	0.221 875		
3	0.053 686	0.111 125	0.175 323	0.245 527	
4	0.052 184	0.103 992	0.167 064	0.234 264	0.299 588
5	0.055 563	0.103 617	0.160 306	0.221 125	0.282 694
6	0.06 232	0.108 497	0.161 432	0.217 746	0.275 936
7	0.072 081	0.117 508	0.16 819	0.222 251	0.277 438
8	0.084 095	0.129 521	0.179 828	0.232 387	0.286 073
9	0.098 361	0.144 914	0.195 220	0.24 778	
10	0.11 488	0.162 934	0.214 367	0.266 926	

coupled modes were initially calculated using a finite element model developed in Patran-Nastran. The coupled problem was then solved in SYSNOISE using modal superposition with a coupled direct approach.[15] In order to obtain reliable results, structural uncoupled modes up to three times the maximum frequency of interest were used. A truncated conical shell with free/clamped and clamped/clamped boundary conditions were modeled using BEM. Structural damping was introduced by considering a complex Young modulus $E'=E(1-j\eta_s)$, where $\eta_s=0.02$ is the structural loss factor. Radiation damping of the fluid was introduced by modifying the speed of sound in the medium in a similar way, that is, $c_f'=c_f(1-j\eta_f)$, where $\eta_f=0.001$ is the damping loss factor of the fluid. The results from SYSNOISE are presented in form of a frequency response function (FRF). Figures 9 and 10 present the FRFs obtained computationally using SYSNOISE and for the analytical model presented here, for a free/clamped shell driven at different axial locations. Comparison of these results shows very good agreement between the analytical and computational models over most of the frequency

TABLE II. Frequency parameters for the cone with SD boundary condition at both ends.

Circumferential mode number n	Frequency parameter Ω_c				
	1	2	3	4	5
0	0.021 775	0.263 547	0.296 209		
1	0.123 514	0.242 899			
2	0.0 443	0.180 954	0.270 305		
3	0.031 536	0.107 747	0.19 522	0.280 066	
4	0.034 914	0.094 982	0.162 558	0.233 513	
5	0.041 672	0.094 982	0.155 425	0.218 872	0.281 192
6	0.049 556	0.100 614	0.156 176	0.214 742	0.273 684
7	0.060 068	0.109 999	0.163 309	0.218 872	0.274 435
8	0.072 457	0.122 388	0.174 572	0.228 633	0.28 269
9	0.087 474	0.138 156	0.189 964	0.243 274	0.296 960
10	0.104 368	0.156 176	0.208 736	0.262 421	

TABLE III. Frequency parameters for the cone with clamped boundary condition at both ends.

Circumferential mode number n	Frequency parameter Ω_c				
	1	2	3	4	5
0					
1	0.158 804	0.278 189			
2	0.084 846	0.175 323	0.277 438		
3	0.056 689	0.122 764	0.20 423	0.287 574	
4	0.052 184	0.104 743	0.168 565	0.238 394	
5	0.055 563	0.103 617	0.160 306	0.221 125	0.283 069
6	0.055 563	0.103 617	0.160 432	0.217 746	0.275 936
7	0.072 081	0.117 508	0.16 819	0.222 251	0.277 438
8	0.084 095	0.129 521	0.179 828	0.232 387	0.286 448
9	0.098 361	0.144 914	0.19 522	0.247 404	
10	0.11 488	0.162 934	0.214 367	0.266 926	

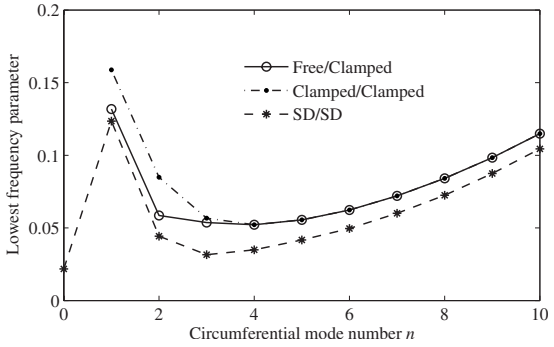


FIG. 7. Lowest frequency parameters for different boundary conditions.

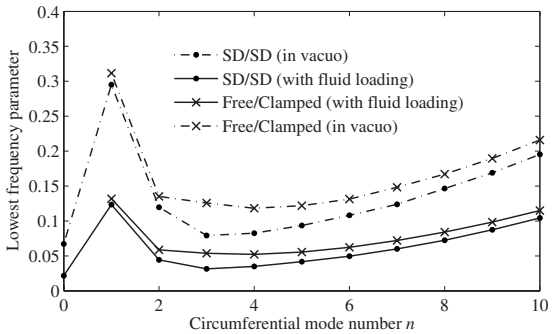


FIG. 8. Lowest frequency parameters for free/clamped and SD/SD boundary conditions.

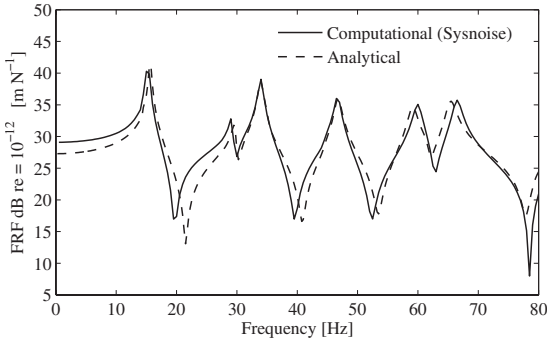


FIG. 9. Frequency response function of u_{x_0} at the driving point (x=0 m), θ=0 (free/clamped boundary conditions).

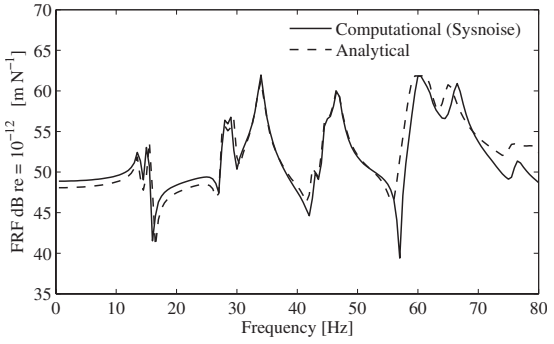


FIG. 10. Frequency response function of u_{x_0} at x=7.33 m, θ=0 (free/clamped boundary conditions).

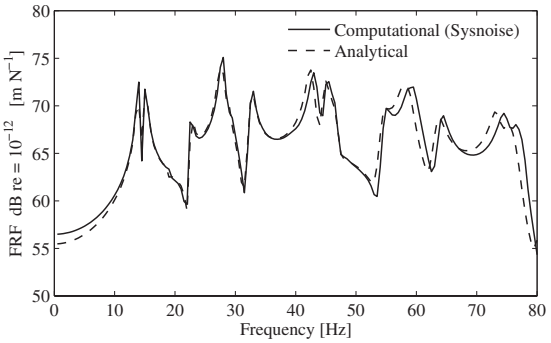


FIG. 11. Frequency response function of u_{x_0} at x=1.69 m, θ=0 (clamped/clamped boundary conditions).

range, although differences in the natural frequencies are observed at higher frequencies. Differences in the curves are attributed to the differing underlying assumptions used for the two techniques, mainly associated with the structural damping and fluid loading.

Figures 11 and 12 present the FRFs obtained analytical and computationally for clamped/clamped and SD/SD boundary conditions, respectively. In the case of the cone with clamped ends, the shell was excited with a force normal to the surface located at x_0=7.33 m. While good agreement is again observed between the analytical and computational results, greater differences can be observed for a shell with shear diaphragm boundary conditions and are attributed to the difficulties associated in modeling these boundary conditions in SYSNOISE.

VI. CONCLUSIONS

A model to predict the dynamic behavior of a truncated conical shell immersed in a fluid has been presented. A power series solution was used to solve for the shell displacements. Fluid loading was included in the analytical model by considering the conical shell to be locally cylindrical. Results have been presented in terms of natural frequencies for different boundary conditions, for representative submarine tail cone parameters. A boundary element model was developed to compare computational results with those obtained from the analytical model.

The analytical model is limited to the low frequency range by several approximations. First, the fluid loading was computed as if the finite shell response was periodically ex-

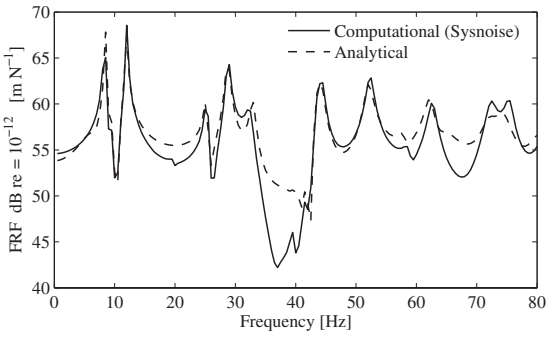


FIG. 12. Frequency response function of u_{y_0} at x=7.33 m, θ=0 (SD/SD boundary conditions).

tended to infinity along a cylindrical baffle. Thus, the real fluid loading on a short shell can vary significantly from the infinite model at higher frequencies. Other limitations that could lead to inaccurate results at higher frequencies arise from the assumption that fluid loading is related to the traveling wave only, instead of all the wave types that occur in the shell. In the range of frequencies considered, the fluid loading is mainly an added mass effect and this approximation is acceptable. At higher frequencies, the damping and mass effect of the fluid loading on the decaying waves could lead to inaccurate results.

Different boundary conditions can be applied to the model presented. The shell can be attached to plates or other shells, such as cones or cylinders, by means of four boundary conditions at each end of the truncated cone.

The model presented in this work is particularly suited for underwater applications where large conical shells are used. A tail cone is often used in a submarine to support the propeller bearing and numerical results have been presented for representative parameters. The principal low frequency excitation forces are at multiples of the blade passing frequency and can be transmitted both through the propeller bearings and the external fluid.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr Roger Kinns for his helpful comments on this paper.

APPENDIX A: RECURRENCE TERMS

For the *in vacuo* case, the recurrence terms $G_{i,j}$ are given by[7]

$$G_{1,1} = -(2m+1)\sin\alpha/(m+2)R_0, \tag{A1}$$

$$G_{1,2} = -m^2\sin^2\alpha/m_{21}R_0^2 + (A_{11}\sin^2\alpha + A_{66}n^2)/(A_{11}m_{21}R_0^2) - \rho h\omega^2/(A_{11}m_{21}), \tag{A2}$$

$$G_{1,3} = -2\rho h\omega^2\sin\alpha/A_{11}m_{21}R_0, \tag{A3}$$

$$G_{1,4} = -\rho h\omega^2\sin^2\alpha/A_{11}m_{21}R_0^2, \tag{A4}$$

$$G_{1,5} = -(A_{12}+A_{66})n/A_{11}(m+2)R_0, \tag{A5}$$

$$G_{1,6} = -[(A_{12}+A_{66})m - (A_{11}+A_{66})]n\sin\alpha/A_{11}m_{21}R_0^2, \tag{A6}$$

$$G_{1,7} = A_{12}\cos\alpha/A_{11}(m+2)R_0, \tag{A7}$$

$$G_{1,8} = (A_{12}m - A_{11})\sin\alpha\cos\alpha/A_{11}m_{21}R_0^2, \tag{A8}$$

$$G_{2,1} = (A_{12}+A_{66})n/A_{66}(m+2)R_0, \tag{A9}$$

$$G_{2,2} = [(A_{12}+A_{66})m + (A_{11}+A_{66})]n\sin\alpha/A_{66}m_{21}R_0^2, \tag{A10}$$

$$G_{2,3} = G_{1,1}, \tag{A11}$$

$$G_{2,4} = -m^2\sin^2\alpha/m_{21}R_0^2 + (A_{66}\sin^2\alpha + A_{11}n^2)/A_{66}m_{21}R_0^2 - \rho h\omega^2/A_{66}m_{21}, \tag{A12}$$

$$G_{2,5} = -2\rho h\omega^2\sin\alpha/A_{66}m_{21}R_0, \tag{A13}$$

$$G_{2,6} = -\rho h\omega^2\sin^2\alpha/A_{66}m_{21}R_0^2, \tag{A14}$$

$$G_{2,7} = -A_{11}n\cos\alpha/A_{66}m_{21}R_0^2, \tag{A15}$$

$$G_{3,1} = A_{12}\cos\alpha/D_{11}(m+4)(m+3)(m+2)R_0, \tag{A16}$$

$$G_{3,2} = (3A_{12}m + A_{11})\sin\alpha\cos\alpha/D_{11}m_{41}R_0^2, \tag{A17}$$

$$G_{3,3} = [3A_{12}(m-1) + 2A_{11}]\sin^2\alpha\cos\alpha/D_{11}m_{41}R_0^3, \tag{A18}$$

$$G_{3,4} = [A_{12}(m-2) + A_{11}]\sin^3\alpha\cos\alpha/D_{11}m_{41}R_0^4, \tag{A19}$$

$$G_{3,5} = A_{11}n\cos\alpha/D_{11}m_{41}R_0^2, \tag{A20}$$

$$G_{3,6} = A_{11}\sin(2\alpha)n/D_{11}m_{41}R_0^3, \tag{A21}$$

$$G_{3,7} = A_{11}n\sin^2\alpha\cos\alpha/D_{11}m_{41}R_0^4, \tag{A22}$$

$$G_{3,8} = -2(2m+1)\sin\alpha/(m+4)R_0, \tag{A23}$$

$$G_{3,9} = \{[2(D_{12}+2D_{66})n^2 + D_{11}\sin^2\alpha]/D_{11}R_0^2 - 6m^2\sin^2\alpha/R_0^2\}/(m+4)(m+3), \tag{A24}$$

$$G_{3,10} = \{[2(D_{12}+2D_{66})n^2 + D_{11}\sin^2\alpha] \times (2m-1)\sin\alpha/D_{11}Ro^3 , - 2m(m-1)(2m-1)\sin^3\alpha/R_0^3\}/(m+4)(m+3)(m+2), \tag{A25}$$

$$G_{3,11} = [(2D_{11}n^2 + D_{11}\sin^2\alpha)(\sin^2\alpha)m(m-2) - A_{11}R_0^2\cos^2\alpha - D_{11}m(m-1)^2(m-2)\sin^4\alpha - D_{11}n^4 + 4D_{11}n^2\sin^2\alpha + \rho h\omega^2R_0^4]/(D_{11}R_0^4m_{41}), \tag{A26}$$

$$G_{3,12} = (4\rho h\omega^2R_0^3\sin\alpha - 2A_{11}R_0\sin\alpha\cos^2\alpha)/(D_{11}R_0^4m_{41}), \tag{A27}$$

$$G_{3,13} = (6\rho h\omega^2R_0^2\sin^2\alpha - A_{11}R_0\sin^2\alpha\cos^2\alpha)/(D_{11}R_0^4m_{41}), \tag{A28}$$

$$G_{3,14} = 4\rho h\omega^2\sin^3\alpha/(D_{11}R_0^3m_{41}), \tag{A29}$$

$$G_{3,15} = \rho h\omega^2\sin^4\alpha/(D_{11}R_0^4m_{41}), \tag{A30}$$

where R_0 is the mean radius of the shell, $D_{11}=Eh^3/12(1-\upsilon^2)$, $D_{12}=\upsilon D_{11}$, $D_{66}=Eh^3/24(1+\upsilon)$, $A_{11}=Eh/(1-\upsilon^2)$, $A_{12}=\upsilon A_{11}$, $A_{66}=Eh/2(1+\upsilon)$, $m_{21}=(m+2)(m+1)$ and $m_{41}=(m+4)(m+3)(m+2)(m+1)$.

[1]A. W. Leissa, *Vibration of Shells* (American Institute of Physics, New York, 1993).
[2]H. Saunders, E. J. Wisniewski, and P. R. Paslay, "Vibrations of conical shells," J. Acoust. Soc. Am. **32**, 765–772 (1960).
[3]H. Garnet and J. Kempner, "Axisymmetric free vibration of conical shells," J. Appl. Mech. **31**, 458–466 (1964).
[4]R. A. Newton, "Free vibrations of rocket nozzles," AIAA J. **4**, 1303–1305 (1966).
[5]C. C. Siu and C. W. Bert, "Free vibrational analysis of sandwich conical shells with free edges," J. Acoust. Soc. Am. **47**, 943–945 (1970).
[6]T. Irie, G. Yamada, and Y. Kaneko, "Free vibration of a conical shell with variable thickness," J. Sound Vib. **82**, 83–94 (1982).
[7]L. Tong, "Free vibration of orthotropic conical shells," Int. J. Eng. Sci. **31**, 719–733 (1993).
[8]Y. P. Guo, "Normal mode propagation on conical shells," J. Acoust. Soc. Am. **96**, 256–264 (1994).
[9]Y. P. Guo, "Fluid-loading effects on waves on conical shells," J. Acoust. Soc. Am. **97**, 1061–1066 (1995).
[10]A. N. Norris and D. A. Rebinsky, "Acoustic coupling to membrane waves on elastic shells," J. Acoust. Soc. Am. **95**, 1809–1829 (1994).
[11]M. C. Junger and D. Feit, *Sound, Structures, and their Interaction* (MIT Press, Cambridge, MA, 1986).
[12]C. R. Fuller, "Radiation of sound from an infinite cylindrical elastic shell excited by an internal monopole source," J. Sound Vib. **109**, 259–275 (1986).
[13]C. R. Fuller and F. J. Fahy, "Characteristics of wave propagation and energy distribution in cylindrical elastic shells filled with fluid," J. Sound Vib. **81**, 501–518 (1982).
[14]P. W. Smith, Jr., "Phase velocities and displacement characteristics of free waves in a thin cylindrical shell," J. Acoust. Soc. Am. **27**, 1065–1072 (1955).
[15]S. Merz, S. Oberst, P. G. Dylejko, N. J. Kessissoglou, Y. K. Tso, and S. Marburg, "Development of coupled FE/BE models to investigate the structural and acoustic responses of a submerged vessel," J. Comput. Acoust. **15**, 23–47 (2007).

Self-learning active noise control

J. Yuan[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong

(Received 12 September 2007; revised 8 May 2008; accepted 16 July 2008)

An important step for active noise control (ANC) systems to be practical is to develop model independent ANC (MIANC) systems that tolerate parameter variations in sound fields. Reliabilities and stabilities of many MIANC systems depend on results of online system identifications. Parameter errors due to system identifications may threaten closed-loop stabilities of MIANC systems. A self-learning active noise control (SLANC) system is proposed in this study to stabilize and optimize an ANC system in case identified parameters are unreliable. The proposed system uses an objective function to check closed-loop stability. If partial or full value of the objective function exceeds a conservatively preset threshold, a stability threat is detected and the SLANC system will stabilize and optimize the controller without using parameters of sound fields. If the reference signal is available, the SLANC system can be combined with a feedforward controller to generate both destructive interference and active damping in sound fields. The self-learning method is simple and stable for many feedback ANC systems to deal with a worst case discussed in this study.
 [DOI: 10.1121/1.2968700]

PACS number(s): 43.50.Ki [BSF] Pages: 2078–2084

I. INTRODUCTION

It has been demonstrated by many researchers that active noise control (ANC) is an effective way of suppressing low-frequency noise. The principle of ANC is either destructive interference by feedforward control, or active damping by feedback control.[1,2] It is possible to combine feedforward/feedback control to generate both destructive interference and active damping in sound fields. Feedforward control is applicable to a sound field if (a) the reference signal is available or recoverable, and (b) primary source and secondary actuator(s) are in upstream locations with respect to an intended quiet zone. In some applications, it is possible to separate reflected waves from incident waves. Feedforward control may be applied to cancel reflected waves for sound absorption. If the reference signal is not available, feedback control is preferred to generate active damping in sound fields. In three-dimensional sound fields or vibration systems, active damping may be more suitable since it does not require the reference signal and separation of incident/reflected waves.

Accurate models of sound fields are very important in generating destructive (instead of constructive) interference or positive (instead of negative) damping. Model independent ANC (MIANC) systems depend on results of online system identifications or invariant properties of sound fields to ensure stabilities. Examples of feedback MIANC systems are direct rate feedback controllers or active resonators, which are theoretically stable in sound fields if dynamics of electronic circuits are negligible.[3–6] In reality, dynamics of electronic circuits are not necessarily negligible and may cause stability problems to direct rate feedback controllers or active resonators. Many direct rate feedback controllers and active resonators require collocated feedback signals. Suppressing collocated feedback signals does not necessarily mean suppressing signals measured away from speakers.

In this study, a self-learning ANC (SLANC) system is proposed as an alternative way to deal with existing problems in feedback MIANC systems. The SLANC system does not require collocated feedback and tolerates dynamic uncertainties (including dynamics of electronic circuits) to ensure closed-loop stability. It is also able to minimize an objective function by tuning controller parameters without using parameters of sound fields. If the reference signal is available, the SLANC system may be integrated with a feedforward controller to generate both destructive interference and active damping in sound fields. Theoretical analysis and experimental results are presented to demonstrate stability and effectiveness of the SLANC system.

II. MODEL OF RESONANT SOUND FIELDS

Since active damping is most suitable for resonant sound fields, the modal theory is adopted to model resonant sound fields. One may consider an ANC system with a primary source at point x_p, a secondary actuator at point x_a, a feedback sensor at point x_f, and an error sensor at point x_e. Let $\phi_k(x)$ denote the kth eigenfunction of a sound field and $q_k(t)$ denote the corresponding temporal coordinate. The pressure signal, measured by the error sensor, may be expressed as[1,2]

$$p(x_e,t)=\sum_{k=1}^{m}\phi_k(x_e)q_k(t). \tag{1}$$

Equation (1) is theoretically an infinite summation. Practically, it is truncated to the first m modes. The signal model is applicable to pressure signals measured at other points such as the feedback point x_f. The only modification is to replace x_e with spatial coordinate of the measurement point such as

a)Electronic mail: mmjyuan@polyu.edu.hk

x_f. The dynamic equation of temporal coordinate $q_k(t)$ is given by[1,2]

$$\ddot{q}_k + 2\xi_k\omega_k\dot{q}_k + \omega_k^2 q_k = \phi_k(x_p)n + \phi_k(x_a)u, \tag{2}$$

where n and u are input signals from the primary source and secondary actuator.

Using Eqs. (1) and (2), one may derive a state space model,[7]

$$\dot{\mathbf{v}} = \mathbf{A}_0\mathbf{v} + \mathbf{B}_{0a}u + \mathbf{B}_{0p}n, \tag{3}$$

$$p_e = \mathbf{C}_{0e}\mathbf{v}, \quad p_f = \mathbf{C}_{0f}\mathbf{v}, \tag{4}$$

where $\mathbf{v}^T=[q_1\ q_2\ \cdots\ q_m\ \dot{q}_1\ \dot{q}_2\ \cdots\ \dot{q}_m]$ is the state vector; p_e and p_f are error and feedback signals. Matrices $\mathbf{A}_0$, $\mathbf{B}_{0a}$, $\mathbf{B}_{0p}$, $\mathbf{C}_{0e}$, and $\mathbf{C}_{0f}$ are given by

$$\mathbf{A}_0 = \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ -\omega_1^2 & & -2\xi_1\omega_1 & & \\ & -\omega_2^2 & & -2\xi_2\omega_2 & \\ & & \ddots & & \ddots \\ & & -\omega_m^2 & & -2\xi_m\omega_m \end{bmatrix}, \tag{5}$$

$$\mathbf{B}_{0a}^T = [0 \quad 0 \quad \cdots \quad 0 \quad \phi_1(x_a) \quad \phi_2(x_a) \quad \cdots \quad \phi_m(x_a)], \tag{6}$$

$$\mathbf{B}_{0p}^T = [0 \quad 0 \quad \cdots \quad 0 \quad \phi_1(x_p) \quad \phi_2(x_p) \quad \cdots \quad \phi_m(x_p)], \tag{7}$$

$$\mathbf{C}_{0e} = [\phi_1(x_e) \quad \phi_2(x_e) \quad \cdots \quad \phi_m(x_e) \quad 0 \quad 0 \quad \cdots \quad 0], \tag{8}$$

$$\mathbf{C}_{0f} = [\phi_1(x_f) \quad \phi_2(x_f) \quad \cdots \quad \phi_m(x_f) \quad 0 \quad 0 \quad \cdots \quad 0]. \tag{9}$$

If an ANC system has more secondary actuators, error, or feedback sensors, then Eqs. (6), (8), and (9) must be modified by matching dimensions of $\mathbf{B}_{0a}$, $\mathbf{C}_{0e}$, and $\mathbf{C}_{0f}$ to respective numbers of actuators, error, or feedback sensors. If n_e error sensors are also feedback sensors, then $\mathbf{C}_{0e}$ shares n_e rows with $\mathbf{C}_{0f}$. If error and feedback sensors are the same, then $\mathbf{C}_{0e}=\mathbf{C}_{0f}$.

Equations (3)–(9) depend on $\phi_k(x)$ and ω_k that are sensitive to boundary conditions or acoustical parameters. Take an L meter one-dimensional duct for example, the kth eigenfunction changes from $\phi_k(x)=\cos(k\pi x/L)$ to $\phi_k(x)=\cos((2k-1)\pi x/2L)$ if boundary conditions of the duct change from completely closed–closed to completely closed–open.[8] Resonant frequency ω_k also changes accordingly. If one end of the duct is partially closed or partially open, then model parameters become more complicated without analytical expressions. Destructive interference or positive damping, generated by a fixed controller for a completely closed–open duct, may become constructive interference or negative damping if boundary conditions of the duct change significantly.[9] It could destabilize an otherwise stable and fixed ANC system.

A practical ANC system must tolerate parameter variations in sound fields. MIANC systems have such an advantage. One can use Eqs. (3)–(9) to show that a direct rate feedback controller is stable in any sound field if the feedback sensor collocates with the actuator ($x_f=x_a$).[3,4] The same condition is required so that the secondary path model is minimum phase for stable operation of active resonators.[5,6] Due to practical restrictions, sensors are placed sufficiently away from actuators and secondary paths are nonminimum phase in many ANC applications. Besides, dynamics of electronic circuits are ignored in the derivation of Eqs. (3)–(9). If dynamics of electronic circuits are taken into account, the model is still represented by Eqs. (3) and (4) but dimensions of $\mathbf{v}$, $\mathbf{A}_0$, $\mathbf{B}_{0a}$, $\mathbf{B}_{0p}$, $\mathbf{C}_{0e}$, and $\mathbf{C}_{0f}$ will increase and the parameters become more complicated. Closed-loop stabilities of direct rate feedback control or active resonators are not guaranteed due to dynamics of electronic circuits, even when feedback sensors collocate with actuators.

III. MODEL OF A CLOSED-LOOP SYSTEM

The SLANC system is able to deal with the abovementioned problems, as a special kind of MIANC system. Although Eqs. (3) and (4) are used in design and analysis of a SLANC system, dynamics of electronic circuits are included and closed-loop stability does not rely on parameters in $\mathbf{A}_0$, $\mathbf{B}_{0a}$, $\mathbf{B}_{0p}$, $\mathbf{C}_{0e}$, and $\mathbf{C}_{0f}$. Feedback sensors are not required to collocate with actuators. The only required information is that all eigenvalues of $\mathbf{A}_0$ are in the negative half of the complex plane. This is an invariant property since all sound fields are open-loop stable with bounded outputs when excited by bounded inputs.

Since the SLANC system may be integrated with a discrete-time feedforward controller to generate both destructive interference and active damping in sound fields, one may adopt a discrete-time model of sound fields[7] for analysis. The model is given by

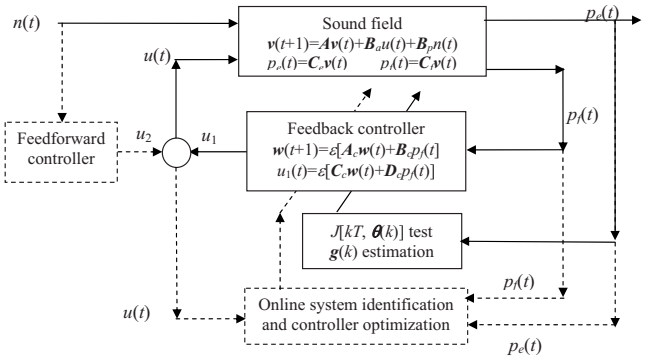


FIG. 1. Block diagram of the closed-loop SLANC system.

$$\mathbf{v}(t+1) = \mathbf{A}\mathbf{v}(t) + \mathbf{B}_a u(t) + \mathbf{B}_p n(t), \tag{10}$$

$$p_e(t) = \mathbf{C}_e \mathbf{v}(t), \quad p_f(t) = \mathbf{C}_f \mathbf{v}(t). \tag{11}$$

Equations (10) and (11) include dynamics of sensors, actuators, and electronic circuits. It is customary to assume $\delta t = 1$, without any negative effects to the general case of $\delta t \neq 1$.

Unlike $\mathbf{A}_0$, $\mathbf{B}_{0a}$, $\mathbf{B}_{0p}$, $\mathbf{C}_{0e}$, and $\mathbf{C}_{0f}$ in the continuous-time model, $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ contain complicated and unknown parameters, some of which may not have analytical expressions. Stability of a SLANC system only depends on eigenvalues of $\mathbf{A}$ that are all inside the unit circle centered at the origin of a complex plane. This is an invariant property of discrete-time models of sound fields that are open-loop stable with bounded outputs when excited by bounded inputs. Mathematically, it implies the existence of positive definite matrices $\mathbf{P} = \mathbf{P}^T > 0$ and $\mathbf{Q} = \mathbf{Q}^T > 0$,[7,10,11,13] such that

$$\mathbf{P} - \mathbf{A}^T \mathbf{P} \mathbf{A} = \mathbf{Q} > 0. \tag{12}$$

The closed-loop SLANC system is described by a block diagram in Fig. 1. The actuation signal consists of two parts $u(t) = u_1(t) + u_2(t)$, where $u_1(t)$ is the feedback part and $u_2(t)$ an optional feedforward part. The signals are synthesized by two blocks in Fig. 1. The dashed block, labeled "feedforward controller," generates the optional $u_2(t)$ only if $n(t)$ is available as the reference signal. The main focus of this study is $u_1(t)$, synthesized by the solid block of "feedback controller." The corresponding control law is

$$\mathbf{w}(t+1) = \varepsilon[\mathbf{A}_c \mathbf{w}(t) + \mathbf{B}_c p_f(t)], \tag{13}$$

$$u_1(t) = \varepsilon[\mathbf{C}_c \mathbf{w}(t) + \mathbf{D}_c p_f(t)], \tag{14}$$

where $\varepsilon > 0$ is a tunable parameter; elements of $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$ may be initialized with random numbers and tuned by two possible algorithms. One is the combination of available online system identification and controller optimization algorithms, represented by the dashed tuning block in Fig. 1; the other is a self-learning algorithm, represented by the solid tuning block in Fig. 1. The self-learning objective is to minimize

$$J(kT, \boldsymbol{\theta}) = \frac{1}{T} \sum_{i=kT}^{(k+1)T} p_e^2(i), \tag{15}$$

where T is the number of signal samples in a sufficiently long averaging period; $\boldsymbol{\theta}$ is a vector with one-to-one mapping to elements of $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$. The proposed system is based on an assumption that

$$J(kT, \boldsymbol{\theta}) \approx J(T, \boldsymbol{\theta}) \approx J(\boldsymbol{\theta}) = E\{p_e^2 | \boldsymbol{\theta}\} \tag{16}$$

for all integer k if the averaging period T is sufficiently long. This is applicable to sound fields with stationary and ergodic noise sources.

One may derive the model of closed-loop system by combining models of the sound field, Eqs. (10) and (11), and the controller, Eqs. (13) and (14). The result is given by [7,10,11,13]

$$\begin{bmatrix} \mathbf{v}(t+1) \\ \mathbf{w}(t+1) \end{bmatrix} = \begin{bmatrix} \mathbf{A} + \varepsilon \mathbf{B}_a \mathbf{D}_c \mathbf{C}_f & \varepsilon \mathbf{B}_a \mathbf{C}_c \\ \varepsilon \mathbf{B}_c \mathbf{C}_f & \varepsilon \mathbf{A}_c \end{bmatrix} \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_p \\ \end{bmatrix} n(t) + \begin{bmatrix} \mathbf{B}_a \\ \end{bmatrix} u_2(t), \tag{17}$$

$$p_e(t) = [\mathbf{C}_e] \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix}. \tag{18}$$

The feedforward signal is switched off $u_2(t) = 0$ if reference signal $n(t)$ is not available. Minimizing $J(kT) \approx E\{p_e^2 | \boldsymbol{\theta}\}$ can be shown mathematically equivalent to minimizing the H_2 norm from $n(t)$ to $p_e(t)$ by tuning $\boldsymbol{\theta}$.[10] It is possible to apply available system identification algorithms to estimate $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$. If online estimates are accurate, the SLANC system may set $\varepsilon = 1$ and apply H_2/H_∞ control algorithms[10,11,13] to optimize $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$. This is a possible version of indirect adaptive control and the dashed tuning block in Fig. 1.

When the reference signal is not available, estimation accuracy is a nontrivial issue.[7,12] The primary source $n(t)$ causes identification errors if it is not available to the ANC system. Closed-loop stabilities of indirect adaptive controllers are not guaranteed when online estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ are not reliable. The SLANC system is able to deal with such a case. It is represented by the solid tuning block in Fig. 1. Switching between the two tuning blocks depends on online values of $J(kT)$.

IV. SLANC MECHANISM

Although the full value of $J(kT)$ is computed periodically with Eq. (15), partial value of $J(kT)$ is available in every step of summation. If closed-loop stability is threatened, the magnitude of $p_e(t)$ grows rapidly and so does partial or full value of $J(kT)$. Whenever partial or full value of $J(kT)$ exceeds a conservatively preset threshold value, the SLANC system will (a) stabilize the closed loop and (b) optimize $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$ without using estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$. This study focuses on the worst case when $n(t)$ is unavailable and so severe that results of online identification are not reliable to ensure closed-loop stability.

A. Stabilizing the closed loop

The first task of the SLANC system is to stabilize the closed-loop system. Initially, the dashed tuning block is active with $\varepsilon = 1$, presuming that estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ are accurate. When partial or full value of $J(kT)$

grows out of bound, the system realizes that estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ are not reliable. It switches from the dashed tuning block to the solid counterpart. Tuning ε becomes the only option to ensure closed-loop stability. The effects may be analyzed with help of positive definite function,[7,10,11]

$$L(t) = \mathbf{v}^T(t)\mathbf{P}\mathbf{v}(t) + \mathbf{w}^T(t)\mathbf{w}(t), \tag{19}$$

where $\mathbf{P}$ is the positive definite matrix in Eq. (12).

If a stability threat is detected, $u_2(t)=0$ is the first action of the SLANC system. Stability of a discrete-time linear system like Eq. (17) may be analyzed using $L(t+1)-L(t)$ by ignoring bounded inputs $[n(t)=0$ in this particular case$]$.[7,10,11] As a result, Eq. (17) becomes

$$\begin{bmatrix} \mathbf{v}(t+1) \\ \mathbf{w}(t+1) \end{bmatrix} = \begin{bmatrix} \mathbf{A}+\varepsilon\mathbf{B}_a\mathbf{D}_c\mathbf{C}_f & \varepsilon\mathbf{B}_a\mathbf{C}_c \\ \varepsilon\mathbf{B}_c\mathbf{C}_f & \varepsilon\mathbf{A}_c \end{bmatrix} \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix}.$$

One may use the above-presented equation to obtain

$$\begin{aligned} L(t+1)-L(t) &= \begin{bmatrix} \mathbf{v}(t+1) \\ \mathbf{w}(t+1) \end{bmatrix}^T \begin{bmatrix} \mathbf{P} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}(t+1) \\ \mathbf{w}(t+1) \end{bmatrix} - \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix}^T \begin{bmatrix} \mathbf{P} & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix}^T \begin{bmatrix} \mathbf{A}^T\mathbf{P}\mathbf{A}-\mathbf{P}+\varepsilon\mathbf{R}_1 & \varepsilon\mathbf{R}_2 \\ \varepsilon\mathbf{R}_2^T & \varepsilon^2(\mathbf{A}_c^T\mathbf{A}_c+\mathbf{C}_c^T\mathbf{B}_a^T\mathbf{P}\mathbf{B}_a\mathbf{C}_c)-\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix}. \end{aligned} \tag{20}$$

It can be verified that both $\mathbf{R}_1=\mathbf{A}^T\mathbf{P}\mathbf{B}_a\mathbf{D}_c\mathbf{C}_f+\mathbf{C}_f^T\mathbf{D}_c^T\mathbf{B}_a^T\mathbf{P}\mathbf{A}+\varepsilon(\mathbf{C}_f^T\mathbf{D}_c^T\mathbf{B}_a^T\mathbf{P}\mathbf{B}_a\mathbf{D}_c\mathbf{C}_f+\mathbf{C}_f^T\mathbf{B}_c^T\mathbf{B}_c\mathbf{C}_f)$ and $\mathbf{R}_2=(\mathbf{A}+\varepsilon\mathbf{B}_a\mathbf{D}_c\mathbf{C}_f)^T\times\mathbf{P}\mathbf{B}_a\mathbf{C}_c+\varepsilon\mathbf{C}_f^T\mathbf{B}_c^T\mathbf{A}_c$ are bounded matrices. As $\varepsilon\to 0$, one can see that

$$\begin{aligned} L(t+1)-L(t) &\to \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix}^T \begin{bmatrix} \mathbf{A}^T\mathbf{P}\mathbf{A}-\mathbf{P} & \\ & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} \\ &= -\mathbf{v}^T(t)\mathbf{Q}\mathbf{v}(t) - \mathbf{w}^T(t)\mathbf{w}(t) < 0, \end{aligned} \tag{21}$$

where Eq. (12) has been substituted.

Equation (21) implies an exponentially decreasing $L(t)$ (magnitudes of closed-loop eigenvalues are smaller than one[7,10,11]), if ε is sufficiently small and yet different from zero. As a result, a SLANC system can stabilize the closed-loop system by (1) turning off $u_2(t)$ and (2) reducing ε and restarting Eq. (15) whenever partial or full value of $J(kT)$ exceeds a conservatively preset threshold.

B. Controller optimization

If $J(kT)$ is found bounded within the preset threshold, the closed-loop system is detected to be stable. A SLANC system will stop reducing ε and switch to the second task of optimizing $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$. In Eq. (16), an one-to-one mapping is created to map elements of $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$ to elements of $\boldsymbol{\theta}$. If estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ are reliable and dimension of $\mathbf{A}_c$ is larger than or equal to dimension of $\mathbf{A}$, there exists an optimal $\boldsymbol{\theta}_{\text{opt}}$ that minimizes $J(kT,\boldsymbol{\theta})\approx E\{p_e^2|\boldsymbol{\theta}\}$.[10,13] If estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ are unreliable, however, the only way to optimize $\boldsymbol{\theta}$ is trial-and-error. The SLANC system works periodically at the beginning of the $(2k+1)$th averaging period of Eq. (15). It is based on two learning rules. The first one is random learning

$$\boldsymbol{\theta}(2k+1) = \boldsymbol{\theta}(2k) + \delta\boldsymbol{\theta}(2k), \tag{22}$$

where $\boldsymbol{\theta}(2k)$ contains parameters of the best available guess; $\boldsymbol{\theta}(2k+1)$ is a tentative parameter vector; and $\delta\boldsymbol{\theta}(2k)$ is a randomly generated vector. At the end of $(2k+1)$th averaging period of Eq. (15), the system either updates $\boldsymbol{\theta}(2k)=\boldsymbol{\theta}(2k+1)$ if $J[2kT,\boldsymbol{\theta}(2k)]>J[(2k+1)T,\boldsymbol{\theta}(2k+1)]$ or otherwise keeps $\boldsymbol{\theta}(2k)$ unchanged. Although in most cases $\boldsymbol{\theta}(2k)$ is the winner of the previous competition, $J[2kT,\boldsymbol{\theta}(2k)]$ is still retested in each new competition since $\boldsymbol{\theta}(2k)$ could represent an unstable controller if parameters of the sound field change suddenly.

Unlike recursive system identifications algorithms, which update the controller in every δt s, the SLANC system waits every $2T\delta t \gg \delta t$ s before it updates $\boldsymbol{\theta}(2k)=\boldsymbol{\theta}(2k+1)$ if $J[2kT,\boldsymbol{\theta}(2k)]>J[(2k+1)T,\boldsymbol{\theta}(2k+1)]$. Otherwise $\boldsymbol{\theta}(2k)$ remains unchanged after $2T\delta t$ s of testing. The value of T is sufficiently large such that $J[2kT,\boldsymbol{\theta}(2k)]\approx E\{p_e^2|\boldsymbol{\theta}(2k)\}$ and $J[(2k+1)T,\boldsymbol{\theta}(2k+1)]\approx E\{p_e^2|\boldsymbol{\theta}(2k+1)\}$ when $n(t)$ is stationary and ergodic. If magnitude of the incremental vector $\delta\boldsymbol{\theta}(2k)$ is small enough, the learning process will reach at least a local minimum in the landscape of $E\{p_e^2|\boldsymbol{\theta}\}$, since each update makes $E\{p_e^2|\boldsymbol{\theta}\}$ smaller than before or at least unchanged.

It is very difficult to evaluate the probability of $J[\boldsymbol{\theta}(2k)]>J[\boldsymbol{\theta}(2k+1)]$ that depends on many factors including primary noise $n(t)$, locations of sensors and speakers, or even the initial values of $\boldsymbol{\theta}$, etc. It is therefore very difficult to predict the convergence speed of random learning. Let $\theta_i(k)$ be the ith element of vector $\boldsymbol{\theta}(k)$, the SLANC system uses $\delta J(k)=J[(k+1)T,\boldsymbol{\theta}(k+1)]-J[kT,\boldsymbol{\theta}(k)]$ and $\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)$ to estimate the gradient vector

$$\mathbf{g}^T = \begin{bmatrix} \dfrac{\partial J}{\partial\theta_1} & \dfrac{\partial J}{\partial\theta_2} & \cdots & \dfrac{\partial J}{\partial\theta_N} \end{bmatrix},$$

such that $\delta J(k)=\mathbf{g}^T[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]$ when magnitudes of $\boldsymbol{\theta}(k+1)$-$\boldsymbol{\theta}(k)$ are sufficiently small. Approximation error

$$\begin{aligned} e(k) &= \delta J(k) - \mathbf{g}^T(k)[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)] \\ &= [\mathbf{g}-\mathbf{g}(k)]^T[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)] \\ &= [\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]^T[\mathbf{g}-\mathbf{g}(k)] \end{aligned} \tag{23}$$

is used to measure the accuracy of estimated gradient $\mathbf{g}(k)$. Although the true gradient $\mathbf{g}$ is never known, $e(k)=\delta J(k)-\mathbf{g}(k)^T[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]$ can be obtained at the end of $(k+1)$th test period using $\delta J(k)=J[(k+1)T,\boldsymbol{\theta}(k+1)]-J[kT,\boldsymbol{\theta}(k)]$. A positive definite function

$$V(k)=[\mathbf{g}-\mathbf{g}(k)]^T[\mathbf{g}-\mathbf{g}(k)] \tag{24}$$

is used to represent estimation error of the gradient. Similar to identity $a^2-b^2=(a-b)(a+b)$, it can be derived that

$$V(k+1)-V(k)=[\mathbf{g}(k)-\mathbf{g}(k+1)]^T[2\mathbf{g}-\mathbf{g}(k+1)-\mathbf{g}(k)]. \tag{25}$$

A recursive algorithm

$$\mathbf{g}(k+1)=\mathbf{g}(k)+\frac{e(k)[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]}{\|\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)\|^2} \tag{26}$$

is proposed in this study to estimate $\mathbf{g}(k)$. It is mathematically equivalent to

$$\mathbf{g}(k+1)-\mathbf{g}(k)=\frac{e(k)[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]}{\|\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)\|^2}, \tag{27}$$

and

$$2\mathbf{g}-\mathbf{g}(k+1)-\mathbf{g}(k)=2[\mathbf{g}-\mathbf{g}(k)]-\frac{e(k)[\boldsymbol{\theta}(k+1)-\theta(k)]}{\|\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)\|^2}. \tag{28}$$

If one substitutes Eqs. (27) and (28) into Eq. (25), the result will be

$$V(k+1)-V(k)=-\frac{e(k)[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]^T}{\|\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)\|^2}\left\{2[\mathbf{g}-\mathbf{g}(k)]-\frac{e(k)[\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)]}{\|\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)\|^2}\right\}. \tag{29}$$

Substituting Eq. (23) into Eq. (29), one can finish the derivation with

$$V(k+1)-V(k)=\frac{-e^2(k)}{\|\boldsymbol{\theta}(k+1)-\boldsymbol{\theta}(k)\|^2}\leqslant 0. \tag{30}$$

It indicates monotonous decrease of $V(k)=[\mathbf{g}-\mathbf{g}(k)]^T[\mathbf{g}-\mathbf{g}(k)]$ until $e(k)\to 0$.

Upon the convergence of $\mathbf{g}(k)\approx\mathbf{g}$, which is signaled by $e(k)\to 0$, the learning rule may be changed from Eq. (22) to

$$\boldsymbol{\theta}(2k+1)=\boldsymbol{\theta}(2k)-\mu\mathbf{g}(2k)+\delta\boldsymbol{\theta}(2k) \quad \text{or} \quad \boldsymbol{\theta}(2k+1)-\boldsymbol{\theta}(2k)=-\mu\mathbf{g}(2k)+\delta\boldsymbol{\theta}(2k), \tag{31}$$

where μ is a small positive constant. One may use Eq. (31) and the gradient to predict

$$\delta J(2k)=\mathbf{g}^T[\boldsymbol{\theta}(2k+1)-\boldsymbol{\theta}(2k)]=-\mu\mathbf{g}^T\mathbf{g}(2k)+\mathbf{g}^T\delta\boldsymbol{\theta}(2k). \tag{32}$$

Due to the convergence of $\mathbf{g}(2k)\approx\mathbf{g}$, it is expected that $\mathbf{g}^T\mathbf{g}(2k)>0$ and hence

$$E\{\delta J(2k)\}=E\{\mathbf{g}^T[\boldsymbol{\theta}(2k+1)-\boldsymbol{\theta}(2k)]\}=-\mu\mathbf{g}^T\mathbf{g}(2k)<0, \tag{33}$$

where all elements of $\delta\boldsymbol{\theta}(k)$ are zero-mean random numbers such that $E\{\mathbf{g}^T\delta\boldsymbol{\theta}(2k)\}=0$. The statistical expectation of the second learning rule is to reduce $J(kT)$ in the averaging period of Eq. (15), as indicated by Eq. (33).

By incorporating online estimation of gradient vector, the second learning rule converges faster since it increases probability of $J[\boldsymbol{\theta}(2k)]>J[\boldsymbol{\theta}(2k+1)]$ as hinted by Eq. (33). However, it is still very difficult to evaluate the probability of $J[\boldsymbol{\theta}(2k)]>J[\boldsymbol{\theta}(2k+1)]$ and the convergence speed of improved self-learning. If online system identification is an optional part of a SLANC system, the identification algorithm never stops even if indirect adaptive control is not stable. The SLANC system keeps using estimates $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_f$, and $\mathbf{C}_e$ to update the indirect adaptive controller in a hope that these matrices will eventually become reliable. Accuracy and speed of convergence of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_f$, and $\mathbf{C}_e$ depend on specific online system identification algorithms. Although there are many available system identification algorithms, it is difficult to say which one is best in terms of accuracy and speed of convergence even for linear systems without disturbances. In many ANC applications, $n(t)$ is a strong disturbance and it is more difficult to predict how long it takes for online estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_f$, and $\mathbf{C}_e$ to become reliable. Periodically, the system allocates test periods of Eq. (15) for the indirect adaptive controller to challenge the best available controller. Performances of controllers are judged by the objective function at the end of test periods. The system keeps the winner that has a smaller $J(*)$ in the competition.

Since an indirect adaptive controller may be unstable if estimates $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_f$, and $\mathbf{C}_e$ are unreliable, the SLANC system will reduce ε and stabilize the indirect adaptive controller if it produces a growing $J(*)$. Once online estimates become accurate, the indirect controller will be optimized by available H_2/H_∞ control algorithms.[10,11,13] It is expected to be adopted by the SLANC system after winning a competition with a smaller $J(*)$.

C. Combination with feedforward control

When $J(kT)$ is found no longer decreasing after repeated testing of controllers, the SLANC system may stop updating the feedback controller by fixing on a controller that produces a smaller $J(kT)$ in the most recent test period. The SLANC system then switches to a third task of incorporating feedforward control if reference signal $n(t)$ is available. The closed-loop system, represented by Eqs. (17) and (18), may be expressed in the discrete-time z-transform domain as

$$p_e(z)=P(z)n(z)+S(z)u_2(z), \tag{34}$$

where $p_e(z)$, $n(z)$, and $u_2(z)$ are z-transform versions of $p_e(t)$, $n(t)$, and $u_2(t)$, respectively; primary path $P(z)$ and secondary path $S(z)$ transfer functions can be obtained from state space models[7,10] as

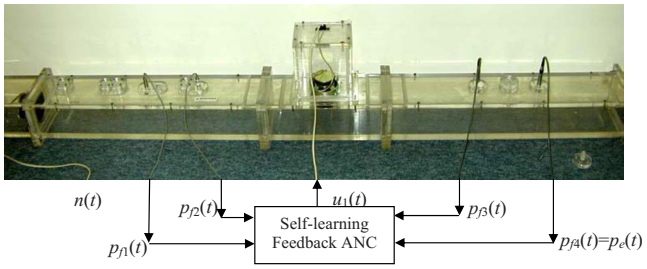


FIG. 2. (Color online) Experiment setup in a 2 m duct.

$$P(z) = [\mathbf{C}_e \quad] \begin{bmatrix} z\mathbf{I} - \mathbf{A} - \varepsilon \mathbf{B}_a \mathbf{D}_c \mathbf{C}_f & -\varepsilon \mathbf{B}_a \mathbf{C}_c \\ -\varepsilon \mathbf{B}_c \mathbf{C}_f & z\mathbf{I} - \varepsilon \mathbf{A}_c \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}_p \\ \end{bmatrix}, \tag{35}$$

and

$$S(z) = [\mathbf{C}_e \quad] \begin{bmatrix} z\mathbf{I} - \mathbf{A} - \varepsilon \mathbf{B}_a \mathbf{D}_c \mathbf{C}_f & -\varepsilon \mathbf{B}_a \mathbf{C}_c \\ -\varepsilon \mathbf{B}_c \mathbf{C}_f & z\mathbf{I} - \varepsilon \mathbf{A}_c \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}_a \\ \end{bmatrix}. \tag{36}$$

Both $P(z)$ and $S(z)$ contain parameters of the feedback controller. Since the feedback controller is tested and updated in online operation, it is not possible to obtain $P(z)$ and $S(z)$ by initial offline identification. A recently developed algorithm, known as orthogonal adaptation,[14,15] can be combined with a SLANC system to generate destructive interference. Let $C_2(z)$ denote transfer function of a feedforward controller such that $u_2(z)=C_2(z)n(z)$, then orthogonal adaptation is able to minimize H_2 norm $\|P(z)+S(z)C_2(z)\|_2$ using online estimates of $P(z)$ and $S(z)$. Details of orthogonal adaptation have been published elsewhere[14,15] and omitted here.

V. EXPERIMENT

An experiment was conducted to test a SLANC system in the worst case when self-learning is the only option for stable and optimal operation. Figure 2 illustrates the experimental setup in a 2 m duct with a cross-sectional area of 12×15 cm^2. The primary source was a 4 in. speaker placed at the upstream end of the duct. It was excited by the pseudorandom noise. The secondary actuator was a 4 in. speaker placed at the middle of the duct. Four feedback sensors were placed on both sides of the actuator. The most downstream sensor was also the error sensor.

The system sampling rate was 2500 Hz. Dimensions of $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}_c$, and $\mathbf{D}_c$ were 200×200, 200×4, 1×200, and 1×4, respectively. Parameters of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ were completely unknown to the SLANC system. Eigenvalues of $\mathbf{A}$ are all inside the unit circle centered at the origin of the complex plane, since the sound field is open-loop stable. The objective function was computed using Eq. (15) with $T=200\,000$ samples. Analog signals in the system were low-pass filtered with a cutoff frequency of 800 Hz before digitized into the SLANC system. The SLANC system was programmed in standard C language and implemented on a dSPACE 1103 board. Although it is possible to combine the SLANC system with indirect adaptive control, the option was disabled to focus on the worst case exclusively. Reference signal $n(t)$ was not available to the controller. Only the feedback control part was tested in the experiment.

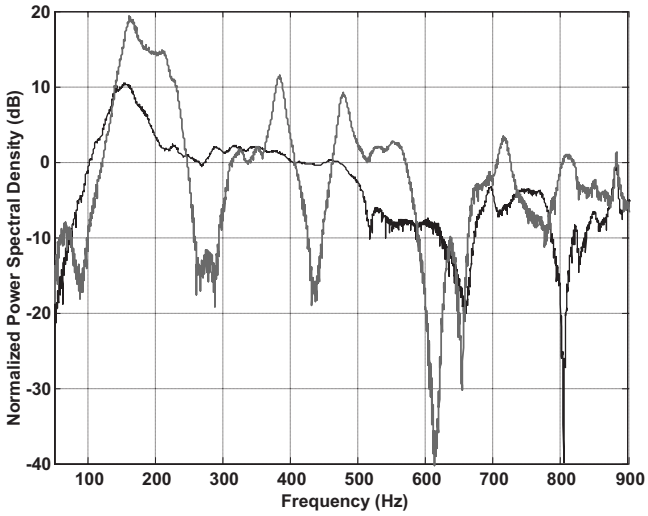


FIG. 3. Normalized power spectral densities of error signal when SLANC was off (gray) and on (black).

Two sets of experimental data were collected for the cases with/without control, respectively. Each set of data was the power spectral density (PSD) of the error signal normalized by the PSD of the primary noise signal. The two normalized PSDs are plotted in Fig. 3, where the black and gray curves represent, respectively, normalized PSD of the error signal for the cases with/without active control. When the controller was off, resonant effects are evidently seen in the gray PSD with significant resonant peaks and antiresonant dips. Since the reference signal was not available to the SLANC system in the experiment, feedback control was only able to introduce active damping to reduce resonant effects. It reduced resonant peaks and filled up antiresonant dips simultaneously, as demonstrated by the black PSD in Fig. 3. The SLANC system worked very well to optimize active damping in a wide frequency range (20–800 Hz), even when parameters of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_f$, $\mathbf{C}_f$, and $\mathbf{C}_e$ were completely unavailable.

Self-learning control is the major difference between a SLANC system and other available MIANC systems. The experiment was intended to test the feedback part of a SLANC system in a worst case when the reference signal was unavailable and online system identification results were unreliable. The feedforward part and indirect adaptive control were not tested because both options are not results of this study. Compared with available feedback MIANC systems such as direct rate feedback control or active resonators, the SLANC system tolerates dynamic uncertainties in the ANC system, avoids sensing near fields, and is able to minimize objective function $J(kT)$, as verified by the experiment.

VI. CONCLUSIONS

The major difference between a SLANC system and available MIANC systems is self-learning in the worst case when (i) model parameters of the sound field are unreliable due to online estimation errors and (ii) reference signal is

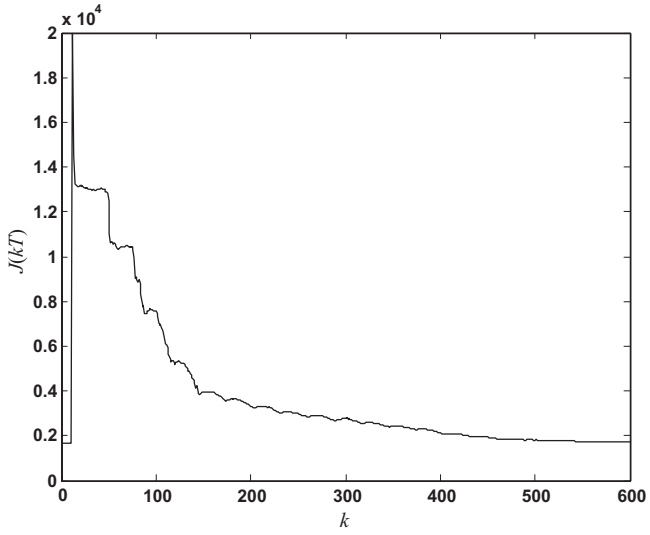


FIG. 4. Values of $J(kT)$ as a function of k.

unavailable. The SLANC system tolerates dynamic uncertainties in sound fields to stabilize the closed-loop system by tuning controller parameters.

A curve of $J(kT)$ is plotted in Fig. 4 as a function of k in the worst case without online system identification. The controller started with parameters learned previously. Its parameters were replaced with random numbers when $k=10$. The system took a very large number of k to converge. This is a drawback of self-learning and also a reason why indirect adaptive control is incorporated in Fig. 1. Available system identification algorithms converge quickly and accurately in the absence of disturbances like $n(t)$. When $n(t)$ acts as the disturbance, there is no guarantee on the accuracy of parameter estimates. If estimates of $\mathbf{A}$, $\mathbf{B}_a$, $\mathbf{B}_p$, $\mathbf{C}_e$, and $\mathbf{C}_f$ are accurate, indirect adaptive control is optimal; otherwise it may be unstable. Self-learning control is a stable backup whenever indirect adaptive control is unstable. Since the system allocates test periods to both tuning blocks periodically, indirect adaptive control is expected to win a competition whenever it is nearly optimal. Afterwards, elements of $\boldsymbol{\theta}(2k)$ become parameters of the indirect adaptive controller. Elements of $\boldsymbol{\theta}(2k+1)$ are either (i) slightly modified elements of $\boldsymbol{\theta}(2k)$ by Eq. (31) or (ii) parameters of a fresher version of the indirect adaptive controller. In case (ii), $\boldsymbol{\theta}(2k+1)$ is not necessarily better than $\boldsymbol{\theta}(2k)$ due to estimation errors. If $\boldsymbol{\theta}(2k+1)$ represents an unstable controller, the SLANC system will stabilize the closed loop by reducing ε. If Eq. (16) were exact, the competitions would either improve $\boldsymbol{\theta}(2k)$ if $E\{p_e^2|\boldsymbol{\theta}(2k)\} > E\{p_e^2|\boldsymbol{\theta}(2k+1)\}$ or otherwise keep $\boldsymbol{\theta}(2k)$ unchanged. Due to finite T and approximation errors in Eq. (16), values of $J(kT,\boldsymbol{\theta})$ fluctuate slightly for different k even when a constant $\boldsymbol{\theta}$ represents a stable controller. The fluctuation may cause sporadic false updates of $\boldsymbol{\theta}(2k)$, but the SLANC system generally remains optimal or nearly optimal until parameters of the sound field change and the entire process starts over again.

If the reference signal is available, a SLANC system may be combined with a feedforward controller to apply both destructive interference and active damping to sound fields. Experimental results are presented to demonstrate the performance of a SLANC system in the worst case. The SLANC system proves to be a simple and stable tool for many feedback ANC systems to deal with the worst case discussed in this study.

ACKNOWLEDGMENTS

The writer thanks the reviewers for their penetrating comments and suggestions that led to an improved version of the manuscript. This project was partially supported by Internal Grant No. G-U392 from the Hong Kong Polytechnic University.

[1]C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E and FN Spon, London, 1997).
[2]P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).
[3]R. L. Clark and D. G. Cole, "Active damping of enclosed sound fields through direct rate feedback control," J. Acoust. Soc. Am. **97**, 1710–1716 (1995).
[4]H. R. Pota and A. G. Kelkar, "Modeling and control of acoustic ducts," ASME J. Vibr. Acoust. **123**, 2–10 (2001).
[5]J. D. Kemp and R. L. Clark, "Noise reduction in a launch vehicle fairing using actively tuned loudspeakers," J. Acoust. Soc. Am. **113**, 1986–1994 (2003).
[6]J. B. Bisnette, J. S. Vipperman, and D. D. Budny, "Active noise control using damped resonator filters," J. Acoust. Soc. Am. **113**, 2228–2234 (2003).
[7]J. Dorsey, *Continuous and Discrete Control Systems: Modeling, Identification, Design, and Implementation* (McGraw-Hill, Boston, 2002).
[8]D. E. Hall, *Basic Acoustics* (Wiley, New York, 1987).
[9]S. Liu, J. Yuan, and K. Y. Fung, "Robust active control of broadband noise in finite ducts," J. Acoust. Soc. Am. **111**, 2727–2734 (2002).
[10]S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory* (SIAM Books, Philadelphia, 1994).
[11]B. A. Frances, "A course in $\mathbf{H}_\infty$ control theory," in *Lecture Notes in Control and Information Sciences*, edited by M. Thoma and A. Wyner (Springer, New York, 1987).
[12]G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
[13]P. P. Khargonekar and M. A. Rotea, "Mixed H_2/H_∞ control: A convex optimization approach," IEEE Trans. Autom. Control **39**, 824–837 (1991).
[14]J. Yuan, "Orthogonal adaptation for active noise control," J. Acoust. Soc. Am. **120**, 204–210 (2006).
[15]J. Yuan, "Orthogonal adaptation for multichannel feedforward control," J. Acoust. Soc. Am. **120**, 3723–3729 (2006).

On the statistical errors in the estimate of acoustical energy density by using two microphones in a one dimensional field

Jean-Claude Pascal[a] and Jean-Hugh Thomas[b]
Laboratoire d'Acoustique de l'Université du Maine (CNRS UMR 6613) and Ecole Nationale Supérieure d'Ingénieurs du Mans (ENSIM), Université du Maine, rue Aristote, 72085 Le Mans Cedex, France

Jing-Fang Li[c]
Visual VibroAcoustics, 51 rue d'Alger, 72000 Le Mans, France

(Received 27 November 2007; revised 15 July 2008; accepted 24 July 2008)

It was recently shown that the statistical errors of the measurement in the acoustic energy density by the two microphone method in waveguide have little variation when the losses of coherence between microphones increase. To explain these intervals of uncertainty, the variance of the measurement is expressed in this paper as a function of the various energy quantities of the acoustic fields—energy densities and sound intensities. The necessary conditions to reach the lower bound are clarified. The results obtained are illustrated by an example of a one-dimensional partially coherent field, which allows one to specify the relationship between the coherence functions of the pressure and particle velocity and those of the two microphone signals.
 [DOI: 10.1121/1.2973232]

PACS number(s): 43.58.Fm, 43.60.Cg, 43.60.Qv [EGW] Pages: 2085–2089

I. INTRODUCTION

From the analysis of the statistical error of the acoustic energy density measured by two microphones, Ghan *et al.*[1] showed that the normalized standard deviation for a Gaussian broadband signal varies weekly in function of the kind of sound field, within an upper bound $1/\sqrt{BT}$ and a lower bound $1/\sqrt{2BT}$ (where B is the bandwidth of the analysis and T is the time length of acquisition). The upper bound is that of the normalized standard deviation of a potential energy density (or quadratic pressure) measured by a single microphone. The fact that in certain cases the standard deviation of the energy density (sum of potential and kinetic energy densities) can even be $\sqrt{2}$ times weaker than the upper bound is surprising at a first look. Indeed, studies of the statistical errors of the acoustic intensity[2–7] using a probe consisting of two microphones have shown that the normalized standard deviation always has $1/\sqrt{BT}$ for lower bound and can take much more significant values being dependent on the phase and the loss of coherence between the two microphone signals. To understand the significance of this result, in this paper, the statistical error of the energy density is analyzed for the case of two partially coherent waves in opposite directions. It is often thought that the use of a discretization scheme by a finite difference approximation is the cause of the increase of statistical errors. It is shown in this paper that the use of the finite difference approximation has a little influence on the statistical errors of the energy density which can entirely be expressed as a function of the quadratic quantities (intensities and energy densities) of acoustic fields.

[a]Electronic mail: jean-claude.pascal@univ-lemans.fr
[b]Electronic mail: jean-hugh.thomas@univ-lemans.fr
[c]Electronic mail: jinfang.li@visualvibroacoustics.com

II. EXPRESSION FOR THE STATISTICAL ERRORS OF THE ENERGY DENSITIES

Given a one-dimensional sound field with only the x-component of the particle velocity u_x, the potential V, kinetic T, and total E energy densities are written by

$$V = \frac{1}{2\rho_0 c^2} G_{pp}(\omega), \quad T = \frac{\rho_0}{2} G_{u_x u_x}(\omega), \quad E = V + T, \tag{1}$$

where ρ_0 is the mass density of the fluid and c is the speed of sound. $G_{pp}(\omega)$ and $G_{u_x u_x}(\omega)$ are, respectively, the autospectral power densities of pressure and x-particle velocity. The pressure measured by two microphones with spacing of Δ ($\Delta < \lambda/3$) allows one to give the finite-difference approximation expressions[8] for the particle velocity $u_x \approx (p_2 - p_1)/(-j\rho_0 c k \Delta)$ and the pressure estimation $p \approx (p_2 + p_1)/2$, with the wavenumber $k = \omega/c$. Using these expressions, the approximation of the energy densities can then be expressed in the following form:

$$D_a(\omega) = \alpha[G_{11}(\omega) + G_{22}(\omega)] + \beta[2C_{21}(\omega)]. \tag{2}$$

The subscript a indicates an approximate quantity. D_a corresponds to V_a, T_a, or E_a according to the values of the coefficients α and β in Table I. The one-sided autospectral $G_{11}(\omega)$, $G_{22}(\omega)$ and cross-spectral $G_{12}(\omega)$ densities are given by

$$G_{ij}(\omega) = \lim_{T_W \to \infty} \frac{2}{T_W} E\{p_j(\omega, T_W) p_i^*(\omega, T_W)\} \quad (i,j = 1,2), \tag{3}$$

$2C_{21}(\omega)$ is the shortened notation for $\mathrm{Re}\{G_{21}(\omega)\} + j\,\mathrm{Im}\{G_{21}(\omega)\}$. In Eq. (3), $p_1(\omega, T_W)$ and $p_2(\omega, T_W)$ are the finite Fourier transforms of length T_W.

TABLE I. Coefficients α and β used in the expressions of the energy densities by the finite-difference approximations (Eqs. (2) and (9)).

Energy quantity	α	β
potential energy V_a	$1/(8\rho_0c^2)$	$1/(8\rho_0c^2)$
kinetic energy T_a	$1/(2\rho_0c^2k^2\Delta^2)$	$-1/(2\rho_0c^2k^2\Delta^2)$
total energy E_a	$1/(8\rho_0c^2)+1/(2\rho_0c^2k^2\Delta^2)$	$1/(8\rho_0c^2)-1/(2\rho_0c^2k^2\Delta^2)$

$E\{p_1(\omega,T_W)p_2^*(\omega,T_W)\}$ denotes the expected value of $p_1(\omega,T_W)p_2^*(\omega,T_W)$.

By evaluating the cross-spectral power densities,

$$\hat{G}_{21}(\omega)=(2/n)\sum_{i=1}^{n}[p_1(\omega,T_W)p_2^*(\omega,T_W)]/T_W, \tag{4}$$

with a time window of length T_W and an average of n records, the statistical errors in the estimator appear if the pressures are represented by random signals. For bivariate random processes, the variance of the energy densities can be written in terms of the covariance of power densities by

$$\mathrm{var}\{\hat{D}_a(\xi_1,\xi_2,\ldots)\}\approx\sum_{i,j}\frac{\partial\hat{D}_a}{\partial\xi_i}\frac{\partial\hat{D}_a}{\partial\xi_j}\mathrm{cov}\{\xi_i,\xi_j\}$$

$$\text{with } \xi_i,\xi_j\in\{\hat{G}_{11}(\omega),\hat{G}_{22}(\omega),\hat{C}_{21}(\omega)\}. \tag{5}$$

Analytical expressions of covariances were given by Jenkins and Watts[9] for Gaussian processes as a function of the factor BT (B is the width of an elementary filter corresponding to a frequency bin of the fast Fourier transform FFT analysis, which leads to $BT_W=1$, $T=nT_W$, and $BT=n$)

$$\mathrm{cov}\{\hat{G}_{ii},\hat{G}_{jj}\}=|G_{ij}|^2/n,$$

$$\mathrm{cov}\{\hat{G}_{ii},\hat{C}_{ij}\}=G_{ii}C_{ij}/n \quad \text{with } i,j=\{1,2\}, \tag{6}$$

and

$$\mathrm{cov}\{\hat{G}_{ii},\hat{G}_{ii}\}=\mathrm{var}\{\hat{G}_{ii}\}=G_{ii}^2/n, \tag{7}$$

$$\mathrm{cov}\{\hat{C}_{21},\hat{C}_{21}\}=\mathrm{var}\{\hat{C}_{21}\}=(G_{11}G_{22}+C_{21}^2-Q_{21}^2)/(2n). \tag{8}$$

When considering the coherence function $\gamma_{21}^2(\omega)$ between the two microphone signals and by using the relation $C_{21}^2+Q_{21}^2=|G_{21}|^2=\gamma_{21}^2G_{11}G_{22}$, the expression for the variance of the energy densities is written in the following form:

$$\begin{aligned}\mathrm{var}\{\hat{D}_a(\omega)\}&\approx\frac{1}{n}[D_a^2(\omega)-2(\alpha^2-\beta^2)(G_{11}(\omega)G_{22}(\omega)\\&\quad-|G_{21}(\omega)|^2)]\\&\approx\frac{1}{n}[D_a^2(\omega)-2(\alpha^2-\beta^2)G_{11}(\omega)\\&\quad\times G_{22}(\omega)(1-\gamma_{21}^2(\omega))].\end{aligned} \tag{9}$$

To simplify the notations, we will consider, in the rest of this paper, the variance $\mathrm{var}\{\hat{D}_a(\omega)\}$ rather than the normalized standard deviation $\varepsilon\{\hat{D}_a(\omega)\}$ by knowing that $\varepsilon\{\hat{D}_a(\omega)\}=(\mathrm{var}\{\hat{D}_a(\omega)\})^{1/2}/\hat{D}_a(\omega)$. By considering the potential and kinetic energy densities for which $\alpha^2=\beta^2$ (Table I), the following equation can be obtained as

$$\mathrm{var}\{\hat{V}_a(\omega)\}\approx\frac{V_a^2(\omega)}{n} \quad \text{and } \mathrm{var}\{\hat{T}_a(\omega)\}\approx\frac{T_a^2(\omega)}{n}. \tag{10}$$

The variance of the statistical errors for a probe with two microphones is the same as that obtained from Eq. (1) for a single pressure sensor $\hat{V}(\omega)=\hat{G}_{pp}(\omega)/(2\rho_0c^2)$ and for a single particle velocity sensor $\hat{T}(\omega)=\rho_0\hat{G}_{uu}(\omega)/2$, because the variance of an autospectrum[9] is always equal to $G_{ii}^2(\omega)/n$. This result had already been obtained by Elko.[5] The loss of coherence between the two microphones has no influence, whether it is caused by the nature of the sound fields or by the independent electronic noise of each channel. However, it is not the same for the total energy density since

$$\alpha^2-\beta^2=1/(2\rho_0c^2k\Delta)^2, \tag{11}$$

and the second term on the right-hand side of Eq. (9) depending on the coherence function is not equal to zero. This term is always positive and will always result in subtraction from E_a^2. Since the probes are used in the range $k\Delta<1$, this term may have significant values at low frequencies, in particular, when the coherence $\gamma_{21}^2(\omega)$ is appreciably lower than 1. Ghan *et al.*[1] studied this expression of the variance of the total energy density and defined its intervals of variation as follows:

$$\frac{E_a^2(\omega)}{2n}\le\mathrm{var}\{\hat{E}_a(\omega)\}\le\frac{E_a^2(\omega)}{n}. \tag{12}$$

The lower bound can seem paradoxical inasmuch as the variance could be smaller than (i) that of the measurement by using only one sensor or (ii) that of a measurement made in a sound field having a full spatial coherence $\gamma_{21}^2=1$ (which corresponds to the upper bound).

A different factorization from Eq. (9) allows the variance of the total energy density to be written in the following form:

$$\begin{aligned}\mathrm{var}\{\hat{E}_a(\omega)\}&\approx\frac{1}{n}\Bigg[V_a^2(\omega)+T_a^2(\omega)\\&\quad+\frac{4Q_{21}^2(\omega)+(G_{11}(\omega)-G_{22}(\omega))^2}{2(2\rho_0c^2k\Delta)^2}\Bigg],\end{aligned} \tag{13}$$

in which one can recognize the approximate expressions, respectively, for the active acoustic intensity $I_a(\omega)=Q_{21}(\omega)/(\rho_0ck\Delta)$ and the reactive acoustic intensity $J_a(\omega)=(G_{11}(\omega)-G_{22}(\omega))/(2\rho_0ck\Delta)$ for a probe with two microphones.[3,10] Thus Eq. (13) takes the following remarkable form:

$$\mathrm{var}\{\hat{E}_a(\omega)\}\approx\frac{1}{n}\Bigg[V_a^2(\omega)+T_a^2(\omega)+\frac{I_a^2(\omega)}{2c^2}+\frac{J_a^2(\omega)}{2c^2}\Bigg]. \tag{14}$$

III. INTERPRETATION

Without considering the finite difference approximations, the variance of the total energy density can be calculated by applying Eq. (5) to Eq. (1):

$$\mathrm{var}\{\hat{E}(\omega)\} \approx \left(\frac{1}{2\rho_0 c^2}\right)^2 \mathrm{var}\{G_{pp}(\omega)\} + \left(\frac{\rho_0}{2}\right)^2 \mathrm{var}\{G_{uu}(\omega)\} + \frac{1}{2c^2}\mathrm{cov}\{G_{pp}(\omega), G_{uu}(\omega)\}. \quad (15)$$

The coherence function between the pressure and the particle velocity in the sound field was defined as[11]

$$\gamma_{up}^2(\omega) = \frac{|G_{up}(\omega)|^2}{G_{pp}(\omega)G_{uu}(\omega)}, \quad (16)$$

and by using Eq. (6), the covariance term becomes $\mathrm{cov}\{G_{pp}(\omega), G_{uu}(\omega)\} \approx (1/n)\gamma_{up}^2(\omega)G_{pp}(\omega)G_{uu}(\omega)$. Using Eq. (7), the variance of the total energy density is finally expressed in the following two forms:

$$\mathrm{var}\{\hat{E}(\omega)\} \approx \frac{1}{n}[V^2(\omega) + T^2(\omega) + 2\gamma_{up}^2(\omega)V(\omega)T(\omega)] \quad (17a)$$

or

$$\mathrm{var}\{\hat{E}(\omega)\} \approx \frac{1}{n}[E^2(\omega) - 2(1-\gamma_{up}^2(\omega))V(\omega)T(\omega)]. \quad (17b)$$

Equation (17b) allows one to obtain the bounds of the variance according to whether the coherence function between pressure and particle velocity takes a value 0 or 1:

$$\frac{V^2(\omega)}{n} + \frac{T^2(\omega)}{n} \le \mathrm{var}\{\hat{E}(\omega)\} \le \frac{E^2(\omega)}{n}. \quad (18)$$

The lower bound is thus reached when the covariance term is zero. It is noted that this lower bound is two times smaller than the upper bound, when the potential energy density is equal to the kinetic energy density. One can also note that a zero value of the potential or kinetic energy density will cancel the cross term and will result in reaching the upper bound.

Now by considering the third term on right-hand side of Eq. (15) which can be written in the form $\mathrm{cov}\{G_{pp}(\omega), G_{uu}(\omega)\} \approx (1/n)|G_{up}(\omega)|^2$ and by expressing the complex intensity[3,10] as $G_{up}(\omega) = I(\omega) + jJ(\omega)$, the variance takes the same form as that in the expression (14) without, however, using the approximation values:

$$\mathrm{var}\{\hat{E}(\omega)\} \approx \frac{1}{n}\left[V^2(\omega) + T^2(\omega) + \frac{I^2(\omega)}{2c^2} + \frac{J^2(\omega)}{2c^2}\right]. \quad (19)$$

Equation (19) shows clearly that the lower bound $E^2(\omega)/2n$ is reached in a sound field where the active and reactive intensities are equal to zero, and where the energy densities are equal to $V(\omega) = T(\omega) = E(\omega)/2$. It now remains to specify this type of sound field.

The probe consisting of two microphones for measurement of the energy densities can be used only in one-dimensional fields because only one component of the particle velocity is measured for computing the kinetic energy density. A general model of a one-dimensional partially coherent field can be represented by two random plane waves traveling in opposite directions:

$$p(x,t) = a\left(t - \frac{x}{c}\right) + b\left(t + \frac{x}{c}\right), \quad (20)$$

where a and b are, respectively, the random amplitudes of two plane waves traveling in opposite directions. The energy quantities can thus be expressed as a function of the auto- and the cross-spectra between the amplitudes of the waves:

$$V(\omega) = \frac{G_{AA}(\omega) + G_{BB}(\omega) + 2G_{AB}(\omega)\cos 2kx}{2\rho_0 c^2},$$

$$T(\omega) = \frac{G_{AA}(\omega) + G_{BB}(\omega) - 2G_{AB}(\omega)\cos 2kx}{2\rho_0 c^2}, \quad (21)$$

$$I(\omega) = \frac{G_{AA}(\omega) - G_{BB}(\omega)}{\rho_0 c}, \quad J(\omega) = 2\frac{G_{AB}(\omega)}{\rho_0 c}\sin 2kx,$$

where $G_{AB}(\omega) = \gamma_{AB}(\omega)\sqrt{G_{AA}(\omega)G_{BB}(\omega)}e^{j\varphi_{AB}}$ (the phase φ_{AB} is set to 0 in what follows in order to simplify the expressions without loss of generality). The coherence $\gamma_{AB}^2(\omega)$ between the two components can vary from 0 (two independent progressive plane waves) to 1 (quasistanding wave). The approximate expression for the energy densities of Eq. (4) is obtained from the computations of the pressure at the two microphone positions $x_1 = x - \Delta/2$ and $x_2 = x + \Delta/2$ by the use of Eq. (20):

$$D_a(\omega) = \alpha[2(G_{AA}(\omega) + G_{BB}(\omega)) + 4G_{AB}(\omega)\cos 2kx \cos k\Delta] + \beta[2(G_{AA}(\omega) + G_{BB}(\omega))\cos k\Delta + 4G_{AB}(\omega)\cos 2kx]. \quad (22)$$

From Eqs. (21) and (22), one can derive the expressions for the potential and kinetic energy densities:

$$V_a(\omega) = V(\omega)\cos^2 k\Delta/2, \quad T_a(\omega) = T(\omega)\left(\frac{\sin k\Delta/2}{k\Delta/2}\right)^2, \quad (23)$$

and the expressions for the active and reactive acoustic intensities:[10]

$$I_a(\omega) = I(\omega)\frac{\sin k\Delta}{k\Delta}, \quad J_a(\omega) = J(\omega)\frac{\sin k\Delta}{k\Delta}. \quad (24)$$

IV. DISCUSSION

According to Eq. (19), the lower bound of the variance $E^2(\omega)/2n$ is reached when the acoustic intensities are equal to zero and when the potential and kinetic energies are equal. Equation (21) shows that the active intensity is canceled when the two opposite waves have identical amplitudes. To obtain, at the same time, a cancellation of the reactive intensity and an equality of the potential and kinetic energy densities, it is necessary that the fluctuations due to the interferences disappear, i.e., that the two waves are completely incoherent. The same reasoning is transposable with the

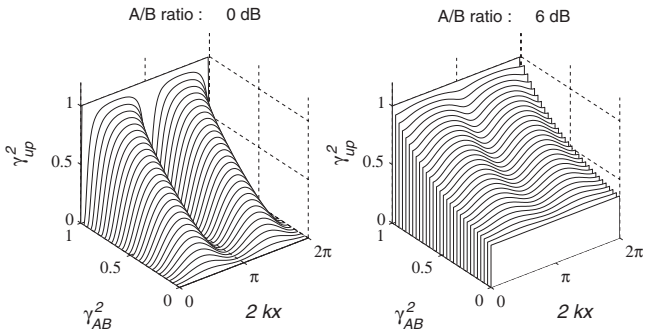


FIG. 1. Pressure-velocity coherence function γ^2_{up} of two waves, respectively, with amplitudes A and B traveling in opposite directions vs the coherence γ^2_{AB} of the two waves and $2kx$ for two A/B ratios indicated.

measurement by a probe consisting of two microphones by considering Eq. (14). Because of the systematic errors due to the finite-difference-approximation method used for evaluation of the potential and kinetic energy densities, the value of the variance of $\hat{E}(\omega)$ is always slightly higher than the theoretical value of the lower bound when $k\Delta$ increases.

According to the above results, the coherence function between the pressure and the particle velocity that appeared in Eqs. (17a) and (17b) and is given by Eq. (16) can also be written as follows:

$$\gamma^2_{up}(\omega) = \frac{I^2(\omega) + J^2(\omega)}{4c^2 V(\omega) T(\omega)}. \tag{25}$$

Uncorrelated two opposite waves ($\gamma^2_{AB}=0$) results in canceling the reactive intensity $J(\omega)$, but it is also necessary that these waves have identical amplitudes, so that the active intensity is canceled. Equation (25) shows that these two conditions lead to $\gamma^2_{up}(\omega)=0$. The null coherence between pressure and velocity is thus a sufficient condition to reach the lower bound of the variance, as it is predicted by Eqs. (17a) and (17b). The function of the coherence between pressure and velocity γ^2_{up} can be expressed in terms of the coherence γ^2_{AB} between the two opposite waves, the A/B ratio $r^2(\omega) = G_{BB}(\omega)/G_{AA}(\omega)$, and the position in the quasistanding wave

$$\gamma^2_{up}(\omega) = \frac{(1 - r^2(\omega))^2 + 4\gamma^2_{AB}(\omega) r^2(\omega) \sin^2 2kx}{(1 + r^2(\omega))^2 - 4\gamma^2_{AB}(\omega) r^2(\omega) \cos^2 2kx}. \tag{26}$$

Figure 1 shows the evolution of the coherence between the pressure and velocity γ^2_{up} versus γ^2_{AB} and kx for two amplitude ratios r^2. When two opposite waves have the same amplitudes ($r^2=0$ dB), the coherent part of the field creates quasistationary waves and γ^2_{up} depends also on the position of nodes ($kx=n\pi/2$, n is an integer number). When two opposite waves are incoherent, the coherence between pressure and velocity γ^2_{up} vanishes (neither active nor reactive intensity exists). When the amplitude of one wave is much greater than the others (r^2 is around 6 dB), the phenomenon is less evident and γ^2_{up} is never zero.

The variance of the energy density is often calculated from Eq. (9) in which the coherence between microphones appears. It is related to the pressure-velocity coherence function but evolves differently. For the two microphones positioned, respectively, at $x_1 = x - \Delta/2$ and $x_2 = x + \Delta/2$, using Eq. (20), the coherence between the two microphones can be obtained:

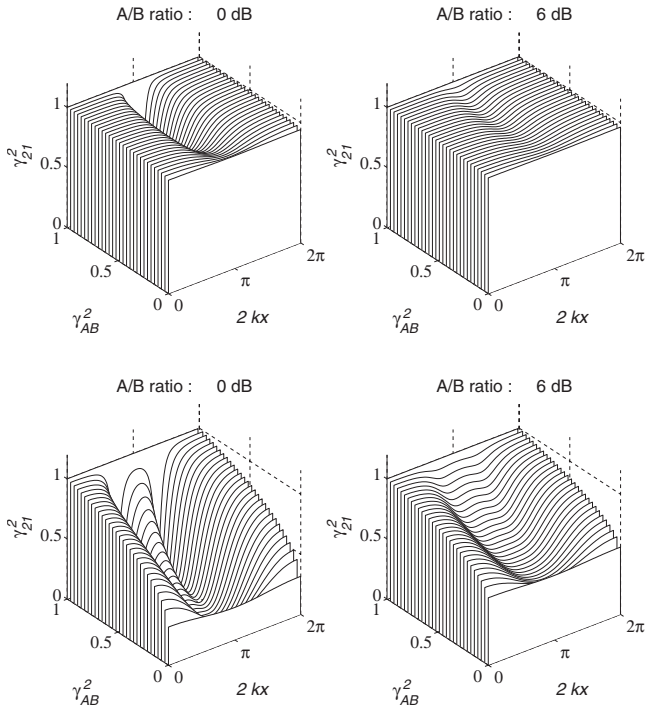


FIG. 2. Coherence function between two microphone signals γ^2_{12} vs the coherence of two waves traveling in opposite directions γ^2_{AB} and $2kx$ for two A/B ratios indicated and when $k\Delta=0.25$ (top), $k\Delta=1$ (bottom).

$$\gamma^2_{21}(\omega) = \frac{|G_{21}(\omega)|^2}{G_{11}(\omega) G_{22}(\omega)} = \frac{f(r(\omega), \gamma_{AB}(\omega), kx, k\Delta) - 4r^2(\omega) \sin^2 k\Delta}{f(r(\omega), \gamma_{AB}(\omega), kx, k\Delta) - 4\gamma^2_{AB}(\omega) r^2(\omega) \sin^2 k\Delta}, \tag{27}$$

where

$$f(r(\omega), \gamma_{AB}(\omega), kx, k\Delta) = (1 + r^2(\omega))^2 + 4\gamma_{AB}(\omega) r(\omega) \times (1 + r^2(\omega)) \cos 2kx \cos k\Delta + 4\gamma^2_{AB}(\omega) r^2(\omega) \cos^2 2kx.$$

γ^2_{21} is equal to 1 when the two opposite waves are totally coherent. If this condition is not satisfied, it tends toward 1 when the microphone spacing becomes small in comparison to the wavelength (in the low frequency range, when $k\Delta \ll 1$). Figure 2 shows the coherence between the two microphones under the same conditions as in Fig. 1 for the values $k\Delta=0.25$ and $k\Delta=1$ (upper bound of the frequency range). The coherence between the two microphones γ^2_{21} is very different from that between the pressure and velocity γ^2_{up}. γ^2_{21} is never zero and is dependent on the microphone spacing $k\Delta$, where $k\Delta = 2\pi\Delta/\lambda$, λ is the wavelength. Any loss of coherence between the signals of the two microphones will tend to approach the lower bound of the statistical error, just like the addition of independent noise to each of the two microphone channels.

V. CONCLUSION

The normalized variance of the total energy density has an upper bound equal to E^2/n ($n=BT$), like any quadratic value of a physical quantity. It is shown that this variance can be expressed as a function of the various energy quantities of sound fields, such as potential and kinetic energy densities, and active and reactive acoustic intensities. It is the nature of the sound field that determines the statistical errors not the use of the discretization by finite difference. It appears that the value of the variance becomes two times lower when the active and the reactive acoustic intensities vanish simultaneously. However, this condition is not sufficient. It is also necessary that the potential and kinetic energy densities are equal at all points of the acoustic field. This can be obtained only when the coherence between the pressure and acoustic particle velocity is equal to zero. In a one-dimensional sound field, this condition is obtained when the two waves traveling in opposite directions are independent and of equal amplitudes. The relationships between the pressure-velocity coherence function and the coherence measured between the two microphone signals, by which the energy quantities can be experimentally determined, were specified.

[1] J. Ghan, B. S. Cazzolato, and S. D. Snyder, "Statistical errors in the estimation of time-averaged acoustic energy density using the two-microphone method," J. Acoust. Soc. Am. **115**, 1179–1184 (2004).

[2] A. F. Seybert, "Statistical errors in acoustic intensity measurements," J. Sound Vib. **75**, 519–526 (1981).

[3] J.-C. Pascal, "Measurements of the active and reactive intensity in various acoustic fields," Proceedings of the Conference on Recent Developments in the Acoustical Intensity Measurements, Senlis (France), 30 September–2 October 1981, edited by M. Bockhoff, Centre Technique des Industries Mécaniques, Senlis, France, pp. 11–19.

[4] O. Dyrlund, "A note on statistical errors in acoustic intensity measurements," J. Sound Vib. **90**, 585–589 (1983).

[5] G. W. Elko, "Frequency domain estimation of the complex acoustic intensity and acoustic energy density." Ph.D. thesis, Pennsylvania State University, State College, 1984.

[6] F. Jacobsen, "Random errors in sound intensity estimation," J. Sound Vib. **128**, 247–257 (1989).

[7] T. Loyau and J.-C. Pascal, "Statistical errors on the amplitude and direction of the complex acoustic intensity vector," J. Acoust. Soc. Am. **97**, 2942–2962 (1995).

[8] F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995).

[9] J. M. Jenkins and D. G. Watts, *Spectral Analysis* (Holden-Day, San Francisco, 1968).

[10] J.-C. Pascal and C. Carles, "Systematic measurement errors with two microphone sound intensity meters," J. Sound Vib. **83**, 53–65 (1982).

[11] J.-F. Li, J.-C. Pascal, and C. Carles, "Acoustic energy fields of partially coherent sources," J. Acoust. Soc. Am. **103**, 962–972 (1998).

Methods for determining infrasound phase velocity direction with an array of line sensors

Kristoffer T. Walker, Mark A. Zumberge, Michael A. H. Hedlin, and Peter M. Shearer
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, MC 0225, La Jolla, California 92093-0225

(Received 26 December 2007; revised 10 July 2008; accepted 10 July 2008)

Infrasound arrays typically consist of several microbarometers separated by distances that provide predictable signal time separations, forming the basis for processing techniques that estimate the phase velocity direction. The directional resolution depends on the noise level and is proportional to the number of these point sensors; additional sensors help attenuate noise and improve direction resolution. An alternative approach is to form an array of directional line sensors, each of which emulates a line of many microphones that instantaneously integrate pressure change. The instrument response is a function of the orientation of the line with respect to the signal wavefront. Real data recorded at the Piñon Flat Observatory in southern California and synthetic data show that this spectral property can be exploited with multiple line sensors to determine the phase velocity direction with a precision comparable to a larger aperture array of microbarometers. Three types of instrument-response-dependent beamforming and an array deconvolution technique are evaluated. The results imply that an array of five radial line sensors, with equal azimuthal separation and an aperture that depends on the frequency band of interest, provides directional resolution while requiring less space compared to an equally effective array of five microbarometers with rosette wind filters. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2968675]

PACS number(s): 43.60.Fg, 43.28.Dm, 43.58.Ta, 43.60.Qv [RR] Pages: 2090–2099

I. INTRODUCTION

A. Global infrasound monitoring

The acoustic frequency band beneath the human detection threshold (~18 Hz) is called infrasound. Infrasound is generated by a variety of natural and manmade sources and can travel thousands of kilometers under favorable atmospheric conditions. Nuclear explosions in the atmosphere produce far-reaching infrasound in the 0.01–10 Hz band (Landau and Lifshitz, 1959). This has recently led to increased interest in infrasound for compliance verification of the Comprehensive Nuclear Test-Ban Treaty (CTBT) [e.g., see Christie, (1999)]. The tool for this monitoring is the International Monitoring System (IMS), which is to include 60 globally distributed infrasound arrays.[1]

B. Wind noise reduction

Highly sensitive microbarometers and microphones with low-frequency response are available from a number of sources. The main difficulty in signal detection is not the sensitivity of the sensors themselves but the noise created by the wind. Increasing wind speeds correlate with increasing noise levels across the entire infrasound band. Specifically, the wind noise spectrum correlates highly with the wind velocity spectrum (Morgan and Raspet, 1992; Raspet *et al.*, 2006). One source of wind noise is the advection of spatially varying pressure or wind velocity anomalies that are frozen in time across the sensor at the local wind speed [Taylor's frozen turbulence hypothesis; see Taylor (1938)]. Because infrasound travels at much faster acoustic speeds (~330–350 m/s) than the wind and propagates across the Earth's surface at grazing elevation angles (nearly horizontal), large-aperture arrays of sensors and associated processing techniques are typically used to separate the rapidly moving signal from the more slowly moving turbulence. These techniques also provide an estimate of the azimuth toward the source and the speed with which the signal propagates across the array.

Wind noise occurs at all frequencies, and therefore the spatial separation of sensors provides only limited wind noise reduction. Spatial wind filters also help reduce wind noise. There are two types that are commonly employed: pipe networks and hose networks. Pipe networks consist of many inlet ports linked by pipes (often buried) to a central microbarometer or microphone. A common design is the "rosette" pipe filter, consisting of several clusters or rosettes of low-impedance pipe network inputs in which all inlet-microbarometer distances are the same (Alcoverro, 1998; Hedlin and Raspet, 2003; Alvocerro and Le Pichon, 2005). These rosettes vary in aperture depending on the frequency band of interest. Hedlin *et al.* (2003) and Hedlin and Alvocerro (2005) evaluated an array of rosettes. As the spatially coherent infrasound wavefront propagates across the inlets of these wind filters, the signals diffuse into the pipes and propagate toward the central sensor at the speed of sound. If wind noise originates from pressure anomalies that are advected across the array (Taylor's hypothesis) with a spatial coherence that is shorter than the separation distance between the inlets, the noise is reduced through averaging by the central sensor. This leads to a standard amplitude signal-to-noise gain of $n^{1/2}$, where n is the number of inlets. The rosette response depends on the apparent speed and frequency of the signal that propagates across the ports as well as on the rosette size. The disadvantages of rosette filters are

 0001-4966/2008/124(4)/2090/10/$23.00

that they are usually large and logistically challenging to build, the rosettes attenuate signal energy depending on the aperture (Alcoverro, 1998; Hedlin and Raspet, 2003), and resonance may occur in the infrasound band (Hedlin *et al.*, 2003). However, these filters do significantly attenuate wind noise and are employed by many of the IMS infrasound stations in arrays of up to nine elements and apertures of up to 2 km across (Hedlin and Alcoverro, 2005).

Another type of wind filter is a network of inexpensive porous hoses connected to a central microbarometer or microphone in a fashion similar to pipe arrays. Because the infrasound diffuses into the porous hose at all points of contact with the wavefront rather than a few discrete points, the signal-to-noise (SNR) ratio may be better than that for a rosette pipe array of the same aperture, depending on the spatial structure of the wind noise. These arrays are relatively simple to deploy, but their efficacy depends on their age and manufacturer (Shields *et al.*, 2007).

A more recent technique is to create a large array of many microphones and to use adaptive array processing techniques to attenuate incoherent wind noise while preserving a coherent signal (e.g., Shields, 2005). The success of this technique depends on the sensitivity of the microphones, the aperture of the array, the number and spacing of the sensors, and the processing technique employed. A disadvantage of using the standard beamforming technique with a large number of sensors is that it can require extensive computational and data-storage resources. Shields (2005) used an array of 28 sensors to investigate the optimum sensor separation distance. He showed that wind turbulence has a consistent spatial coherence for wind speeds in the 4–7 m/s range, allowing one to select the optimum sensors in the array to use in the beam for a known wind speed and direction to achieve better than $n^{1/2}$ noise reduction (where $n < 28$).

C. Motivation for this study

The optical fiber infrasound sensor (OFIS) is described by Zumberge *et al.* (2003) as a different approach to reducing wind noise. The OFIS is a compliant sealed tube that instantaneously measures pressure variation along its length by interferometrically sensing pressure-induced diameter change (Fig. 1). It spatially integrates pressure change and therefore attenuates spatially incoherent wind noise. Acoustic resonance inside the tube is not detected since standing waves are averaged to zero by the instantaneous integration along the length. These sensors can be oriented in any configuration (e.g., circle, rosette, or line). Zumberge *et al.* (2003) and Walker *et al.* (2007) showed that a single 90 m line OFIS, buried beneath 15 cm of coarse gravel, attenuates wind noise as well as a larger 70 m rosette pipe filter for winds of ~3.5 m/s. Laboratory and field measurements indicate that below 20 Hz, there is no attenuation of infrasound by the gravel (Zumberge *et al.*, 2003). For acoustic attenuation in a porous medium, the exponential decay with depth constant is proportional to the square root of the effective air flow resistance (Herrin *et al.*, 2001), suggesting that the effective flow resistance of the gravel is low.

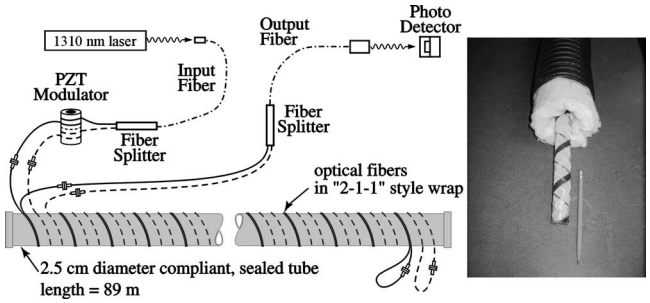


FIG. 1. Diagram and photo of an optical fiber infrasound sensor (OFIS; modified from Zumberge *et al.*, 2003). A laser illuminates two optical fibers that are helically wrapped around a sealed silicone tube. Differences in the coupling between the two optical fibers and the tube result in their responding differently to pressure-induced diameter changes, allowing an optical fiber interferometer to transform pressure changes into an optical signal. A photodetector converts the optical signal to an electrical one, which is analyzed in real time by a digital signal processor to generate a pressure recording (Zumberge *et al.*, 2004).

While the linear OFIS of length L is efficient in attenuating wind noise with less required real estate than a rosette of diameter L, its feasibility as an operational solution for infrasound data collection also depends on its ability to record infrasound reliably and to determine the direction from which the signal originated (back azimuth and elevation angle). In this paper we demonstrate three techniques for determining the phase velocity direction from an array of linear OFIS arms laid on the surface of the ground (unburied), which can generally occupy the same space as a single rosette filter. We also show how one can recover the unattenuated infrasound waveform from the OFIS recorded signals using an array deconvolution technique. The theory is presented along with applications to synthetic data and actual recordings, and conclusions are presented regarding useful OFIS array configurations.

II. SENSOR DIRECTIVITY

The velocity vector of the propagating wavefront in free space is the "phase velocity" c. The intersection of the wavefront with the Earth's surface advances at the "apparent velocity." The response of the instrument to an impulse, R, is the ratio of the recorded signal to the true pressure signal. A linear OFIS relative to a central reference point has the amplitude and phase response given by

$$R_a(f) = \mathrm{sinc}\left\{\frac{L\pi f}{c}\cos(\theta)\right\} \tag{1}$$

and

$$R_p(f) = \frac{L\pi f}{c}\cos(\theta), \tag{2}$$

where f is the frequency, L is the length of the OFIS, and θ is the angle between the phase velocity direction and the OFIS (Zumberge *et al.*, 2003, and adapted from Sheriff and Geldart (1995), p. 204). This angle is given by

$$\theta = \cos^{-1}(\cos(\theta_H)\cos(\theta_V)), \tag{3}$$

where θ_H are θ_V are the horizontal and vertical angles, respectively (Fig. 2). Equation (3) can be obtained by express-

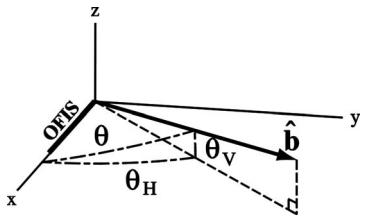


FIG. 2. Three-dimensional view of the relationship between the angles in Eq. (3). The OFIS is parallel to the x-axis, and the vector $\hat{\boldsymbol{b}}$ is antiparallel to the incident ray vector.

ing the length of the projection of $\hat{\boldsymbol{b}}$ onto the x-y plane in terms of θ_V and θ_H and by comparing this to $\hat{\boldsymbol{b}}\cdot\hat{\boldsymbol{i}}$, where $\hat{\boldsymbol{i}}$ is the unit vector in the x direction and $\hat{\boldsymbol{b}}$ is the unit vector antiparallel to the incident ray.

The phase response is just a time shift from the center of the OFIS to the end of the OFIS. For typical infrasound signals from distant sources, $\theta_V=0°-30°$ (grazing angles) and $\theta\cong\theta_H$. Figure 3(a) shows R_a as a function of θ for three typical infrasound signal frequencies. Figure 3(b) shows the variation in R_a with frequency for several sample angles. For low frequencies (or orthogonal arrival angles), the OFIS responds like a point sensor and has a flat instrument response. For all other conditions, a frequency-dependent attenuated version of the signal is recorded because the OFIS averages over some number of wave cycles. In other words, the fingerprint of the directionally dependent instrument response is recorded in the amplitude spectrum of the signal.

The OFIS impulse response can be verified numerically in the same way that the impulse response of a pipe rosette wind filter is calculated (Hedlin *et al.*, 2003). Specifically, for rays perpendicular to the OFIS, the impulse is recorded as a delta function (flat amplitude spectrum). For rays at some oblique angle to the OFIS, the impulse is averaged over time into a lower amplitude boxcar function and therefore a sinc function for the amplitude spectrum.

One might consider it a disadvantage that an OFIS spatially averages the signal. This averaging is only for wavelengths shorter than about $4L$ and only for certain directions. By comparison, a rosette filter of diameter L attenuates all signals having a wavelength shorter than about $4L$, regardless of the azimuth [Fig. 3(b); Hedlin *et al.*, 2003]. It is the

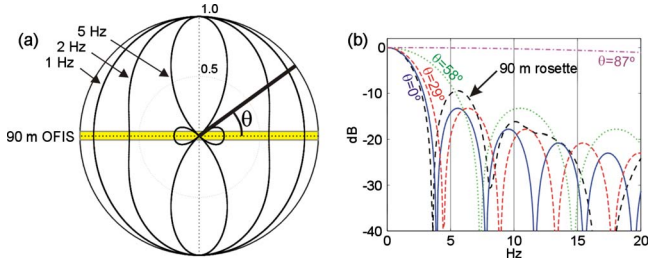


FIG. 3. (Color online) (a) Directivity and (b) frequency response R_a for a 90 m long OFIS as a function of θ and frequency [Eqs. (1) and (3)]. In (a) R_a is plotted in polar coordinates as a function of θ for three sample frequencies. In (b) R_a is plotted in dB as a function of frequency for four sample angles. For comparison purposes, the omnidirectional plane-wave response for a 90 m rosette with eight 16 m diameter secondary rosettes is also shown for grazing angles in (b).

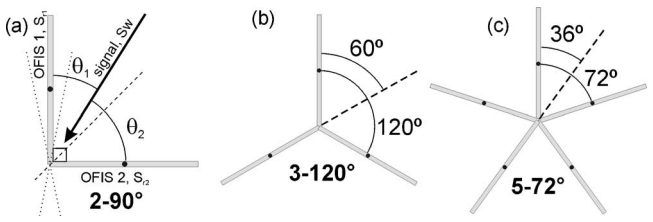


FIG. 4. OFIS array configuration naming terminology. The dotted and dashed lines in (a) are azimuths of ambiguity. The dotted azimuths can only be resolved with the separation of the OFIS centers. The dashed azimuths cannot be resolved with instrument response discrimination or separation of the OFIS centers. The angles to the dashed lines in (b) and (c) indicate the effective azimuthal separation in terms of phase-velocity-direction resolution.

directional dependence of attenuation inherent in a linear OFIS that provides additional information (besides time separation) about the direction from which a signal originated. This naturally leads to the question: Does a linear OFIS record as much directional information in the recorded amplitude spectrum as a collocated string of microphones via time separation?

In the frequency domain, each recorded OFIS signal S_r is a convolution of the signal waveform spectrum S_w with the instrument response $S_r=S_wR$. Conversely, if θ is known, one can deconvolve the instrument response to determine the signal spectrum $S_w=S_r/R$. Although the forward step is stable, the inverse step is not in this case because the denominator R has near zeros for certain frequencies (Fig. 3). To get around this problem, one can use water-level deconvolution (WLD) [e.g., see Langston 1979], which increases the amplitude of R in these troughs to, say, 2% of the maximum.

III. PHASE VELOCITY DETERMINATION TECHNIQUES

In this section, we describe the theory behind three different techniques used to derive the phase velocity direction. These techniques are a type of directional instrument-response-dependent beamforming or beam-steering. Figure 4 shows different OFIS configurations that are appropriate to use with these techniques if the length of the OFIS arms is longer than a quarter wavelength of the signal.

A. Water-level deconvolution: Predicted OFIS comparison technique

One usually wants to measure θ and the signal with the highest SNR possible. One approach is to have an array of several circular OFISs with an aperture optimized for beamforming in the frequency band of interest. However, one can exploit the directionality property of a linear OFIS by forming an array of OFIS arms in different orientations. For each possible ray orientation, an $R(\theta)$ exists that relates S_W to what should be recorded by each OFIS. Because the recorded OFIS signal is $S_r=f(S_w,\theta,L,c)$, where only S_w and θ are the unknowns, one can estimate S_w and θ if one records the signal on two OFIS with different orientations with $S_{r1}=f(S_w,\theta_1,L_1,c)$ and $S_{r2}=f(S_w,\theta_2,L_2,c)$, where θ_1 and θ_2 are related to the phase velocity direction by the array configuration [Fig. 4(a)]. One can do this by substitution, i.e., using S_{r1} to compute a predicted wave spectrum S_w^p using WLD,

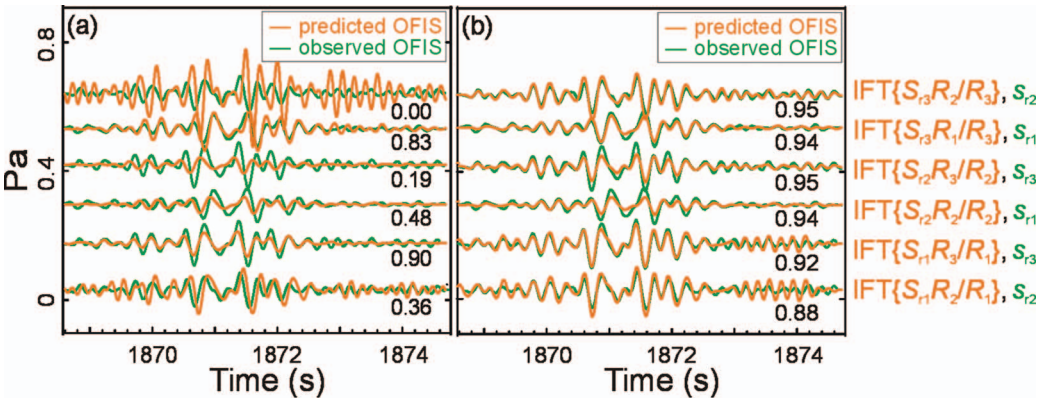


FIG. 5. Analysis of a 3°–120° OFIS signal recorded at PFO on 2005/09/20 (263) 02:31:09 UTC using the WLD technique. The predicted OFIS time series are compared with that observed for two trial back azimuths and elevation angles: (a) BAZ=316° and ELEV=44° and (b) the correct BAZ=286° and ELEV=44°. IFT is the inverse Fourier transform function. The numbers next to each pair are the correlation coefficients.

and then $S_{r2}{}^{p}=S_{w}{}^{p}R_2=S_{r1}R_2/R_1$, and performing a grid search over trial θ_t to minimize the sum of squares of the misfit between the maximum-normalized inverse Fourier transforms of S_{r2} and $S_{r2}{}^{p}$. The misfit is then given by

$$M(\theta_t)=\sum_{i=1}^{m}\left[\frac{s_{r2}{}^{p}}{\max(s_{r2}{}^{p})}-\frac{s_{r2}}{\max(s_{r2})}\right]^2, \tag{4}$$

where s_{r2} and $s_{r2}{}^{p}$ are the inverse Fourier transforms of S_{r2} and $S_{r2}{}^{p}$, respectively, and m is the number of points in the analyzed time series. The global minimum of the misfit function M corresponds to the best phase-velocity-direction estimate. We can calculate error bars by assuming that M is a measurement of a well behaved noise process [χ^2 distribution; see Jenkins and Watts (1968)] and determine the number of degrees of freedom based on the average recorded signal bandwidth (Silver and Chan, 1991), freeing us from the requirement that the observed signal samples be independent (white spectrum).

The precision using the WLD technique is mostly dependent on the SNR and the orientation of the ray with respect to the OFIS arms that are being used to estimate the waveform signal (deconvolution step). For example, in Fig. 4(a) the deconvolution of $R(\theta_1)$ from OFIS 1 (using OFIS 1 to predict OFIS 2) is more prone to noise amplification than the deconvolution of $R(\theta_2)$ from OFIS 2 (using OFIS 2 to predict OFIS 1) for the sample ray shown. There are three ways to compensate for this sensitivity. First, one can use each OFIS arm as the reference and sum both misfit functions throughout the grid search. Another approach is to skip grid points for which the trial ray is within some angle of the reference OFIS orientations (this requires more than two OFIS arms). Finally, one can use only the OFIS arm that is most perpendicular to the trial ray as the reference OFIS. Throughout this paper, we only use the first approach and sum both misfit functions for each reference OFIS. Therefore, for each grid point, $n(n-1)$ pairs of waveforms are compared, where n is the number of OFIS arms.

The 2–90° OFIS configuration has a special property [Fig. 4(a)]. The 90° separation offers identical spectral resolutions for phase velocity directions in all of the quadrants. In addition, it is computationally the fastest in part because one only needs to do the grid search within one of those quadrants to get the spectral resolution for all possible phase velocity directions. For each trial, we can therefore obtain the spectral constraints and then time shift the resulting inverse Fourier transformed waveform by $dt=(L_2\cos\theta_{H2}-L_1\cos\theta_{H1})\cos\theta_V/2c$ to calculate four separate quadrant misfit values (if we set the phase response $R_p=0$). The 2–90° configuration has been tested by analyzing recorded signals at the Piñon Flat Observatory (PFO) with the WLD method. The configuration generally only works for near-horizontally propagating signals with SNR > 12 dB. In addition, a fundamental limitation of any two-arm OFIS configuration is that it cannot be used to determine a phase velocity direction within the vertical plane defined by the azimuth between the two OFISs [Fig. 4(a)]. Only OFIS configurations with additional elements can resolve that ambiguity.

The WLD technique can be applied to OFIS arrays having more than two elements. There is, however, a doubling of the CPU time because quadrant symmetry in a 2–90° configuration is replaced by hemisphere symmetry in larger arrays (with OFIS arms that span an azimuthal range larger than a single quadrant). For example, a real signal recorded by a 3–120° OFIS array is shown in Fig. 5. Because there are three OFIS arms, there are six pairs of traces compared for similarity at each grid point. Using a trial phase velocity direction in Fig. 5(a) that is 30° from the true direction, the predicted OFIS recordings are quite different from the observed recordings in shape and timing. A good fit (average correlation coefficient >0.9) is only obtained when the trial phase velocity direction is correct [Fig. 5(b)].

B. Simultaneous array deconvolution technique

The WLD technique can amplify noise because of the deconvolution step. A more stable alternative is simultaneous array deconvolution (SAD). One estimates the original signal spectrum S_w by weighting, as a function of frequency and trial phase velocity direction, the recorded spectra of the OFIS arms. For each particular frequency, some OFIS arms are positioned to provide a more accurate spectral estimate of S_w than others. For example, consider four OFIS arms (1, 2, 3, and 4) oriented at trial azimuths of 0°, 29°, 58°, and 87°,

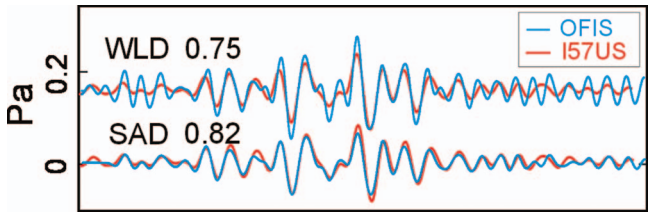


FIG. 6. Comparison of the WLD and SAD deconvolution techniques in removing the instrument responses from a 3°–120° OFIS. The resulting time series signal s_w from both techniques is compared (with correlation coefficients) to the beamformed signal from a microbarometer array.

respectively [i.e., the sample angles in Fig. 3(b)]. One should weight their estimates of the signal waveform based on the value of each response function at each frequency. In this example, the order from the highest weight to the lowest at 5 Hz would be OFIS 4, 3, 1, and 2 [comparing the responses in Fig. 3(b) at 5 Hz]. Such weighting makes the SAD technique more stable than the WLD approach because recorded spectra values around troughs in the relevant response functions are effectively eliminated from the signal estimation computation. This approach is a least-squares inversion. Assume that $\mathbf{D}$ is an $m \times n$ matrix containing the complex spectra computed from OFIS recordings, where m is the number of OFIS arms and n is the length of the spectra. The forward problem is represented in matrix notation as $\boldsymbol{GX}=\boldsymbol{D}$, where $\boldsymbol{X}=\boldsymbol{xI}$ contains S_w along the diagonal and $\boldsymbol{G}$ is an $m \times n$ matrix containing the m response functions of length n for the trial phase velocity direction. The least-squares solution for $\boldsymbol{X}$ is $\boldsymbol{X}=[\boldsymbol{G}^T\boldsymbol{G}]^{-1}\boldsymbol{G}^T\boldsymbol{D}$. Expanding for the ith frequency component, we have

$$x_i = \frac{G_{1i}D_{1i} + G_{2i}D_{2i} + \cdots + G_{mi}D_{mi}}{G_{1i}^2 + G_{2i}^2 + \cdots + G_{mi}^2}. \tag{5}$$

One can see from Eq. (5) that each x_i is a sum over the elements of $\boldsymbol{D}$ weighted by the elements of $\boldsymbol{G}$ (the response functions appropriate for that trial). This equation, which provides a least-squares estimate of S_w, is the most accurate technique when each OFIS has the same absolute noise level. From here, one can calculate the n OFIS predicted recordings in the same manner as discussed above (with the added computational load of a least-squares fit for each trial velocity direction) and derive a misfit function by comparing the n pairs of waveforms following Eq. (4). Alternatively, another measure of the misfit is the residual function in frequency space. For the trial phase velocity direction θ_t, the misfit function is given by

$$M(\theta_t) = \sum_{i=1}^{m}\sum_{j=1}^{n}(D_{ij} - G_{ij}x_j)^2. \tag{6}$$

The advantage of using the more CPU-intensive time-domain misfit of Eq. (4) is that one can normalize each waveform before taking the misfit, which allows for compensation of calibration problems and can sometimes improve stability.

A comparison of WLD and SAD for simply estimating the unattenuated signal waveforms is shown in Fig. 6. The instrument response was deconvolved from the recorded waveforms in Fig. 5. The resulting waveforms are compared with an estimate of the unattenuated signal provided by a collocated microbarometer array. WLD is performed on each OFIS recording, and their suite is averaged to give the graphed waveform. As evidenced by the correlation coefficient, the SAD waveform estimate matches the true signal more closely than does the WLD estimate, which has higher amplitudes for certain frequencies, presumably due to noise amplification.

C. Common spectrum identification technique

The common spectrum identification (CSI) technique is a search for the common signal spectrum recorded by all the OFIS arms. It does not involve deconvolution. Recall that the observed signals for two OFISs are $S_{r1}=S_wR_1$ and $S_{r2}=S_wR_2$. One can multiply both sides of these equations by the opposite instrument response function $S_{r1}R_2=S_wR_1R_2$ and $S_{r2}R_1=S_wR_2R_1$. Because of the commutative property of multiplication in the frequency domain, the right hand sides of these two equations are equal, and $S_{r1}R_2=S_{r2}R_1=Z$. We refer to Z as the common spectrum recorded by both OFIS arms. For n OFIS arms, one can perform the grid search within a hemisphere, where for each trial direction all OFIS recordings are convolved with their $n-1$ neighbor instrument responses and transformed back to the time domain (z) to evaluate the misfit function,

$$M(\theta_t) = \sum_{j=1}^{n}\sum_{i=1}^{m}\left[\frac{z_{ij}}{\max(z_j)} - \frac{\bar{z}_i}{\max(\bar{z})}\right]^2, \tag{7}$$

where $\bar{z}$ is the mean of the n estimates of z and m is the number of points in the analyzed time series. The CSI technique compares calculated waveforms that are more smoothed (because they result from convolution with multiple response functions) than any of the predicted individual OFIS recordings from the WLD or the SAD methods.

IV. RESOLUTION AND SIGNAL-TO-NOISE RATIO

Throughout this paper, SNR is referred to as excellent (SNR > 20 dB), good (SNR > 6 dB), and poor (SNR > 3.5 dB). The above OFIS array methods rely on two aspects of a recorded signal to determine its phase velocity direction: "time separation" and "instrument response discrimination." Time separation is the property that is exploited in classical array processing in either time or frequency space. Instrument response discrimination is the use of an OFIS to determine phase velocity direction based on the angular dependence of its frequency response, a characteristic of any sensor that averages along a line (Fig. 3). For example, consider a signal with a flat spectrum arriving at an OFIS with an incidence angle of 29° (one of the sample angles in Fig. 3). The spectrum of that signal recorded by that OFIS would resemble the cartoon in Fig. 7(a), where the 29° response curve from Fig. 3 leaves a fingerprint in the recorded spectrum unique to that arrival angle (or its 180° counterpart). In practice, the spectrum of the signal is unknown, which is why an array of sensors is needed to solve for both the arrival angle and the signal spectrum. Indeed,

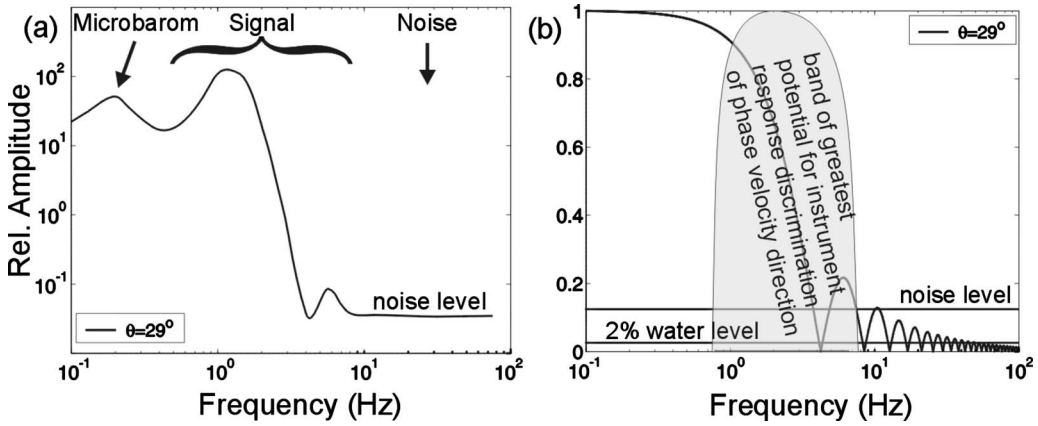


FIG. 7. Cartoon of the infrasound band that has the most leverage in determining the phase velocity direction by means of instrument response discrimination for a 90 m OFIS arm and a heuristic noise level.

the phase velocity direction of signals having spectral content mainly below $c/4L$ (around 1 Hz for a 90 m OFIS) cannot be determined with instrument response discrimination.

Noise affects the process of determining phase velocity direction in various ways for the different techniques discussed. Assuming a certain heuristic white noise level and recalling the amplitude decay of the response function at higher frequencies, one can see that certain regions of the spectrum have more leverage in determining the arrival angle for a particular SNR. As the noise floor increases, the higher signal frequencies recorded by OFIS arms at oblique angles become obscured. For this reason, it often helps to filter the recorded waveforms to isolate the band where there is the greatest potential for instrument response discrimination. Figure 7(b) depicts this region for a given SNR.

If the noise floor is above the water level used in the WLD technique, noise will be amplified, degrading the resolution by an amount dependent on the orientation and number of elements in the array. The SAD technique, on the other hand, can be more stable for higher noise floors because it assigns a weight to the *i*th frequency of all OFIS spectra based on the amplitude of the associated predicted instrument response function. This results in less contribution in the deconvolved signal from energy below noise floors that are significantly lower than the peak of the signal spectrum. However, if the absolute noise floor is much higher on one of several OFIS arms, this noise will be mapped into the deconvolved signal because SAD is a least-squares technique that can weight greatly the spectrum at certain frequencies of an uncharacteristically noisy OFIS. This noise can degrade the resolution since all OFIS comparisons are made using the noise-contaminated deconvolved signal spectrum.

The CSI method is the most stable for higher noise floors since there is no potential to amplify noise. However, because this technique heavily filters all of the OFIS recordings (including those from OFIS arms aligned parallel to the wavefront), lower frequencies (at wavelengths that are much longer than the OFIS length and separation distance) can dominate the spectrum and lead to poor phase-velocity-direction resolution. One should take care to isolate the signal energy with the appropriate high pass filter to minimize this effect.

V. APPLICATION TO REAL DATA

PFO in the southern California high desert is ensonified by many types of infrasound signals, including those originating from quarry mines, bombing ranges, sonic booms, rocket launches, and aircraft flying out of Los Angeles. During the summer of 2005, we recorded 81 infrasound signals with varying SNRs with a 3–120° OFIS array that was comprised of 90 m arms lying on the ground (unburied) at azimuths of 109.4°, 220.9°, and 330.7° (Fig. 8). The OFIS array was approximately collocated with an IMS infrasound array [I57US; see Hedlin *et al.* (2003)]. Comprised of eight elements fitted with rosette pipe filters, the IMS array data served as a reference with which to compare computed signal phase velocity directions [e.g., see Hedlin and Alcoverro (2005)]. Figure 8 shows a schematic of the high-frequency (HF) I57US subnetwork, located at the center of the I57US array with four 18 m diameter pipe rosette wind filters, each with 96 inlets spread over four 24-inlet clusters.

We analyzed all the signals recorded by the HF array and the OFIS array to empirically determine which of the OFIS methods described above yield phase velocity directions that are the most consistent with those estimated from

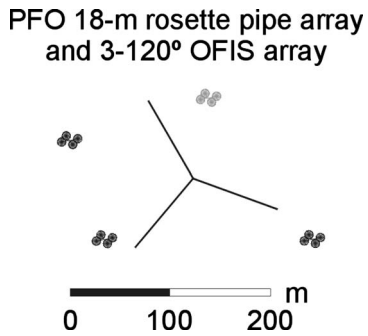


FIG. 8. PFO surface OFIS and partially buried pipe array (I57US, part of the IMS network) used in this study. The gray 18 m rosette was not used. The arrays are drawn to scale but are shown collocated for aperture comparison purposes.

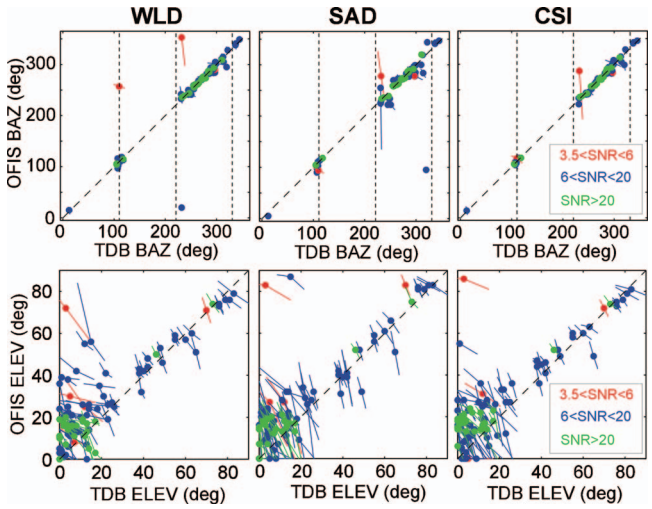


FIG. 9. Comparison of real signal back azimuths and elevation angles (recorded by a 3-120° OFIS and collocated pipe array I57US) using three OFIS methods and TDB for the pipe array. The back azimuth is the clockwise angle with respect to north. The elevation angle is the vertical angle with respect to the horizontal. SNR is the average SNR (dB) recorded by the OFIS. The 2σ error bars are the vector sums of the asymmetric error bars determined for each technique. Statistics that summarize these results are presented in Table I.

I57US array data (a known timing problem with one of the IMS data loggers limited our analysis to only three HF rosette elements). The low-frequency elements of I57US (four outer sensors with 70 m rosette filters) were also excluded from the analysis because they usually had poor SNR in the frequency band of interest.

We used a zero-phase fifth-order Butterworth bandpass filter to limit our analyses of the OFIS data to the 2–5 Hz band for the WLD and SAD techniques. The low-cut and high-cut frequencies correspond to amplitudes of −3 dB relative to the pass-band amplitude. For the CSI technique, we had the same low-cut frequency but chose a high-cut frequency of 10 Hz. This is permissible because the CSI technique does not involve deconvolution (unstable at the higher frequencies) and because it is a multifilter technique that naturally attenuates higher frequencies that are not common to all OFIS sensors. Because most of the OFIS techniques required a 2–5 Hz bandpass filter, we chose the same filter for the reference pipe array. We show results for all the signals with elevation angles of 0°–60° as determined using the pipe array. The sound speed used in the calculations was corrected for temperature.

In the following two sections, the OFIS back azimuths and elevation angles from the three techniques are compared against those obtained with time-domain beamforming (TDB) using the pipe array (Fig. 9). The TDB uncertainty is obtained using the same technique (mentioned above) as that for the OFIS methods and is therefore directly comparable. The error bars indicate the vector sum of the OFIS and TDB error bars and should intersect the unity-slope dashed line. The vertical dashed lines indicate the azimuths of the OFIS arm, where one might expect to find difficulty for some OFIS methods for low SNR. The means (μ) and standard deviations (σ) for the differences are presented in Table I. The back azimuth and elevation angle results assume that outliers are defined outside the ±25° and ±50° ranges, respectively.

TABLE I. Statistics for differences between the TDB and OFIS back azimuths and elevation angles from Fig. 9 (minus the outliers and low SNR measurements).

		Back azimuth			Elevation angle		
Method	SNR bin	μ	σ	n	μ	σ	n
WLD	All	0.4°	5.2°	79	−7.7°	11.4°	90
	SNR > 20	0.3°	2.4°	23	−6.1°	7.8°	24
	6 < SNR < 20	0.4°	6.0°	54	−8.2°	12.6°	62
SAD	All	1.7°	8.2°	77	−6.1°	8.1°	88
	SNR > 20	−0.1°	3.4°	23	−7.6°	6.1°	24
	6 < SNR < 20	1.8°	9.2°	51	−4.8°	8.6°	60
CSI	All	0.6°	4.9°	81	−6.6°	8.6°	89
	SNR > 20	0.2°	2.5°	23	−9.8°	6.7°	24
	6 < SNR < 20	0.7°	5.3°	55	−5.4°	9.0°	61

A. Back azimuth

None of the OFIS techniques shows a significant back azimuth bias. The range of standard deviations (spreads) for the WLD technique is 2°–6° depending on the SNR. Not using reference OFIS arms that are within 20° of the trial rays reduces the problems observed near the OFIS arm azimuths (because the deconvolution step is theoretically less unstable in the 2–5 Hz pass-band) but generally performs the same. Using only the most perpendicular OFIS in the deconvolution step, while theoretically better, also performs about the same. This can be explained by the reduction in the number of compared waveforms used in the misfit function (two pairs in this modified WLD versus six pairs in the unmodified WLD per trial ray).

The SAD technique has a larger spread of 3°–9°, but the outliers in WLD for azimuths close to the OFIS arms have been recovered. The SAD technique uses a more stable deconvolution method than the WLD technique. The SAD spread may be larger because one OFIS may have had an unusually high level of noise compared to the others, as discussed above.

The CSI technique has a spread of 3°–5°, comparable to the WLD spread. However, this technique eliminates all the outliers and yields measurements with associated error bars that appropriately span the unity-slope line (Fig. 9).

B. Elevation angle

There appears to be a bias of up to 10° between the OFIS estimates of elevation angle and those obtained using the pipe array. The bias, being generally less than one standard deviation, is barely significant statistically. The bias is also approximately the same for all SNRs, but it is better defined for the higher SNR signals because the error bars are smaller. It is difficult to discriminate among low elevation angles with sensors constrained to a horizontal plane; the addition of an OFIS oriented along a vertical line (on a tower, for example) may help, as would separating the OFIS arms by greater distances. It is possible that an underestima-

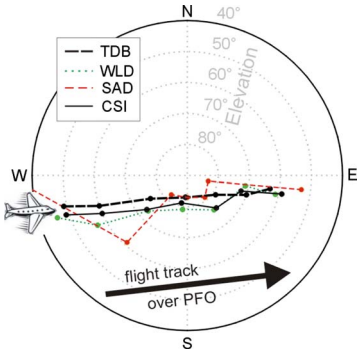


FIG. 10. (Color online) The analysis of infrasound created by an aircraft flying over PFO on 2005/09/18 (261) 19:52 UTC tracked by the OFIS array and the I57US high-frequency (HF) array 900 m to the southeast.

tion of the elevation angle by the three HF elements of the I57US array (Fig. 8) may also account for this discrepancy. Regardless, above an elevation of 30°, the differences are roughly symmetric about the unity-slope line. We performed numerical experiments that suggested that reasonable inaccuracies in the temperature-corrected sound velocity could not account for this bias. The OFIS array is located on an approximately flat surface about 900 m to the west-northwest of the I57US HF array, which is located on a slope dipping 5°–6° west. The CSI bias for easterly and westerly back azimuths is −11° and −7°, respectively. Although the slope explains the increase in the bias for easterly directions, the bias does not flip signs between the east and west as one would expect for a slope explanation.

The spread range of the elevation angles for the WLD technique is 8°–13° depending on the SNR. Above a 30° elevation, the results match better, with most of the error bars overlapping the unity-slope line. The SAD method has a spread range of 6°–9° and appears to be better at lower elevations than WLD. At higher elevations, SAD shows more scatter. The CSI technique has a comparable spread to the SAD method; the improvement seen at higher elevations mostly balances out with deterioration at lower elevations.

Many of the infrasound signals observed at PFO originate from airplanes flying overhead. These are interesting sources because the temporal variation in phase velocity direction provides a different type of comparison than that presented earlier. Figure 10 shows a polar plot of the phase velocity directions from one aircraft wavetrain broken into seven 6 s time windows. Both sensor systems tracked the plane across the sky. With respect to the other OFIS methods, the CSI method provided a track that was the most linear across the sky and the most similar to that from the pipe array. Most of the OFIS estimates are south of the estimates from the pipe array. This is likely because the arrays are not exactly collocated.

The structure of the misfit function can provide insight into the resolving power of each technique. The ideal technique results in a smooth misfit with a large dynamic range and a small confidence region. Figure 11 shows the misfit function variance (normalized sum of squares of the deviation between predicted and observed waveforms) for the three OFIS techniques compared with that provided by TDB for the signal in Figs. 5 and 6. Although the TDB estimate is well resolved, the local minima in the misfit could have been problematic had the SNR been lower. The WLD technique provides a smaller confidence region, but it too has local minima. Although the SAD method has a better dynamic range, it has prominent vertical ridges in the misfit and a vertical elongation of the confidence region, consistent with the larger elevation angle uncertainties seen in Fig. 9. The CSI misfit function is the smoothest, has the greatest dynamic range, and has the smallest confidence region.

VI. APPLICATION TO SYNTHETIC DATA

The above success in determining phase velocity direction confirms the validity of Eq. (1). Therefore, it is appropriate to create synthetic data and use the same equation to explore the power of other OFIS array configurations in determining the phase velocity direction. We created synthetic waveform data from a Vandenberg rocket launch signal recorded at I57US by an OFIS that was oriented perpendicular to the arriving infrasound wavetrain. We filtered the time series with a 2–5 Hz five-pole Butterworth bandpass filter [Fig. 12(a)]. Assuming a phase velocity direction, the filtered waveform was convolved with the appropriate response functions. White noise at a −12 dB level with respect to the original 2–5 Hz filtered signal (before convolution with the instrument response) was band limited using the same bandpass filter and was added to the result. Therefore, OFIS arms that are perpendicular to the input arrival direction have SNR = 12 dB, while OFIS arms that are at an oblique angle to the arrival direction have SNR < 12 dB.

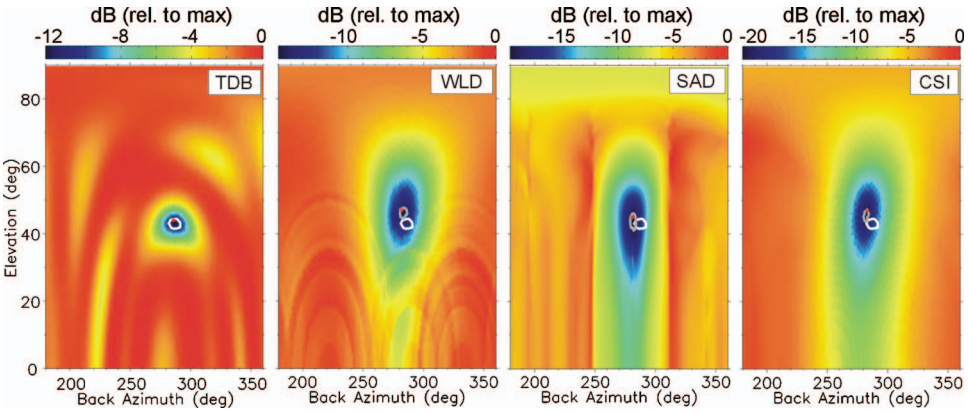


FIG. 11. Misfit functions for the three OFIS techniques and TDB (I57US HF array) used to analyze the PFO signal in Figs. 5 and 6. The 95% confidence regions of the reference TDB method (white contour) and individual OFIS methods (gray contour) are shown. The dot indicates the best estimate of the phase velocity direction (global minimum).

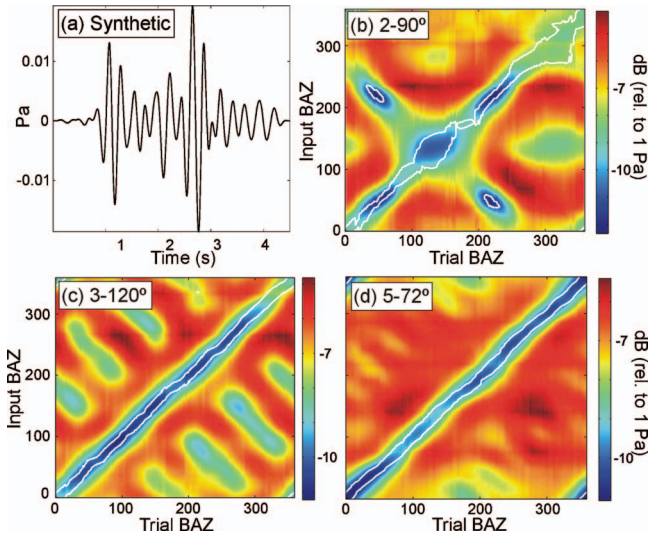


FIG. 12. Synthetic data experiments demonstrating the resolving power for phase velocity direction of different OFIS configurations using the WLD technique. The synthetic source signal is shown in (a). A misfit grid summarizing 360 different synthetic experiments (one experiment per degree input back azimuth) is shown for three different OFIS configurations in (b)–(d). White contours indicate the 95% confidence region, which ideally should enclose the unity-slope line that intersects the origin. The OFIS arms are 60 m in length.

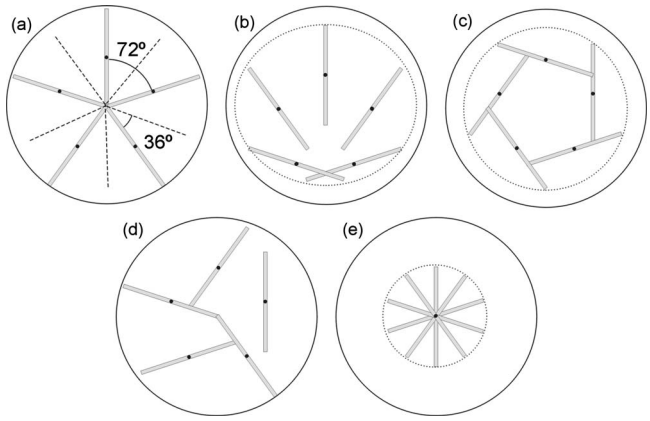


FIG. 13. Different configurations of OFIS arms of length L (drawn to scale). The circle that defines the maximum aperture of these configurations is drawn for each configuration. Configurations in (a)–(d) have the same theoretical resolving power for phase velocity direction (spectral and time separation resolution). Configuration (e) has no time separation resolution but is almost equally effective with a 75% footprint reduction. However, this configuration cannot distinguish between signals propagating from a given back azimuth and its 180° counterpart.

Comparisons up to this point have been between a 3–120° OFIS array and a three element rosette filter array. Figure 12 shows the results of analyzing horizontally propagating synthetic signals from the 0–360° back azimuth range for 2–90°, 3–120°, and 5–72° configurations. There were 360 synthetic tests (1° increment for the input back azimuth domain). For each test, the one-dimensional misfit is normalized by the 95% confidence contour. All 360 tests are plotted as a single image, the color scale of which indicates the variance of the misfit function in decibel relative to 1 Pa. For a perfect resolution, there would be a line of minima with a slope of unity intersecting the origin, reflecting that the input back azimuth for each synthetic test was recovered. In addition, the 95% confidence contours would collapse to a single line collocated with the line of minima. Finally, the variation of the misfit (or "roughness") would be insignificant compared to the depth of the minima. If the roughness were significant, then the additional off-axis local minima would tend to become possible solutions as the noise level increases.

The 2–90° configuration recovers most of the input back azimuths, but the precision varies with input back azimuth. In addition, one still observes local minima corresponding to poorly resolved ambiguities (Fig. 4). Spatial separation of the OFIS centers provides enough time separation resolution to keep these local minima above the 95% confidence contour. However, the ambiguity problem for the azimuth between the two OFIS arms is still apparent; the OFIS arms are at 0° and 90°, and the unresolved ambiguity angles are 45° and 225°.

The 3–120° configuration recovers most of the back azimuths with higher precision that does not vary with that of the input back azimuth. Although there are no unresolved ambiguities, the misfit has a high roughness.

The 5–72° configuration yields about the same precision as the 3–120° configuration. However, the misfit has a low roughness, suggesting that it would be more stable given higher noise levels. The 5–72° configuration is superior to the others because there are more OFIS arms recording the signal and because the effective azimuthal arm separation is smaller.

One can trade azimuths between OFIS arms and still preserve the original phase-velocity-direction resolving power. Although one generally wants to maximize the distance over which the coherent signal is recorded to attenuate incoherent wind noise [Fig. 13(a) radial geometry], this flexibility to trade azimuths can be attractive when logistics require the most compact footprint [36% less area than radial configuration, Fig. 13(c)] or flexibility in array installation [e.g., see Fig. 13(d)]. Since most of the resolving power of the 5–72° configuration is by means of spectral discrimination, one can collocate the OFIS centers [Fig. 13(e)] to make the smallest footprint (75% smaller than the radial configuration) as long as some method is provided to determine the hemisphere from which the signal originated.

VII. CONCLUSIONS

An array of n OFIS arms with length L inherits additional phase velocity resolution in a particular frequency band that an array of n point sensors separated by a distance L does not. One can exploit this directional fingerprint in the recorded spectra with instrument-response-dependent beamforming techniques. Each trial direction not only predicts time shifts to align the waveforms but also predicts waveform shape and amplitude changes. Real data recorded at PFO by a 3–120° OFIS array suggest that CSI outperforms other beamforming techniques that we have tested for signals with various SNRs. Of the two deconvolution techniques we tested to correct the recorded signals for the instrument response, SAD is usually better.

The PFO phase velocity directions, determined by CSI using a three-OFIS array (3–120°) with a 180 m aperture, compare favorably with those provided by three rosette elements of a roughly collocated, certified[2] IMS array with an aperture of ~250 m. This is significant because the three 90 m OFIS arms require less dedicated space than the three rosette filters.

The favorable results above are for a surface 3–120° array. An OFIS is more affected by noise on the surface than when it is buried. A single buried 90 m OFIS has a noise floor that is comparable to that provided by an entire 70 m rosette (Zumberge *et al.*, 2003; Walker *et al.*, 2007). These facts, our comparisons between different algorithms, and our synthetic tests predict that a buried 5–72° OFIS array will perform even better than a 3–120° OFIS array for infrasound measurement and characterization.

ACKNOWLEDGMENTS

We thank Petar Durdevic, Pat Walsh, Clint Coon, Eric Blum, Robin Matoza, and Joel White for their assistance with field and laboratory work. We also thank David Horwitt and David Chavez for providing data acquisition and real-time data transfer software and support. Jon Berger, Robin Matoza, and Scott DeWolf provided fruitful and inspiring discussions. We thank Mark Pickens, Bill Andre, and the U.S. Infrasound Team for their suggestions and feedback. This research was supported by the U.S. Army Space and Missile Defense Command. The High-Performance Wireless Research and Education Network (HPWREN) provided internet access to the OFIS array and microbarometer data recorded at PFO. This manuscript was improved by reviews from Richard Raspet and two anonymous reviewers.

[1]A description of the IMS is provided at http://www.seismo.ethz.ch/bsv/ctbto/ims.html (last viewed May 23, 2008).

[2]The certification of IMS arrays is discussed at http://www.ctbto.org/verification/overview.html (last viewed May 23, 2008).

Alcoverro, B. (**1998**), "Proposition d'un systeme de filtrage acoustique pour une station infrason IMS (Proposal of a filtering system for an acoustic IMS infrasound station)," Scientific Report No. 241, Commissariat a l'Energie Atomique, Departement Analyse Surveillance Environnement (Atomic Energy Commission, Department of Analysis, Surveillance, and Environment), Bruyéres-le-Châtel, France.

Alcoverro, B., and Le Pichon, A. (**2005**), "Design and optimization of a noise reduction system for infrasound measurements using elements with low acoustic impedance," J. Acoust. Soc. Am. **117**, 1717–1727.

Christie, D. (**1999**). "The infrasound segment of the CTBTO's international monitoring system," IUGG, XXII General Assmbly, B.7 (abstract).

Hedlin, M. A., Alcoverro, B., and D'Spain, G. (**2003**), "Evaluation of rosette infrasonic noise reducing spatial filters," J. Acoust. Soc. Am. **114**, 1807–1820.

Hedlin, M. A. H., and Alcoverro, B. (**2005**), "The use of impedance matching capillaries for reducing resonance in rosette infrasonic spatial filters," J. Acoust. Soc. Am. **117**, 1880–1888.

Hedlin, M. A. H., and Raspet, R. (**2003**), "Infrasonic wind noise reduction by barriers and spatial filters," J. Acoust. Soc. Am. **114**, 1379–1386.

Herrin, E., Sorrells, G. C., Negraru, P., Swanson, J. G., Golden, P., and Mulcahy, C. (**2001**), "Comparative evaluation of selected infrasound noise reduction methods," *Proceedings of the 23rd Seismic Research Review: Worldwide Monitoring of Nuclear Explosions*, 2–5 October, pp. 131–139.

Jenkins, G. M., and Watts, D. G. (**1968**), *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, **1968**).

Landau, L. D., and Lifshitz, E. M. (**1959**), *Fluid Mechanics* (Pergamon, Oxford).

Langston, C. (**1979**), "Structure under Mount Rainier, Washington, inferred from teleseismic body waves," J. Geophys. Res. **84**, 4749–4762.

Morgan, S., and Raspet, R. (**1992**), "Investigation of the mechanisms of low-frequency wind noise generation outdoors," J. Acoust. Soc. Am. **92**, 1180–1183.

Raspet, R., Webster, J., and Dillion, K. (**2006**), "Framework for wind noise studies," J. Acoust. Soc. Am. **119**, 834–843.

Sheriff, R. E., and Geldart, L. P. (**1995**), *Exploration Seismology* (Cambridge University Press, Cambridge, England).

Shields, F. D. (**2005**), "Low-frequency wind noise correlation in microphone arrays," J. Acoust. Soc. Am. **117**, 3489–3496.

Shields, F. D., Dillon, K., and Howard, W. (**2007**), "Acoustical properties of porous hose wind noise filters," in Proceedings of the Acoustical Society of America, New Orleans, LA, 28 November–1 December 2007.

Silver, P. G., and Chan, W. W. (**1991**), "Shear wave splitting and subcontinental mantle deformation," J. Geophys. Res. **96**, 16429–16454.

Taylor, G. I. (**1938**), "The spectrum of turbulence," Proc. R. Soc. London, Ser. A **164**, 476–490.

Walker, K. T., Zumberge, M., Hedlin, M., Berger, J., and Shearer, P. (**2007**), "Resolving infrasound signals with arrays of optical fiber infrasound sensors (OFIS): Wind noise reduction, instrument responses, and phase velocity direction resolution," in Proceedings of the Acoustical Society of America, New Orleans, LA, 28 November–1 December 2007.

Zumberge, M. A., Berger, J., Hedlin, M., Husmann, E., Nooner, S., Hilt, R., and Widmer-Schnidrig, R. (**2003**), "An optical fiber infrasound sensor: A new lower limit on atmospheric pressure noise between 1 Hz and 10 Hz," J. Acoust. Soc. Am. **113**, 2474–2479

Zumberge, M. A., Berger, J., Dzieciuch, M. A., and Parker, R. L. (**2004**), "Resolving quadrature fringes in real time," Appl. Opt. **43**, 771–775.

Performance analysis of the *Ormia ochracea's* coupled ears

Murat Akcakaya[a] and Arye Nehorai[b]
Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, Missouri 63130

(Received 14 January 2008; revised 30 June 2008; accepted 7 July 2008)

The *Ormia ochracea* is able to locate a cricket's mating call despite the small distance between its ears compared with the wavelength. This phenomenon has been explained by the mechanical coupling between the ears. In this paper, it is first shown that the coupling enhances the differences in times of arrival and frequency responses of the ears to the incoming source signals. Then, the accuracy of estimating directions of arrival (DOAs) by the *O. ochracea* is analyzed by computing the Cramér–Rao bound (CRB). The differential equations of the mechanical model are rewritten in state space and its frequency response is calculated. Using the spectral properties of the system, the CRB for multiple stochastic sources with unknown directions and spectra is asymptotically computed. Numerical examples compare the CRB for the coupled and the uncoupled cases, illustrating the effect of the coupling on reducing the errors in estimating the DOAs.
 [DOI: 10.1121/1.2967862]

PACS number(s): 43.60.Jn, 43.80.Lb [JAS] Pages: 2100–2105

I. INTRODUCTION

Most available array processing methods employ the time differences of arrival between the elements of a sensor array to estimate the directions of arrival (DOAs) of the incoming waves. Since the performance of such arrays is directly proportional to the size of the array's aperture, large aperture arrays are often required. However, this is costly and may be impractical in many tactical and mobile applications. This paper demonstrates a high-performance array with very small aperture, namely, of the parasitic fruit fly called the *Ormia ochracea*.

The *O. ochracea* is known to have a mechanical coupling between its ears to enhance its hearing. There are also other small animals having interactions between their ears for the same purpose,[1–5] but the mechanical coupling is unique for the *O. ochracea*. This coupling is necessary for the *O. ochracea's* perpetuation. The female *O. ochracea* must locate and deposit her parasitic maggots on or near a male field cricket, relying on the cricket's mating call, which is relatively pure in frequency (peak frequency 4.8 kHz). However, there is a tremendous incompatibility between the distance between the two ears (≈1.2 mm) and the wavelength (≈7 cm) of the cricket's mating call. This disaccord leads to extremely small interaural intensity and time of arrival differences between the ear closest to (ipsilateral) and the ear farthest from (contralateral) the sound source. It is believed that the coupling mechanism magnifies these binaural differences and subsequently improves the sound source localization accuracy.[3,6–10]

A mechanical system that models the mechanical coupling between the ears of the *O. ochracea* is introduced in Ref. 6. Miles *et al.*[6] showed experimentally that this model is well matched to the fly's ear in terms of frequency and transient responses.

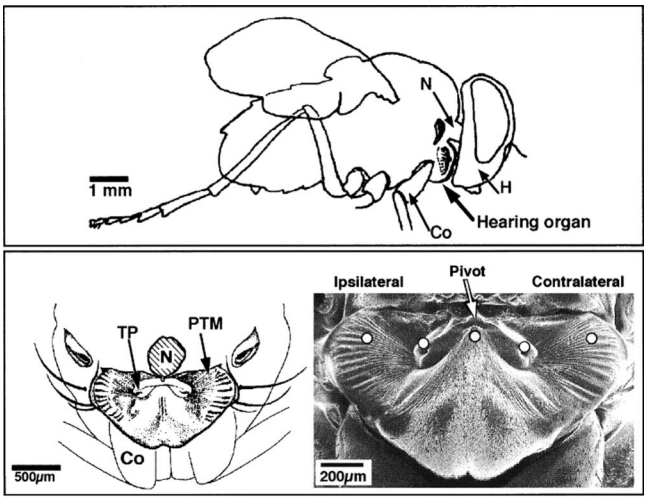
In this paper, it is analytically shown that the mechanical coupling between the *O. ochracea's* ears is significant for localization accuracy. First, the system in Ref. 6 is represented in state space. Then the model is solved and its time and spectral properties are illustrated. The impulse response of the system is found, and after sampling it, by taking its discrete-time Fourier transform (DTFT), the frequency response is computed. The frequency and impulse responses for coupled and uncoupled versions of the system are compared to demonstrate the effect of coupling on the intensity and time differences between the two ears. A statistical model with multiple stochastic sources and measurement noise is then developed to analyze the effect of the coupling on localization performance. The asymptotic Cramér–Rao bound (CRB) on estimating the source DOA is computed in two-dimensional space, i.e., for simplicity the estimation of only the azimuth angle is considered. It is assumed that the DOA and the power spectra of the input signals are unknown. Finally, numerical examples that compare the CRB's of the coupled and the uncoupled systems are presented, showing the improvement in the localization accuracy due to coupling.

II. MODELING

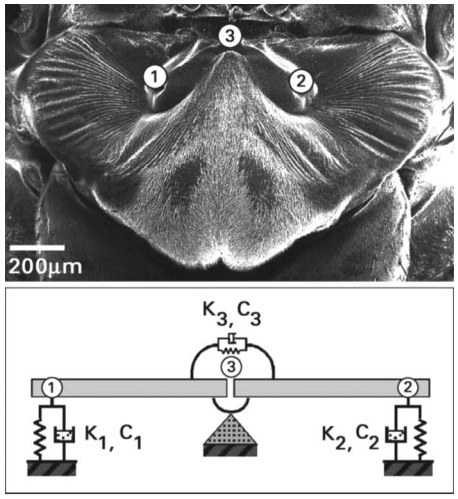
In this section, a brief review of the anatomy of the female *O. ochracea's* ears is provided. The mechanical model used in Ref. 6 is described, associating its parameters with the parts of the ear. Then, to demonstrate the effect of the coupling, the impulse and frequency responses of the coupled and the uncoupled systems for a far-field source are computed and compared with each other.

Figure 1(a) shows the female *O. ochracea* and its ear structure. We observe the following.

[a]Electronic mail: makcak2@ese.wustl.edu
[b]Electronic mail: nehorai@ese.wust.edu



(a)



(b)

FIG. 1. (a) Anatomy of the female *Ormia ochracea's* ear. Top: side view of the *y*. Bottom: front view of the ear after the head was removed. (b) Top: front view of the ear after the head was removed. Bottom: mechanical model (Ref. 6).

- The ear is located on the front face of the thorax, which is behind the head.
- Prosternal tympanal membranes serve for hearing.
- Bulbae acustica (sensory organs) are connected to the tympanal pit.
- The tympanal pits and the pivot point are connected to each other by a cuticular structure referred as intertympanal bridge. This improves the usage of interaural differences.

A simple mechanical model, composed of springs and dash-pots, is proposed in Ref. 6 to explain the mechanical coupling between the ears [Fig. 1(b)] with k_i's and c_i's ($i=1,2,3$) as the spring and dash-pot constants, respectively. In this model, the intertympanal bridge is assumed to consist of two rigid bars connected at the pivot point through a coupling spring k_3 and dash-pot c_3. The springs and dash-pots at the extreme ends of the bridge approximately represent the dynamic properties of the tympanal membranes, bulbae acustica, and surrounding structures. The numerical values of the above model were empirically found for a 45° incident angle, but they were shown to hold also for a wide range of angles.[6]

The differential equations for the mechanical model in Fig. 1(b) can easily be written in matrix form following Ref. 6:

$$\begin{bmatrix} k_1+k_3 & k_3 \\ k_3 & k_2+k_3 \end{bmatrix}\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} + \begin{bmatrix} c_1+c_3 & c_3 \\ c_3 & c_2+c_3 \end{bmatrix}\begin{bmatrix} \dot{z}_1(t) \\ \dot{z}_2(t) \end{bmatrix} + \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}\begin{bmatrix} \ddot{z}_1(t) \\ \ddot{z}_2(t) \end{bmatrix} = \begin{bmatrix} f_1(t,\Delta) \\ f_2(t,\Delta) \end{bmatrix} = \boldsymbol{f}(t,\Delta), \quad (1)$$

where

- $f_i(t,\Delta)=p_i(t,\Delta)s$, $i=1,2$, where $p_1(t,\Delta)$ and $p_2(t,\Delta)$ both correspond to the same input sound source and are the pressure waves at the ipsilateral and the contralateral ears, respectively, and s is the surface area of each tympanal membrane (see Fig. 1);
- $z_1(t)$ and $z_2(t)$ are displacements at the first and second ends of the intertympanal bridge, respectively [see Fig. 1(b)];
- m is the effective mass of all moving elements and it is assumed to be concentrated at each end of the intertympanal bridge; and
- Δ corresponds to the time difference of arrival between the two ears: $\Delta=d\sin\phi/v$, where $\phi\in[-90°,90°]$ is the DOA, d is the distance between force locations, and v is the speed of sound, which is roughly 344 m/s.

In order to find the solutions for $z_1(t)$ and $z_2(t)$, Eq. (1) is rewritten as a state space model

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{f}(t,\Delta),$$

$$\boldsymbol{y}(t) = C\boldsymbol{x}(t), \quad (2)$$

where $\boldsymbol{x}(t)=[x_1(t),x_2(t),x_3(t),x_4(t)]^T=[z_1(t),z_2(t),\dot{z}_1(t),\dot{z}_2(t)]^T$ is the state variable vector, A and B are constant matrices, which are functions of the model parameters in Eq. (1), and C is a constant matrix depending on the observations; see, for example, Ref. 11. C is chosen such that $\boldsymbol{y}(t)=[y_1(t),y_2(t)]^T=[z_1(t),z_2(t)]^T$. Using the variation of constants formula,[12] the solution for the state space model can be computed by

$$\boldsymbol{x}(t) = \Phi(t,t_0)\boldsymbol{x}(t_0) + \int_{t_0}^{t} \Phi(t,\tau)B\boldsymbol{f}(\tau,\Delta)d\tau,$$

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

Here, t_0 is the initial time referencing the instant when the input signal first arrives at the ipsilateral ear and $\Phi(t,t_0)$ is the transition matrix depending on the matrix A.

Figures 2–4 show the impulse, amplitude, and phase responses, respectively, for both the coupled and uncoupled systems. Figure 2 illustrates the impulse responses, $\boldsymbol{h}(t,\Delta)=[h_1(t,\Delta),h_2(t,\Delta)]^T$, calculated for $\phi=45°$ using Dirac delta function in Eq. (3) as an input. The responses $h_1(t,\Delta)$ and $h_2(t,\Delta)$ correspond to the ipsilateral and contralateral ears, respectively. It is apparent that the interaural differences be-

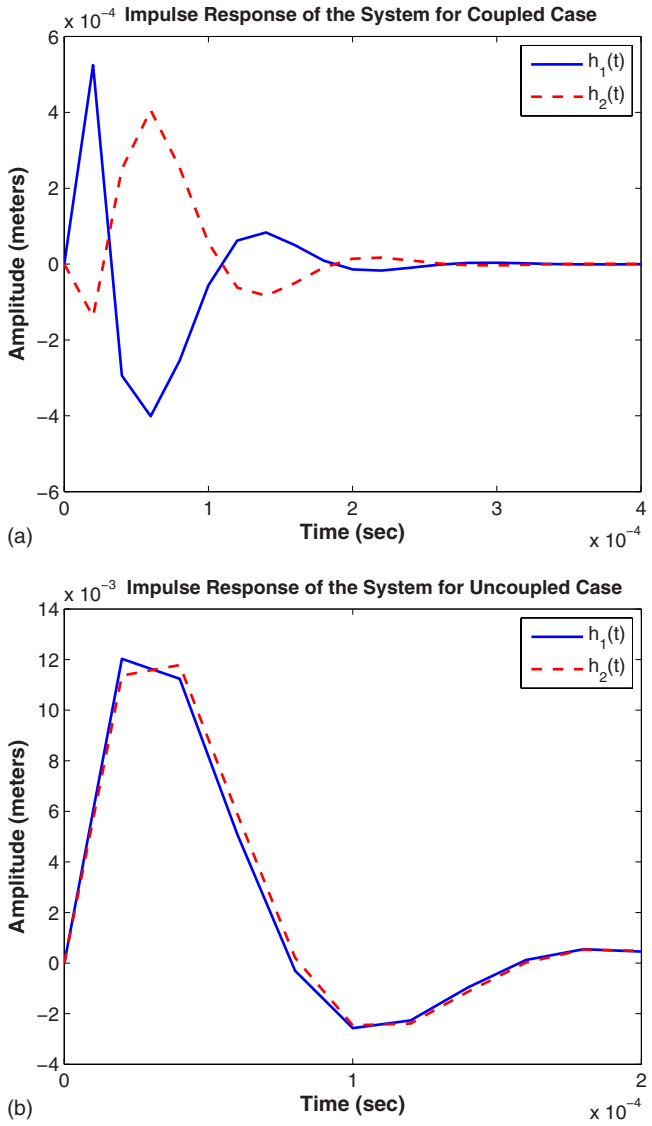


FIG. 2. (Color online) Impulse responses of the *Ormia ochracea's* two ears. (a) Coupled system. (b) Uncoupled system.

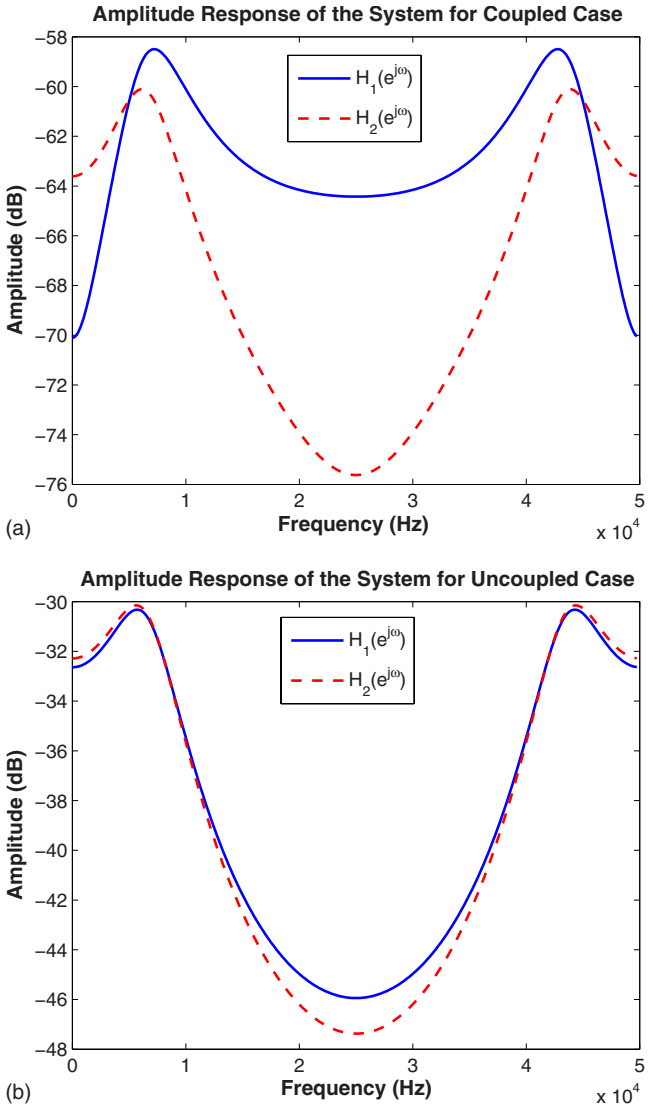


FIG. 3. (Color online) Amplitude responses of the *Ormia ochracea's* two ears. (a) Coupled system. (b) Uncoupled system.

tween the two ear outputs are enhanced for the coupled system [Figs. 2(a) and 2(b)]. These differences can be explained more clearly in the frequency domain. Therefore, the amplitude and phase responses are calculated by taking the DTFT (Ref. 13) of the sampled impulse responses. Figure 3 shows that the gap between the ipsilateral $(H_1(e^{j\omega},\Delta))$ and contralateral $(H_2(e^{j\omega},\Delta))$ amplitude responses is bigger for the coupled system [Figs. 3(a) and 3(b)]. This confirms the improvement of the intensity differences between the ear outputs. Similarly, Fig. 4 demonstrates how the phase difference between the responses of the two ears is amplified, so is the difference in the arrival time of the sound source to the two ears for coupled system [Figs. 4(a) and 4(b)]. This analysis may explain how the extremely small interaural differences in intensity and arrival time are increased by the coupling to a level that the *O. ochracea* could use the improved binaural cues to process the information more effectively.

III. PERFORMANCE ANALYSIS

In this section, a statistical model is presented and the CRB on DOA estimation of model (3) is computed.

A. Statistical model

The model consists of M multiple stochastic inputs $p_m(t)$ $(m=1,2,\ldots,M)$ and additive measurement noise $\boldsymbol{e}(t)$ $=[e_1(t),e_2(t)]^T$ with $e_1(t)$ and $e_2(t)$ corresponding to the measurement noise at the ipsilateral and contralateral ears, respectively (Fig. 5). That is, M different angles for M different input signals are chosen to model the environment. This model gives rise to

$$\boldsymbol{y}(t,\boldsymbol{\Delta}) = \sum_{m=1}^{M} p_m(t) \times h(t,\Delta_m) + \boldsymbol{e}(t), \quad t = 1,2,\ldots,N, \quad (4)$$

where $\boldsymbol{\Delta}=[\Delta_1,\Delta_2,\ldots,\Delta_M]^T$ and Δ_m is the time difference between two ears corresponding to the incidence angle ϕ_m of the input signal $p_m(t)$. The impulse response $h(t,\Delta_m)$ depends on the angle ϕ_m, due to the fact that for any signal $p_m(t)$ the equality $p_m(t-\Delta_m)*h(t)=p_m(t)*h(t-\Delta_m)$ always holds. Thus, it can be concluded that the system has different impulse responses and respective frequency responses for different incidence angles.

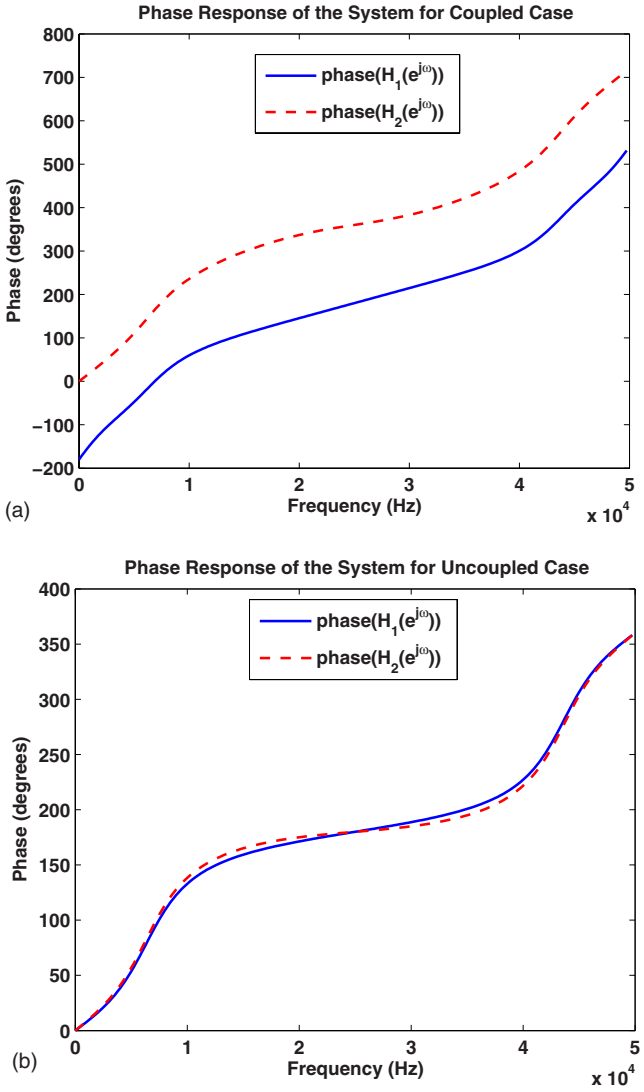


FIG. 4. (Color online) Phase responses of the *Ormia ochracea's* two ears. (a) Coupled system. (b) Uncoupled system.

It is assumed that $p_m(t)$ and $\boldsymbol{e}(t)=[e_1(t),e_2(t)]^T$ are zero-mean wide-sense stationary (WSS) Gaussian random processes; thus, $\boldsymbol{y}(t)$ is also a WSS Gaussian process since the system in Eq. (3) is linear and time invariant for zero initial state. These assumptions are used to asymptotically compute the CRB on the variance of the error of estimating the input signal incidence angle ϕ_m when there is an unbiased estimator $\hat{\phi}_m$ available.[14,15] However, for simplicity, the following further assumptions are made: $e_1(t)$ and $e_2(t)$ are white, have the same variance σ_e^2, and are uncorrelated with each other as well as with $p_m(t)$ $(m=1,2,\ldots,M)$. These assumptions result in

$$S_y(\omega,\boldsymbol{\Theta})=\sum_{m=1}^{M}\boldsymbol{H}(e^{j\omega},\Delta_m)S_{p_m}(\omega)\boldsymbol{H}^H(e^{j\omega},\Delta_m)+\sigma_e^2 I, \quad (5)$$

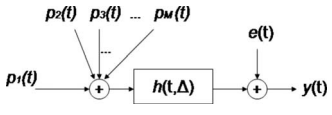


FIG. 5. Measurement model.

where

(i) $S_{p_m}(\omega)$, $\sigma_e^2 I$, and $S_y(\omega,\boldsymbol{\Theta})$ are the power spectral densities of $p_m(t)$, $\boldsymbol{e}(t)$, and $\boldsymbol{y}(t,\boldsymbol{\Delta})$, respectively;

(ii) $\boldsymbol{\Theta}=[\Delta_1,\Delta_2,\ldots,\Delta_M,S_{p_1},\ldots,S_{p_M},\sigma_e^2]^T=[\Theta_1,\ldots,\Theta_{2M+1}]$; and

(iii) $\boldsymbol{H}(e^{j\omega},\Delta_m)=[H_1(e^{j\omega},\Delta_m),H_2(e^{j\omega},\Delta_m)]^T$ $(m=1,2,\ldots,M)$ is the frequency response vector of the system related to the input signal $p_m(t)$ with incidence angle ϕ_m and "$(.)^H$" denotes the Hermitian transpose.

B. Cramér–Rao bound

Let $\boldsymbol{y}(t,\boldsymbol{\Delta})$ be defined as in Eq. (4) and satisfy the assumptions defined in Sec. III A. Then, the elements of the Fisher information matrix corresponding to unknown parameter vector $\boldsymbol{\Theta}=[\Delta_1,\Delta_2,\ldots,\Delta_M,S_{p_1},\ldots,S_{p_M},\sigma_e^2]^T$ can be found (for large N) as follows:[14,15]

$$[J(\boldsymbol{\Theta})]_{kl}=\frac{N}{4\pi}\int_{-\pi}^{+\pi}\mathrm{tr}\left\{\frac{\partial S_y(\omega,\boldsymbol{\Theta})}{\partial\Theta_k}S_y^{-1}(\omega,\boldsymbol{\Theta})\frac{\partial S_y(\omega,\boldsymbol{\Theta})}{\partial\Theta_l}\times S_y^{-1}(\omega,\boldsymbol{\Theta})\right\}d\omega,$$

$$[J(\boldsymbol{\Theta})]_{kl}\approx\frac{1}{2}\sum_{n=1}^{N}\mathrm{tr}\left\{\frac{\partial S_y(n,\boldsymbol{\Theta})}{\partial\Theta_k}S_y^{-1}(n,\boldsymbol{\Theta})\frac{\partial S_y(n,\boldsymbol{\Theta})}{\partial\Theta_l}S_y^{-1}(n,\boldsymbol{\Theta})\right\}, \quad (6)$$

where $S_y(n,\boldsymbol{\Theta})$ is the discrete Fourier transform (DFT) of the system output with frequency index n, which is obtained by sampling $S_y(\omega,\boldsymbol{\Theta})$ in the frequency domain.[13]

Recall that $\Delta_m=d\sin(\phi_m)/v$. Hence, to find the CRB of estimating the DOA, ϕ_m, the transformation formula[15] is utilized as follows:

$$\mathrm{Var}(\hat{\phi}_m)\geq\left[\left(\frac{\partial g(\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}\right)J^{-1}(\boldsymbol{\Theta})\left(\frac{\partial g(\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}\right)^T\right]_{mm}, \quad (7)$$

where

(i) $\boldsymbol{g}(\boldsymbol{\Theta})=[g_1(\boldsymbol{\Theta}),\ldots,g_M(\boldsymbol{\Theta})^T]$, with the mth element defined as $g_m(\boldsymbol{\Theta})=\arcsin(v\Delta_m/d)$, and

(ii) $J(\boldsymbol{\Theta})$ is the Fisher information matrix with $[J(\boldsymbol{\Theta})]_{kl}$ as defined in Eq. (6) $(1\leq k,\ l\leq 2M+1)$.

IV. NUMERICAL RESULTS

The CRBs of estimating the DOAs for the coupled and uncoupled systems are compared to show the effect of the coupling. The following scenario for different signal to noise ratios (SNRs) is used.

It is assumed that in Eq. (4), $p_1(t)$ with incidence angle ϕ_1 is the incoming signal that is to be localized and similarly $p_m(t)$ and $\phi_m(t)$ $(m=2,3,\ldots,M)$ are the signal from the environment and corresponding incidence angle, respectively. For simulation purposes only, ϕ_m is randomly chosen (ϕ_m is uniform in $[-90°,90°]$). Accordingly, SNR is defined as

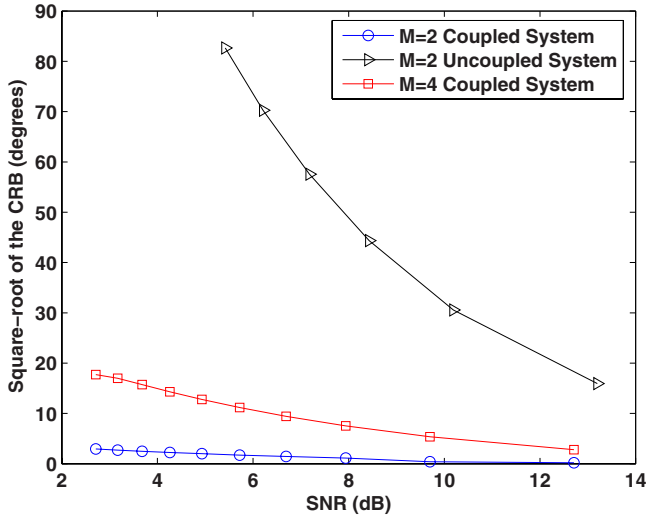


FIG. 6. (Color online) Square root of the CRB on the DOA estimation vs SNR.

$$\text{SNR} = \frac{\mathrm{Tr}\{\Sigma_{n=1}^{N}\boldsymbol{H}(n,\Delta_1)S_{p_1}(n)\boldsymbol{H}^H(n,\Delta_1)\}}{\mathrm{Tr}\{\Sigma_{n=1}^{N}(\Sigma_{m=2}^{M}\boldsymbol{H}(n,\Delta_m)S_{p_m}(n)\boldsymbol{H}^H(n,\Delta_m)) + S_e(n)\}}, \tag{8}$$

that is, $p_m(t)$ for $m=2,3,\ldots,M$ contribute to the noise power. Note also that similar to Eq. (6) $\boldsymbol{H}(n,\Delta_m)=[H_1(n,\Delta_m),H_2(n,\Delta_m)]$ and $S_{p_m}(n)$ are the DFT of $\boldsymbol{H}(e^{j\omega},\Delta_m)$ and $S_{p_m}(\omega)$ and n denotes the frequency index.

In addition to the assumptions in Sec. III A, for further simplification, it is assumed that the spectrum corresponding to $p_m(t)$ is constant over the entire frequency range, i.e., $S_{p_m}(w)=\sigma_{p_m}$ $(\omega\in[-\pi,\pi])$. This simplification is reasonable in this problem since the amplitude response of the mechanical model is already selective (see Fig. 3).

- Model parameters[6] are chosen as follows:
  (1) $c_1=c_2=1.15\times10^{-5}$ Ns/m and $k_1=k_2=0.576$ N/m,
  (2) $c_3=3.88\times10^{-5}$ N s/m and $k_3=5.18$ N/m, and
  (3) $m=2.88\times10^{-10}$ kg.
- Note also that in these examples
  (1) $\phi_1=45°$,
  (2) $t_0=0$ [See Eq. (3)], and
  (3) $T_s=0.02$ ms is the sampling period.

In Fig. 6, the improvement on the lower bound of the error in estimating the DOA when there is coupling in the system is shown by plotting the square root of the CRB with respect to the SNR. This figure first illustrates the case where the number of the time samples is 2000 $(N=2000)$, which correspond to a time range of 40 ms, and the number of input signals is 2 $(M=2)$, i.e., there is only one interference signal other than the input itself. Note that the number of the time samples is chosen to obtain CRB values of roughly 1°–2° for the coupled system to match the experimental results in Ref. 10. When the CRBs of the coupled and the uncoupled system when $M=2$ are compared, a big difference is observed for the same SNR. This clear difference shows how the mechanical coupling improves the localization ability of the *O. ochracea*. For the coupled system the error in the azimuth angle drops down to 1°–2°, which is also in agreement with the experimental results demonstrated in Ref. 10. Note that for the uncoupled system when $M=2$, the variance of the error is markedly above the experimental results for SNR less than 12–14 dB.

Figure 6 also demonstrates the performance of only the coupled system with $N=5000$ and $M=4$. For the chosen values of M (the number of the input signals) and N (number of the time samples), the CRB values for the uncoupled system are too high, so the CRB does not hold anymore. $N=5000$ for $M=4$ are chosen in the same way described for $M=2$ to reach the desired CRB values. It is observed that as the number of interference signals increases, so does the number of time samples for the same CRB values, when the plots of the coupled system for $M=2$ and $M=4$ are compared. For the fly this may correspond to requiring more time to locate the source when background noise increases. It can also be seen from the same figure that to achieve the same CRB values as when $M=2$, higher SNR values are needed for when $M=4$.

V. CONCLUSION

The effect of coupling on the hearing system of the *O. ochracea* was analyzed. By examining the impulse and frequency responses of the model, it was shown that the coupling increased the time and amplitude differences between the two ears. Using our statistical model, the CRB of estimating the DOA for unknown input signal spectra was asymptotically computed. Comparing the CRB of the coupled with the uncoupled system, the enhancement provided by the mechanical coupling was illustrated. The experimental results regarding the azimuthal localization capacity of the *O. ochracea* were analytically proven. Future work will analyze the system and calculate the CRB in continuous time.

ACKNOWLEDGMENT

The authors are thankful to Dr. Chong Meng See of DSO, Singapore, for introducing them to the subject of the *Ormia ochracea's* hearing system. This work was supported by the DARPA Grant No. HR0011-07-1-0036.

[1]N. H. Fletcher and S. Thwaites, "Acoustical analysis of the auditory system of the cricket *Teleogryllus commodus* (walker)," J. Acoust. Soc. Am. **66**, 350–357 (1979).
[2]A. Michelsen, A. V. Popov, and B. Lewis, "Physics of directional hearing in the cricket *Gryllus bimaculatus*," J. Comp. Physiol., A **175**, 153–164 (1994).
[3]D. Robert, R. N. Miles, and R. R. Hoy, "Tympanal mechanics in the parasitoid fly *Ormia ochracea*: Intertympanal coupling during mechanical vibration," J. Comp. Physiol., A **183**, 443–452 (1998).
[4]A. R. Palmer and A. Pinder, "The directionality of the frog ear described by a mechanical model," J. Theor. Biol. **110**, 205–215 (1984).
[5]A. Michelsen, "Hearing and sound communication in small animals: Evolutionary adaptations to the laws of physics," *The Evolutionary Biology of Hearing* (Springer-Verlag, New York, 1992).
[6]R. N. Miles, D. Robert, and R. R. Hoy, "Mechanically coupled ears for directional hearing in the parasitoid fly *Ormia ochracea*," J. Acoust. Soc. Am. **98**, 3059–3070 (1995).
[7]D. Robert, R. N. Miles, and R. R. Hoy, "Directional hearing by mechanical coupling in the parasitoid fly *Ormia ochracea*," J. Comp. Physiol., A

179, 29–44 (1996).
[8]D. Robert, M. P. Read, and R. R. Hoy, "The tympanal hearing organ of the parasitoid fly *Ormia ochracea* (diptera, tachinidae, ormiini)," Cell Tissue Res. **275**, 63–78 (1994).
[9]D. Robert, M. J. Amoroso, and R. R. Hoy, "The evolutionary convergence of hearing in a parasitoid fly and its cricket host," Science **258**, 1135–1137 (1992).
[10]A. C. Mason, M. L. Oshinsky, and R. R. Hoy, "Hyperacute directional hearing in a microscale auditory system," Nature (London) **410**, 686–690 (2001).
[11]K. Ogata, *Modern Control Engineering*, 4th ed. (Prentice-Hall PTR, Upper Saddle River, NJ, 2001).
[12]W. E. Boyce and R. C. Diprima, *Elementary Differential Equations and Boundary Value Problems*, 7th ed. (Wiley, New York, 2000).
[13]A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. (Prentice-Hall PTR, Upper Saddle River, NJ, 1999).
[14]P. Whittle, "The analysis of multiple stationary time series," J. R. Stat. Soc. Ser. B (Methodol.) **15**, 125–139 (1953).
[15]S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall PTR, Upper Saddle River, NJ, 1993).

Effects of maturation on tympanometric wideband acoustic transfer functions in human infants[a)]

Chris A. Sanford[b)]
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98105

M. Patrick Feeney
Otolaryngology, Head and Neck Surgery, Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, Washington 98195

(Received 3 January 2008; revised 30 June 2008; accepted 8 July 2008)

Wideband acoustic transfer function (ATF) measurements of energy reflectance (ER) and admittance magnitude ($|Y|$) were obtained at varying static ear-canal pressures in 4-, 12-, and 27-week-old infants and young adults. Developmental changes in wideband ATF measurements varied as a function of frequency. For frequencies from 0.25 to 0.75 kHz there was as much as a 30% change in mean ER and $|Y|$ with changes in static ear-canal pressure between 4 and 24 weeks of age. From 0.75 to 2 kHz, the effects of pressure produced a small number of significant differences in ER and $|Y|$ with age, suggestive of a developmentally stable frequency range. Between 2 and 6 kHz, there were differential effects of pressure for the youngest infants; negative pressures caused increased ER and $|Y|$ and positive pressures caused decreased ER and $|Y|$; the magnitude of this effect decreased with age. Findings from this study demonstrate developmental differences in wideband tympanometric ATF measurements in 4-, 12- and 24-week-old infants and provide additional insight on the effects of static ear-canal pressure in the young infant's ear. The maturational effects shown in the experimental data are discussed in light of known age-related anatomical changes in the developing outer and middle ear.
 [DOI: 10.1121/1.2967864]

PACS number(s): 43.64.Ha, 43.58.Bh [BLM] Pages: 2106–2122

I. INTRODUCTION

Acoustic ear-canal measurements are dependent on the physical properties of the external and middle ear, properties that undergo developmental changes in human infants over the first several months of life. Some of these changes include (1) an overall increase in the size of the ear canal and middle-ear space, (2) a decrease in the length of the cartilaginous portion of the ear-canal wall due to growth of the bony portion of the canal wall, (3) changes in ossicular bone density due to loss of residual mesenchyme and ossification, and (4) changes in the orientation of the tympanic membrane and ossicular chain (Ikui *et al.*, 1997; Ruah *et al.*, 1991; Saunders *et al.*, 1983). These anatomical changes in infant ears are thought to contribute to conventional single-frequency tympanometry patterns that are not adultlike over the first 6 months of life (Holte *et al.*, 1991; Hunter and Margolis, 1992; Paradise *et al.*, 1976). The general conclusion from these developmental studies is that the overall impedance characteristics of infants' ears at low frequencies are dominated more by the effects of mass, in contrast to the middle-ear system in children and adults which is dominated by the effects of stiffness (Holte *et al.*, 1991; Meyer *et al.*, 1997; Roush *et al.*, 1995). However, the extent to which specific developmental changes in the infant external and middle ears contribute to specific variations in acoustic middle-ear measurements is not fully understood. A multifrequency approach to middle-ear assessment would be ideal for infants in the neonatal to 7 month age range in light of the shifting maturational gradient of acoustic ear-canal-response properties (Feeney and Sanford, 2005; Holte *et al.*, 1991; Keefe *et al.*, 1993; Meyer *et al.*, 1997).

From both research and clinical perspectives, it is important to distinguish between variations in middle-ear measurements attributable to developmental effects and those attributable to dysfunction or pathology. Furthermore, understanding how developmental changes influence acoustic response properties of the infant ear have important implications for interpretation of middle-ear assessment measures as well as behavioral and physiologic measures of infant hearing (Holte *et al.*, 1991; Keefe *et al.*, 1993; Roush *et al.*, 1995; Holmer, 2001[1]). While more recent studies report normative data for 1 kHz probe-tone tympanometry for infants younger than 7 months of age (Alaerts *et al.*, 2007; Baldwin, 2006; Calandruccio *et al.*, 2006; Kei *et al.*, 2003; Margolis *et al.*, 2003), a complete definition of what constitutes a normal or disordered infant middle ear is currently lacking.

Wideband acoustic transfer function (ATF) measurements, such as energy reflectance (ER) and acoustic admittance, are emerging as attractive alternatives to traditional

[a)]Portions of this work were presented at the American Auditory Society, Scientific and Technology Meeting, Scottsdale, AZ, March 2006 and the 30th Meeting of the Association for Research in Otolaryngology, Denver, CO, February 2007.

[b)]Present address: Boys Town National Research Hospital, Omaha, NE; electronic mail: sanfordc@boystown.org

admittance tympanometric measurements (Feeney *et al.*, 2003; Keefe *et al.*, 2000; Keefe *et al.*, 1993; Keefe *et al.*, 2003; Piskorski *et al.*, 1999; Voss and Allen, 1994). ER is the ratio of sound power reflected from the middle ear to the incident sound power presented by a probe stimulus in the ear canal. ER at ambient ear-canal pressure has been used to examine normal and disordered middle-ear function (Allen *et al.*, 2005; Feeney *et al.*, 2003), the acoustic stapedial reflex (Feeney and Keefe, 1999; Feeney and Keefe, 2001; Feeney *et al.*, 2004), neonatal middle-ear function (Keefe *et al.*, 1993; Keefe *et al.*, 2000), and middle-ear development (Keefe *et al.*, 1993; Keefe and Levi, 1996; Keefe and Abdala, 2007). Specifically, Keefe *et al.* (1993) investigated the effects of middle-ear maturation on wideband ATF measurements in adults and in five groups of infants and children (1, 3, 6, 12, and 24 months of age). Age-related variations in ER over a frequency range of 0.125–10.7 kHz were reported by Keefe *et al.* (1993), with ER patterns changing most significantly between 1 and 6 months of age. The greatest changes in ER occurred at frequencies below 0.5 kHz where ER increased by up to 30% between 1 and 6 months of age. Keefe *et al.* (1993) concluded that significant resonant amplification of ear-canal wall motion occurred for lower frequencies in young infants and that these effects may account for variability encountered with 226 Hz admittance tympanograms in infants.

Additional studies have expanded the use of wideband ambient pressure ATF measurements to include measurements made over a range of static ear-canal pressures (ATF tympanometry) (Keefe and Levi, 1996; Keefe and Simmons, 2003; Margolis *et al.*, 1999). Margolis *et al.* (1999) obtained conventional multi-frequency-admittance tympanometry and wideband ATF tympanometry results from 20 adults. At frequencies above 2 kHz, multifrequency admittance tympanograms did not follow an orderly pattern, unlike ER tympanograms, which progressed in an orderly fashion as a function of frequency from 0.25 through 11.3 kHz. The authors suggested that this orderly behavior might prove useful for the assessment of pathologic ears. Of particular clinical interest were data presented by Margolis *et al.* (1999) from one 10-year-old boy with recurrent otitis media. The boy presented with a unilateral 35 dB conductive hearing loss, negative tympanometric peak pressure (TPP) (−250 daPa), and normal 226 Hz static acoustic admittance (SAA). Otoscopy revealed a retracted tympanic membrane, but no evidence of middle-ear effusion. ATF measures at ambient pressure resembled ER typically observed in normal ears under pressurized conditions. However, when the ear canal was pressurized to compensate for the negative middle-ear pressure, an abnormal ER pattern was obtained, indicating the possibility of an additional middle-ear disorder, concurrent with negative middle-ear pressure. Margolis *et al.* (1999) concluded that due to its apparent sensitivity to middle-ear disorders, ER tympanometry could be a useful tool for clinical middle-ear assessment.

Keefe and Simmons (2003) investigated the test performance of 226 Hz tympanometry, energy transmittance (which they defined as 1 minus ER) at ambient ear-canal pressure, and energy transmittance at static ear-canal pressures as predictors of conductive hearing loss in children 10 years of age and older. Data were obtained from 42 normal-functioning ears and 18 ears with conductive hearing loss. For a fixed specificity of 0.90, the sensitivity of these measures to detect conductive hearing loss was 0.28 for SAA (226 Hz), 0.72 for ambient pressure transmittance, and 0.94 for pressurized transmittance (ATF tympanometry), suggesting that wideband ATF tympanometry may be a useful diagnostic tool.

Based on these results which demonstrate high sensitivity of wideband ATF measurements to changes in middle-ear characteristics due to middle-ear pathology, it is hypothesized that maturational effects of the infant ear canal and middle ear may be better characterized using wideband ATF measurements. Investigation of these responses in young infants is of interest in light of the need for better understanding of infant middle-ear development and for a valid test of middle-ear function for young infants. However, to date, no published studies have described young infant middle-ear characteristics using wideband ATF responses under pressurized ear-canal conditions. The current study was designed to further investigate developmental changes in acoustically based middle-ear measurements in infants and to characterize wideband ATF tympanometry results in infants of 4, 12 and 24 weeks of age.

II. METHOD

A. Subjects

One hundred and one infants were recruited for the present study. Cross-sectional data are reported for 60 infants (one ear from each infant), with 20 infants in each age group (4, 12, and, 24 weeks). Mean age and standard deviation (SD) and gender distribution (F=female, M=male) for each age group are as follows: 4 weeks: 4.5 (SD=0.34) weeks, 8F and 12M; 12 weeks: 12.0 (SD=0.34) weeks, 6F and 14M; and 24 weeks: 24.1 (SD=0.31) weeks, 14F and 6M. Infant participant selection criteria included (1) full term birth (40 week gestation, ±2 weeks), (2) birth history free of complications, (3) good health on the test date, (4) no risk factors for hearing loss, (5) no history of otitis media, and (6) clear entrance to the ear canals assessed by visual examination. An approximately equal number of left (L) and right (R) ears were included for each age group (4 weeks: 9 R, 11 L; 12 weeks, 9 R, 11 L, and 24 weeks: 12 R and 8 L. Assignment of right and left ears for data analysis was counterbalanced, alternating with each subject. Because data were not obtained from all ears of every infant, proportions of left and right ears were not equal. Infants were recruited with the assistance of a database maintained at the University of Washington and funded through a grant from the National Institutes of Health.

Screening of infants for inclusion in the study consisted of distortion product otoacoustic emissions (DPOAEs) and 1 kHz admittance tympanometry. DPOAE ($2f1$-$f2$) levels for $f2$ frequencies of 2, 3, and 4 kHz needed to meet criteria of ⩾6 dB signal-to-noise ratio and an absolute level of −10 dB sound pressure level (SPL) or higher. At the time data were collected for the present study, two published stud-

ies reported normative data for infants using 1 kHz tympanometry (Margolis *et al.*, 2003; Kei *et al.*, 2003). Based on the 10th–90th percentiles for positive tail peak-compensated SAA reported by these studies, a SAA criterion of $\geqslant$0.3 mmho was used as an infant inclusion criterion. Data from 41 infants were not included in the final analysis due to failure to obtain a complete data set. Reasons for not obtaining a complete data set included cooperation issues such as fussiness or excessive movement (27 infants), failure to meet tympanometric criteria (7 infants), failure to pass DPOAE screening (3 infants), or experimenter error in data acquisition (4 infants). A complete data set was obtained in as short as 20 min if the infant was cooperative. However, additional time was often required for the infant to go to sleep, be fed, or become settled. It was not required that the infant be asleep for testing, and many infants were tested while awake.

Twenty-one adult participants were recruited using flyers posted at various locations on the University of Washington campus. The final adult data set included a total of 20 adults (15 female, 5 male) with a mean age of 24.6 (SD =2.9) years. Adult participation selection criteria included (1) good general health, (2) a negative history of chronic middle-ear disease, (3) 226 Hz tympanometric criteria within normal limits (Margolis and Heller, 1987), (4) hearing within normal limits (air conduction thresholds $\leqslant$15 dBHL and air bone gap $\leqslant$10 dB), and (5) having ear canals free of wax and tympanic membranes free of scarring as assessed using otoscopy. One subject was excluded due to history of a middle-ear disorder, which was revealed after testing was completed. All screening and experimental measurements were obtained from subjects while seated in a double-walled sound treated booth.

B. Instrumentation and signal processing

Experimental measurements were made using a Welch Allyn prototype diagnostic middle ear (DME) analyzer system. A similar PC-based prototype system had been used for conducting ATF measurements in adults and children (Keefe and Simmons, 2003). The PC housed a multichannel digital signal processor, with one channel providing the wideband probe signal output and the other channel digitizing the response from the probe microphone. A second output channel from the DME system was connected to a modified Grason Stadler immittance instrument (GSI-33) enabling the static ear-canal pressure to be controlled via the DME system.

The DME probe consisted of a custom made receiver/microphone assembly leading to a probe tip with three separate ports for receiver output, microphone input, and pressure adjustments. Modified (shortened length) GSI tympanometry ear tips were placed on the end of the probe tip to allow for a pressure seal when placed in the participant's ear canal.

The broadband stimuli were output from a digital-to-analog converter at a 24 kHz sampling rate. The broadband acoustic chirps used for infant measurements had a duration of 40 ms with a frequency bandwidth extending from 0.25 to 8 kHz. The stimuli used with adults were acoustic clicks with bandwidths of 0.25–8 kHz. The broadband click stimuli were presented at a rate of 11.7 per second. The output levels of the clicks and chirps were 57 and 60 dB SPL, respectively, as measured in a 2 cm^3 coupler. The differences in stimuli used for infants and adults were based on conventions of the DME prototype (separate probe calibration procedures existed for adults and infants using click and chirp stimuli, respectively). Differences in ER measurements arising from using different stimuli were not expected as long as adequate calibrations were obtained (see Sec. II C for details). The microphone signal was digitized at a 24 kHz sampling rate using an analog-to-digital converter, high-pass filtered to eliminate low-frequency noise below 0.25 kHz and stored for data analysis. The software incorporated a noise-rejection algorithm, such that only ear-canal-response levels that were at or below a specific criterion were included in the signal averaging. A time-averaged-waveform response (an average of eight stimuli) for each individual pressure condition was obtained and stored for analyses.

C. Calibration procedure and reflectance calculations

The ER and admittance measurements were made based on a calibration procedure in which the Thevenin source impedance and source pressure of the DME probe were calculated. For a detailed review of the calibration procedure theory for ATF measurements see Keefe and Simmons (2003). Calibration procedures were performed only at ambient pressure. This decision was based on previous work that confirmed that only slight changes in calibration results occurred under pressurized conditions (Keefe and Simmons, 2003). The daily calibration procedure for the DME probe involved calculating the Thevenin source impedance and pressure by modeling the acoustic wave propagation inside two rigid-walled cylindrical calibration tubes. Each tube was open at one end and closed at the other end with a steel plug. During calibration measurements, the DME probe was coupled to a given tube using a modified GSI rubber ear tip. With the probe placed in the tube opening, the broadband stimuli were presented, and the response was fitted to a model of acoustic wave propagation for each smooth-walled cylindrical tube. This method estimated the Thevenin equivalent parameters of the probe system. The DME software executed all calculations involved in the calibration procedure. Sets of "small" and "large" calibration tubes were used to obtain Thevenin equivalent parameters for infant and adult wideband ATF measurements, respectively. The infant calibration tubes had lengths of 236.7 and 61 mm, each with a diameter of 4.8 mm, which is a close approximation to ear-canal diameter in infants. The adult calibration tubes had lengths of 296.2 and 84 mm, each with a diameter of 7.9 mm, which is a close approximation to ear-canal diameter in adults. To assess the validity of the calibration procedure, a software-implemented root mean squared (rms) reflectance error function was generated to determine the error in the acoustical estimate of the length of each cylindrical tube relative to the acoustic wave propagation model. A rms reflectance error of $\leqslant$0.01 was required for a successful calibration. A successful calibration measurement was completed within 2–3 min.

The DME software used the source impedance and pressure from the calibration procedure and subsequent ear-canal measurements to calculate the pressure reflection coefficient (R_p). R_p was calculated as

$$R_p = (Z - Z_c)/(Z + Z_c), \quad (1)$$

where Z is the impedance measured in the ear canal and Z_c is the characteristic impedance of the ear canal (Keefe *et al.*, 1992). The characteristic impedance of the ear canal is needed to calculate R_p and is equal to

$$\rho c/S, \quad (2)$$

where ρ (rho) is the density of air, c is the phase velocity of sound in air, and S is the area of the ear canal. Since an actual area measurement of the ear canal was not practical, the DME system utilized an estimate of ear-canal area equal to that of each calibration tube to calculate Z_c. The squared magnitude of the pressure reflectance,

$$|R_p|^2, \quad (3)$$

was used to calculate ER.

D. Wideband tympanometry measurements

External-ear-canal pressure changes were accomplished using the GSI-33 pressure pump, controlled by the DME system. Measurements with the DME system were made at static pressures of −200, −100, −50, −25, 0, +25, +50, +100, and +200 daPa relative to TPP values obtained during screening measurements with 226 Hz or 1 kHz tympanometry for adults and infants, respectively. For instance, if a participant presented with a TPP of −10 daPa, the extreme static pressures were set to −210 and +190 daPa, with interim static pressures adjusted accordingly. An ambient pressure ATF measurement was obtained before making the static pressure measurements. Following placement of the ear tip and during the initial ambient ATF test, the DME system performed an acoustic leak check. An acoustic leak was determined by an estimation of an appropriate ear-canal volume. If a leak was detected, which was usually the result of an inadequate probe fit in the ear, the experimenter was prompted by the DME software to check the probe fit and run the test again. When an appropriate fit was obtained, all ambient and static pressure conditions were then run. Measurements from the single ambient and nine ear-canal pressure conditions constituted one "trial" and took approximately 1 min to complete. However, if an infant was sufficiently noisy, averaging would be suspended by the DME system's artifact rejection algorithm; therefore, some trials with infant participants took longer to complete. The probe was kept in place for the entire trial. In the event a probe seal was lost during a particular trial, the entire trial was repeated.

E. Coupler measurements

Over the course of five months of data collection, wideband ER measurements were made periodically (11 times) in a Zwislocki coupler (Knowles Electronics, DB-100 with DB-050 canal extension) to ensure consistency of stimulus presentation and response characteristics. These measurements were made in a sound treated booth, with the probe inserted into the Zwislocki coupler, terminated by a $\frac{1}{2}$ in. microphone, and placed on a foam rubber pad. Table I presents ER SD values by frequency for 11 coupler measurements. Small SD values (all less than 5%) suggest minimal change in probe calibration characteristics over the course of the study.

TABLE I. Energy reflectance (ER) standard deviation (SD) values at octave frequencies for 11 ER measurements made with the same probe over the course of five months in a Zwislocki coupler (Knowles Electronics, DB-100 with DB-050 canal extension).

Frequency (Hz)	ER SD
250	0.016
500	0.026
1000	0.011
2000	0.026
4000	0.013
8000	0.046

III. RESULTS

Multiple acoustic variables may be derived from wideband ATF measurements and examination of more than one variable is useful when interpreting wideband ATF results. ER and acoustic admittance measurements, including admittance magnitude ($|Y|$), admittance phase, and conductance (G), are examined in the following sections.

A. Wideband energy reflectance

Figure 1 (left panel) illustrates one-third-octave averaged ER results for each age group at nine ear-canal pressures (−200, −100, −50, −25, 0, +25, +50, +100, and +200 daPa, relative to TPP). In addition, ER pressure differences are presented in Fig. 1 (right panel). ER pressure differences are defined as ER at a given static pressure condition (ER_{SP}) minus ER at TPP (ER_{TPP}). Therefore, a positive value on the y axis for Fig. 1 (right panel) indicates an increase in ER at a given static pressure relative to ER at TPP, while a negative value indicates a decrease in ER relative to ER at TPP.

Ear-canal pressure generally resulted in increased ER relative to TPP. Examination of 0.25 kHz in Fig. 1 (right panel) for each age group shows that the degree of change in ER with pressure steadily decreased as age increased, with an approximately 30% change in ER at 0.25 kHz for 4-week-old infants and less than a 10% change at the same frequency for adults. ER difference patterns for 4-week-old infants exhibited a multipeaked effect across frequency, with maxima in the ER difference functions occurring near 0.25 and 1.5 kHz. This multi-peaked pattern became less distinct for the 12- and 24-week-old infants and shifted to a single-peaked function in the adult data. The greatest pressure induced increases in ER for all age groups occurred at frequencies between 0.75 and 2 kHz (see Fig. 1, right panel). For 4-week-old infants, between 2 and 6 kHz, and to a lesser extent 8 kHz, the effects of positive and negative ear-canal

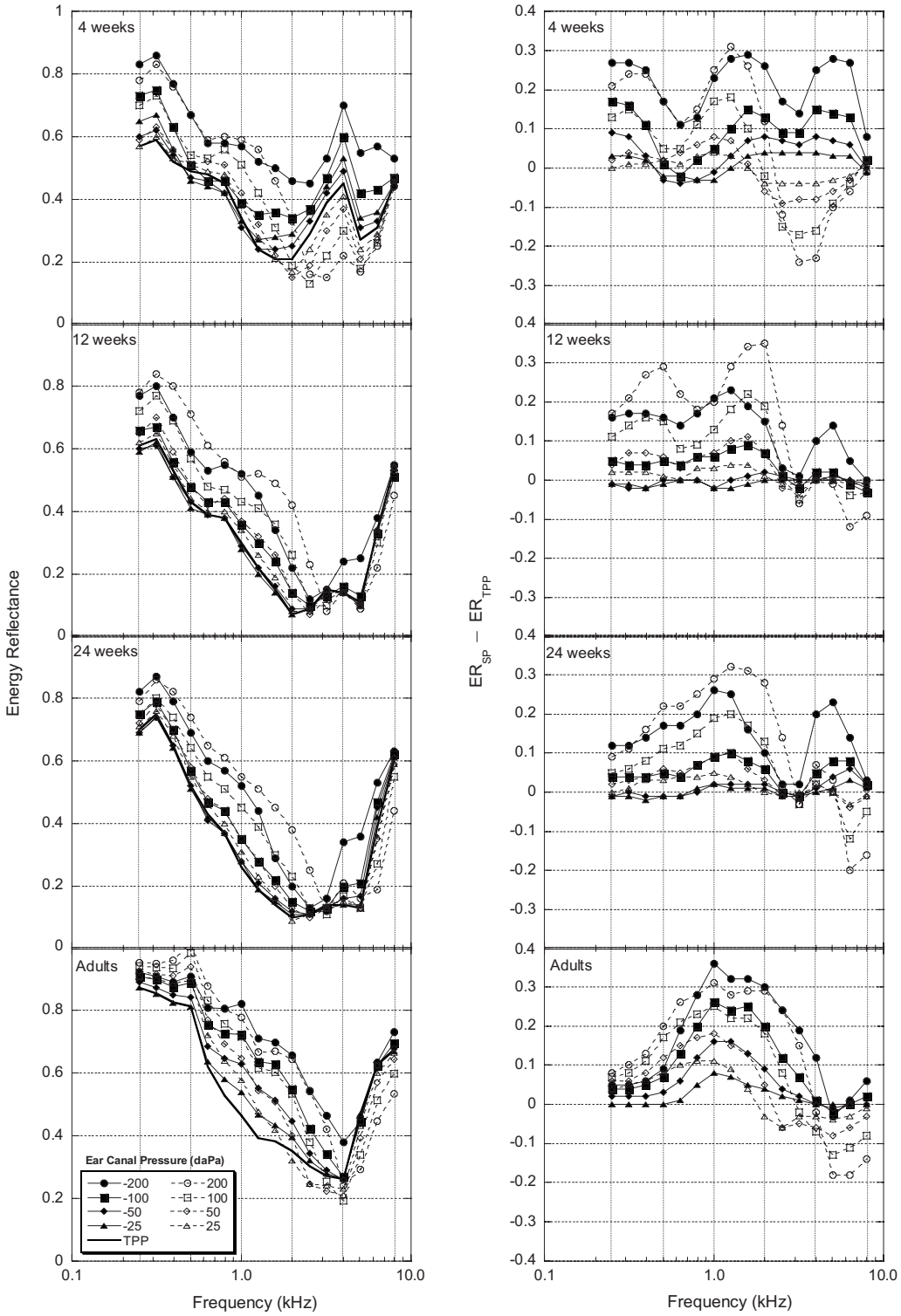


FIG. 1. Group mean one-third-octave energy reflectance (ER) and ER differences (left and right panels, respectively) plotted as a function of frequency for all age groups ($N=20$ for each group). ER pressure differences are defined as ER at an individual static pressure (ER_{SP}) minus ER at TPP (ER_{TPP}).

pressures of the same magnitude (i.e., ±25, ±50, etc.) were opposite in nature, with negative pressures causing increased ER and positive pressures resulting in decreased ER. This effect was generally limited to more extreme ear-canal pressures (i.e., ±200 and ±100 daPa) for frequencies greater than 5 kHz for the two older infant age groups. For adults, on average, both positive and negative pressures resulted in similar ER increases. However, some positive pressures, for frequencies above 2 kHz, resulted in decreased ER while negative pressures had minimal effects.

As a measure of variability, ER SDs were calculated (see Fig. 2). Overall, across frequency and pressure conditions, the adult data showed the smallest SDs. With the exception of the 12-week-old infants, SDs generally increased

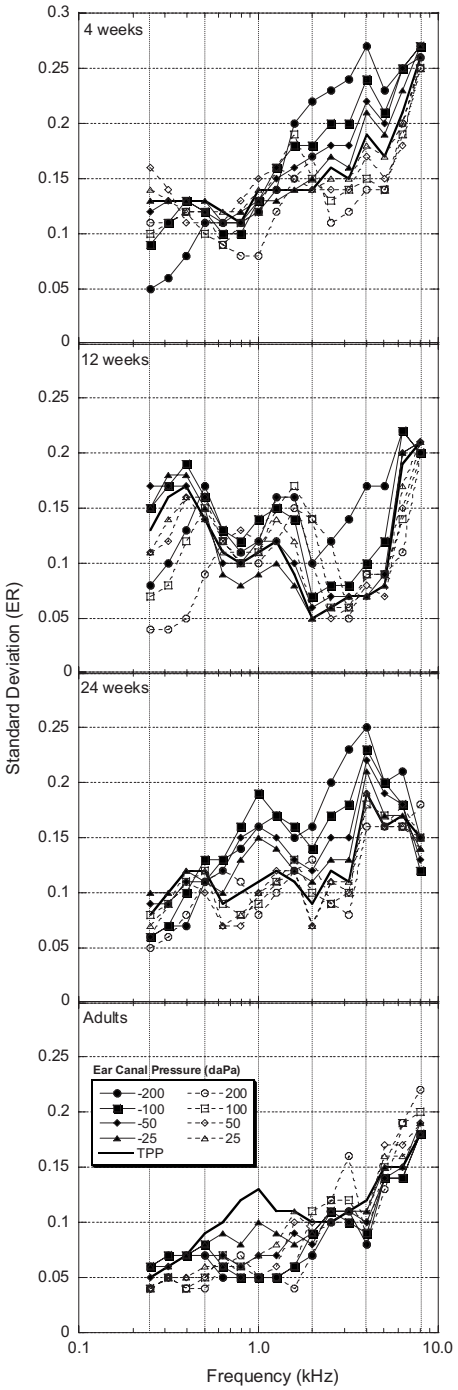


FIG. 2. Energy reflectance (ER) standard deviations (SDs) as a function of frequency for all age groups. The parameter is ear-canal pressure.

as frequency increased. For the infant groups, negative ear-canal pressures tended to cause SDs to increase, relative to SDs at TPP. Some pressure conditions (mostly positive) caused SDs to decrease for the infant groups while nearly all pressure conditions for the adult group resulted in either no change or decreases in SD relative to SDs at TPP.

1. Statistical analyses

Separate two-way analyses of variance (ANOVAs) were calculated at 16 one-third-octave center frequencies to examine the effects of age and pressure on ER pressure differences. An alpha level of 0.05 was used for all analyses. Age and pressure were significant factors at all frequencies. *Post hoc* analyses (Tukey tests) were computed to test for significant differences between age groups. Statistically significant differences in changes to ER as a result of ear-canal pressure were observed between infants and adults. Table II shows the ages and frequency/pressure conditions for which infant ER pressure differences were significantly different from adults. At frequencies below 0.5 kHz ER pressure differences for the youngest infants were significantly different from adults, with most differences found at extreme (±200 daPa) ear-canal pressures. For this same frequency region, no significant differences in ER pressure differences were found between 24-week-old infants and adults for any pressure condition. Significant age and pressure effects were more prevalent at frequencies between 0.63 and 1.26 kHz with more significant differences detected between infants and adults. At frequencies from 1 to 6.35 kHz there tended to be greater differences between adults and infants for negative-pressure differences than for positive-pressure differences.

Next an analysis was made of significant ER differences between infant groups using *post hoc* analyses (Tukey tests). Table III shows ages (in weeks) for which there were significant differences in ER for different frequency/pressure conditions. The majority of significant age and pressure effects were found between 4-week-old infants and the two older infant groups. Specifically, below 0.5 kHz, significant differences were found across nearly all frequency/pressure conditions. At frequencies between 0.79 and 1.58 kHz, age and pressure were found to be significant factors in only three frequency/pressure conditions. However, this is a frequency region where there were many adult-infant differences. For the frequency region from 2.5 to 6.35 kHz, the effects of pressure for 4-week-old infants were significantly different from the effects of pressure for 12- and 24-week-old infants. In contrast, across this same frequency range (with the exception of 6.35 kHz), there were no significant differences for the effects of pressure between 12- and 24-week-old infants. In fact, for all frequency/pressure conditions, only nine conditions exhibited significant differences between 12- and 24-week-old infants, with approximately half of them occurring below 0.8 kHz for pressures at +100 and +200 daPa and the other half at 6.35 kHz.

B. Wideband admittance

One-third-octave averaged wideband admittance magnitude ($|Y|$) values in mmhos are presented in Fig. 3 (left panel). Like ER pressure differences in Fig. 1, $|Y|$ pressure differences were also calculated and are presented in Fig. 3 (right panel). It should be noted that, unlike ER measurements which are relatively unaffected by probe position in the ear canal, admittance measurements are affected by the immittance qualities of the air space between the probe and the tympanic membrane (Stinson *et al.*, 1982). Therefore, some of the absolute admittance differences observed between groups are due to differences in ear-canal volumes (Keefe *et al.*, 1993). Figure 3 (left panel) shows double-peaked $|Y|$ transfer functions for infants in the 1.5–6 kHz range compared to a single-peaked transfer function for the

TABLE II. Results of *post hoc* analyses (Tukey tests) indicating the ages (in weeks) at which infant ER pressure differences (ERat a given static ear-canal pressure minus ER at TPP) were significantly different from adults (α=0.05). A blank space indicates the lack of a statistically significant difference between any of the infant groups and the adult group.

Frequency (Hz)	Static ear-canal pressure (daPa) relative to tympanometric peak pressure (TPP)							
	−200	−100	−50	−25	25	50	100	200
250	4 12	4	4					4 12
315	4 12	4	4					4 12
397	4 12				4			4 12
500			4		4 12 24	4 12 24	4	12
630		4 12 24	4 12 24	4	4 12 24	4 12 24	4 12 24	4
794	4 12	4 12 24	4 12 24	4 12	4 12	4 12	4 12	4 12
1000	4 12	4 12 24	4 12 24	4 12 24	4 12 24	12	12	12
1260		4 12 24	4 12 24	4 12 24	4 12 24	4		
1587	12 24	4 12 24	12 24	12		4		
2000	12 24	12 24	12 24	12	12	4	4	4
2520	12 24	12 24			12 24	24	4	4
3175	12 24	12 24		4	12 24		4	4 12 24
4000	4	4	4	4	12 24	12 24	4 12 24	4 24
5040	4 12 24	4 24	4 24	4	12 24	12 24	12 24	4 12 24
6350	4 24	4	4 24	24				4
8000						4	4	4

adults. Also, as age increased, $|Y|$ increased, with as much as a 27 mmho increase in peak $|Y|$ between 4-week-old infants and adults. Negative pressures, with few exceptions, created decreased $|Y|$ in the infant groups. The most pronounced effect of pressure on $|Y|$ for the infant groups was for positive pressures (mainly 100 and 200 daPa) from 1.5 to 4 kHz, with as much as a 6 mmho increase in magnitude across this frequency range. Especially for the 24-week-old infants, this pressure induced increase in $|Y|$ created a transfer function similar in shape to the adults.

As a measure of the variability in $|Y|$, SDs were calculated across subjects for each pressure condition and at each frequency (see Fig. 4). Overall, there was a nonmonotonic increase in SD as frequency increased. Gradually increasing SDs, with similar values across pressure conditions, were found for all age groups for frequencies from 0.25 to 1.5 kHz. Infant groups exhibited a double-peaked SD function for frequencies between 2 and 6 kHz similar to mean $|Y|$ values. While the lower-frequency SD peak for 4-week-old infants showed similar variability across pressure conditions, the lower-frequency SD peak for 12- and 24-week-old infants was dominated by positive ear-canal pressures.

TABLE III. Results of *post hoc* analyses (Tukey tests) indicating the ages (in weeks) at which infant ER pressure differences (ER at a given static ear-canal pressure minus ER at TPP) were significantly different from each other (α=0.05). A blank space indicates the lack of a statistically significant difference between any of the infant groups.

Frequency (Hz)	Static-ear-canal pressure (daPa) relative to tympanometric peak pressure (TPP)							
	−200	−100	−50	−25	25	50	100	200
250	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24			4–24	4–24 12–24
315	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24			4–24 12–24	4–12 12–24
397	4–12	4–12						12–24
500							4–12	4–12
630					12–24		4–24	4–12 4–24
794								4–24
1000								
1260								
1587	4–12					4–12		
2000	4–24				4–12	4–12 4–24	4–12 4–24	4–12 4–24
2520	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12	4–12 4–24	4–12 4–24
3175	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24
4000	4–12	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24
5040	4–12	4–12	4–12	4–12	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24
6350	4–12	4–12 12–24	4–12 12–24	12–24	12–24			4–24
8000							4–24	4–12 4–24

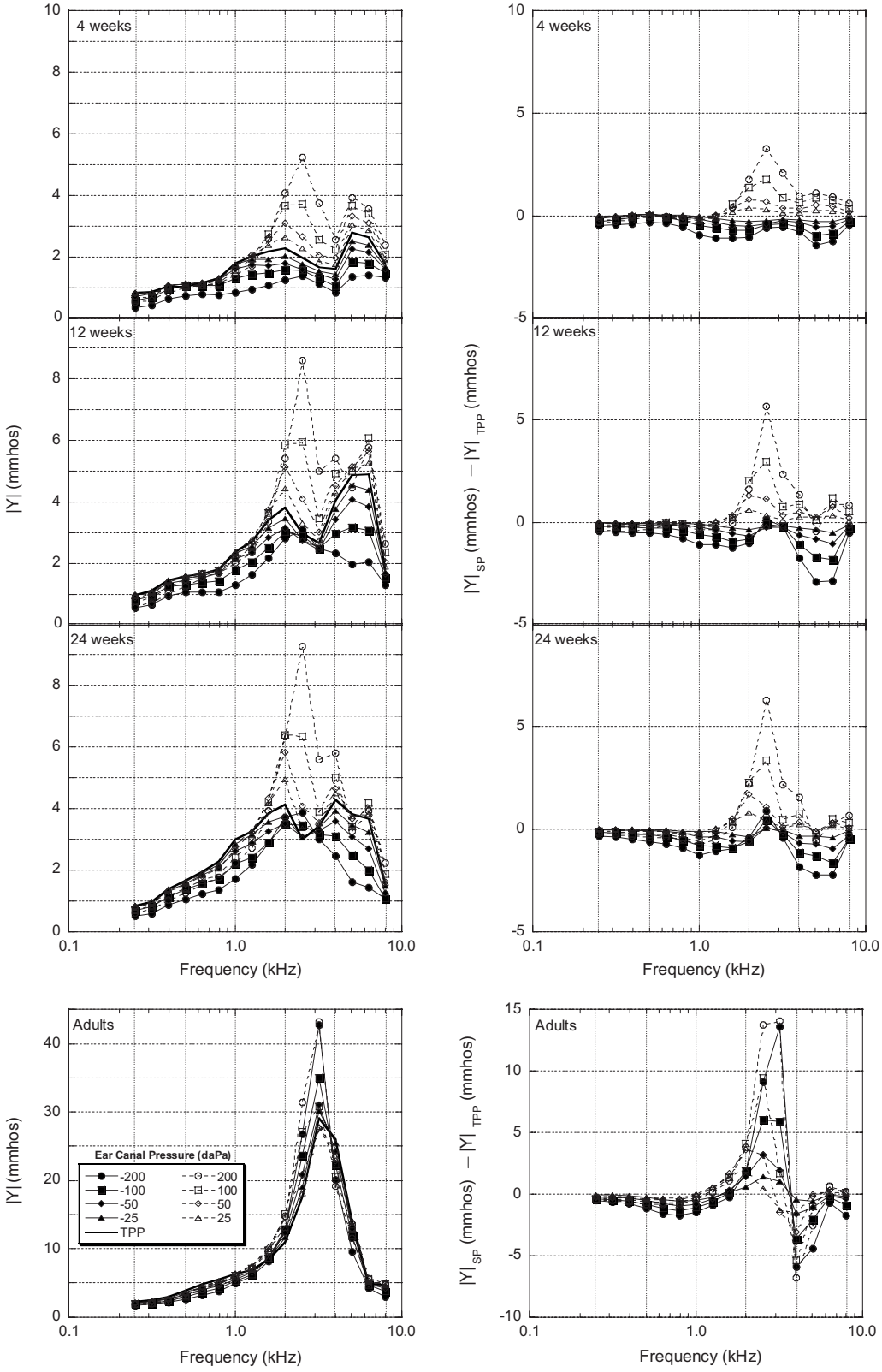


FIG. 3. Group mean one-third-octave admittance magnitude ($|Y|$) and $|Y|$ pressure differences in mmhos (left and right panels, respectively) plotted as a function of frequency for all age groups (N=20 for each group). $|Y|$ pressure differences are defined as $|Y|$ at an individual static pressure ($|Y|_{SP}$) minus $|Y|$ at the TPPs ($|Y|_{TPP}$). Note the difference in scales on the y axes for infants and adults.

1. Statistical analyses

Separate two-way ANOVAs and *post hoc* analyses (Tukey tests) were calculated at 16 one-third-octave center frequencies for admittance data. Tables IV and V show the frequency/pressure regions where changes in $|Y|$, relative to TPP, were statistically significant. Table IV shows results for comparisons between adult and infant groups and Table V shows the results for comparisons between infant groups.

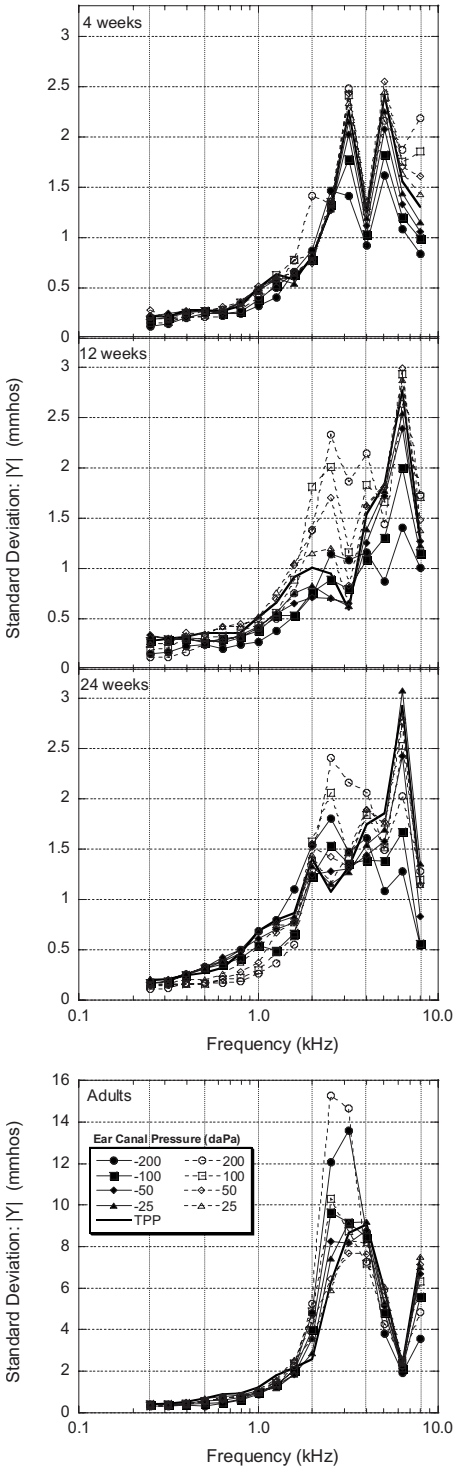


FIG. 4. Admittance magnitude $|Y|$ standard deviations (SDs) as a function of frequency for all age groups. The parameter is ear-canal pressure. Note the difference in scales on the y axes for infants and adults.

The patterns of significant differences for $|Y|$ are similar to those of ER (see Table II), with the midfrequency region for both ATF measures showing fewer significant differences between ages. However, more differences between 12- and 24-week-old infants were found for $|Y|$ pressure differences compared to ER pressure differences, especially at 1.58 and 2 kHz. Compared to ER pressure differences between infants and adults shown in Table II, significantly different $|Y|$ pressure differences shown in Table IV tend to group more toward frequencies ranging from 1.26 to 5.4 kHz.

2. Admittance phase

Mean one-third-octave admittance phase is illustrated in Fig. 5. The phase at TPP (solid line) for the infant groups revealed a general negatively directed monotonic shift from stiffness to mass controlled phase as frequency increased from 0.25 to 2 kHz, with nonmonotonic changes in phase for frequencies from 2 to 8 kHz. Adult phase followed a similar pattern out to 5 kHz, and then shifted toward a positive direction out to 8 kHz. At frequencies below 1 kHz for the 4-week-old infants, admittance phase at TPP was near +45°, indicating equal contributions of conductance and susceptance components. Admittance phase angle shifted in a positive direction with increasing age at frequencies below 1 kHz for the infant groups, indicating increased stiffness with increasing age. Data in Fig. 5 illustrate differences in the effects of ear-canal pressure on phase, relative to phase measurements at TPP. Positive ear-canal pressures created the greatest change in phase for the infant groups, with the largest changes occurring between 1 and 3 kHz. The most common effect of static ear-canal pressure changes for the infants was an increase in phase, indicating an increase in the stiffness component. Ear-canal pressure had smaller effects on phase for the adult group, with positive and negative ear-canal pressures effecting similar changes in phase.

3. Conductance

Mean one-third-octave conductance (G), the real component of admittance, is presented in Fig. 6. Four-week-old infant G values from 0.5 to 2 kHz were reduced (<3 mmho) with the introduction of either negative or positive ear-canal pressures. At frequencies greater than 2 kHz for the 4-week-old group, positive pressures created the greatest changes in G, but only by amounts no greater than 3 mmho. The effects of positive pressures were larger (up to 5 mmho) in the 12- and 24-week-old infants. In contrast, both positive and negative ear-canal pressures had similar effects across frequency for the adult group.

4. Wideband admittance tympanograms

Mean wideband $|Y|$ tympanograms are plotted at octave frequencies from 0.25 to 4 kHz in Fig. 7. The tympanograms for the adult group (bottom panels) were single peaked from 0.25 to 1 kHz, followed by a 2 kHz tympanogram with a "notch" (an abrupt decrease in admittance followed by an increase as a function of pressure), and then returning to single-peaked functions at 4 kHz. $|Y|$ tympanograms for 12- and 24-week-old infants were similar in morphology to the adult tympanograms with respect to the absence of notching up to 2 kHz. In contrast, the 4-week-old infant data showed a single-peaked tympanogram at 0.25 kHz with broad notching appearing at 0.5 kHz and a single-peaked tympanogram at 1 kHz. The tympanograms at these frequencies were consistent with previous research in infants (Holte *et al.*, 1991). Wideband admittance tympanograms for 4 week old at 2 and 4 kHz exhibited different patterns than those observed in

TABLE IV. Results of *post hoc* analyses (Tukey tests) indicating the ages (in weeks) at which infant admittance magnitude ($|Y|$) pressure differences ($|Y|$ at a given static ear-canal pressure minus $|Y|$ at TPP) were significantly different from adults (α=0.05). A blank space indicates the lack of astatistically significant difference between any of the infant groups and the adult group.

Frequency (Hz)	Static ear-cannal prassure (daPa) relative to tympanometric peak pressure (TPP)							
	−200	−100	−50	−25	25	50	100	200
250	4 12 24	4						4 12
315	4 12 24				24			4 12
397	4		4	4	12	4	4	
500		4	4	4	12	4	4	4
630		4	4 12 24	4	12	12	12	4
794			4 12	4 12				
1000	4			4				4 24
1260	4 12 24	4 12 24	4		4 24			4 24
1587	4 12	4 12	4 12	4 12				
2000	4 12	4 12	4 12	4 12	24	24	24	4 24
2520	4 12	4 12	4 12 24	4 12	4 24	4	4 12 24	4 12 24
3175	4 12 24	4 12 24	4 12 24	4 12	4 12 24	4 12	4	4
4000	4 12 24	4 24	4 12 24	4 24	4 12 24	4 12 24	4 12 24	4 12 24
5040	4 12 24	4 12 24	4 12 24	4 12 24	4 12 24	4 12 24	4 12 24	4 12 24
6350	4 24							
8000								

adults. Compared to the adult tympanograms, infant tympanograms were less symmetrical around TPP and the negative $|Y|$ tails exhibited lower $|Y|$ values relative to positive $|Y|$ tails.

IV. DISCUSSION

A. Wideband energy reflectance

1. Ambient pressure conditions

As ER measurement methods evolve, it is important to compare results and identify similarities and differences across studies. Mean one-third-octave *ambient* ER data for 1-, 3-, and 6-month-old infants and adults from Keefe *et al.* (1993) and for adults from Feeney *et al.* (2003) are plotted along with mean one-third-octave ambient data from the present study in Fig. 8 (error bars indicate ±1 standard deviation from the mean). The ambient ER measurements from the present study illustrated in Fig. 8 were obtained prior to the ER measurements obtained at static ear-canal pressures. While there are some overall magnitude differences in ER results between the different studies at each age, the patterns of the ER responses from the present study are consistent with data from Keefe *et al.*(1993) and Feeney *et al.* (2003). The largest differences between ER results for the two older infant groups and adults are mostly found below 2 kHz. The differences at lower frequencies may have been influenced

TABLE V. Results of *post hoc* analyses (Tukey tests) indicating the ages (in weeks) at which infant admittance magnitude ($|Y|$) pressure differences ($|Y|$ at a given static ear-canal pressure minus $|Y|$ at TPP) were significantly different from each other (α=0.05). A blank space indicates the lack of a statistically significant difference between any of the infant groups.

Frequency (Hz)	Static ear-canal pressure (daPa) relative to tympanometric peak pressure (TPP)							
	−200	−100	−50	−25	25	50	100	200
250	4–12 4–24	4–12 4–24	4–12 4–24				4–24	4–24
315	4–24		4–12 4–24					
397			4–12				4–12 4–24	4–12
500		4–12 4–24	4–12 4–24	4–12 4–24			4–24	4–12 4–24
630		4–24					12–24	4–24 12–24
794							12–24	12–24
1000								
1260	4–12 4–24							
1587	4–12 4–24 12–24	4–24 12–24	4–24 12–24	4–24 12–24	4–24 12–24	4–24 12–24	4–24 12–24	4–24 12–24
2000	4–12 4–24 12–24	4–24 12–24	4–24 12–24	4–24 12–24	4–24			
2520	4–12 4–24	4–12 4–24	4–12 4–24	4–24				
3175		4–12 4–24	4–24	4–24	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24
4000				4–12	4–12 4–24	4–12 4–24	4–12 4–24	4–12 4–24
5340								
6350					4–12	4–12	4–12	4–12
8000								

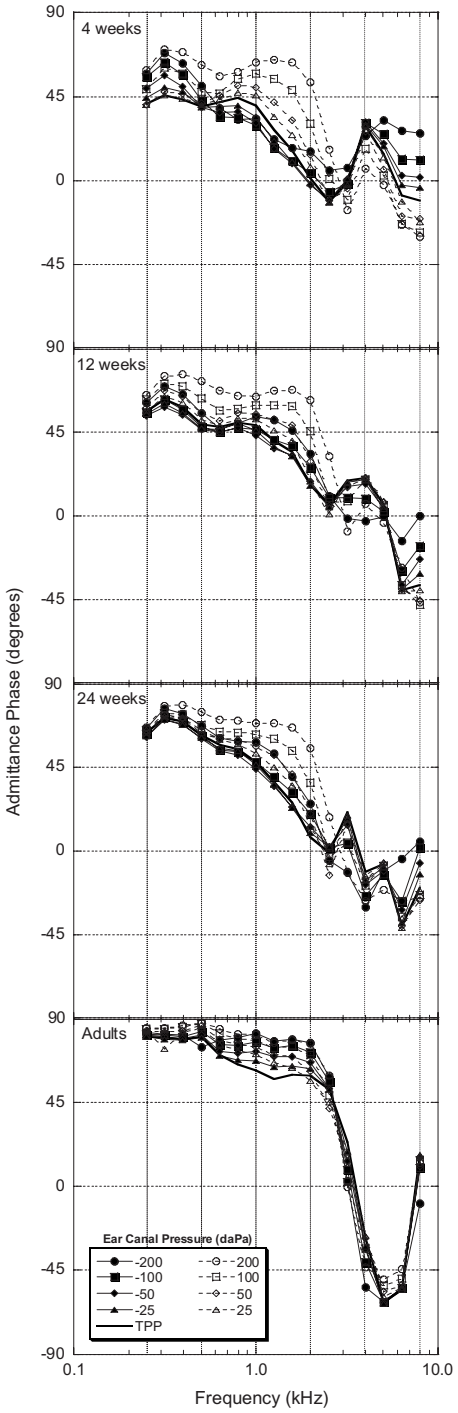


FIG. 5. Group mean one-third-octave admittance magnitude ($|Y|$) phase in degrees as a function of frequency for all age groups (N=20 for each group).

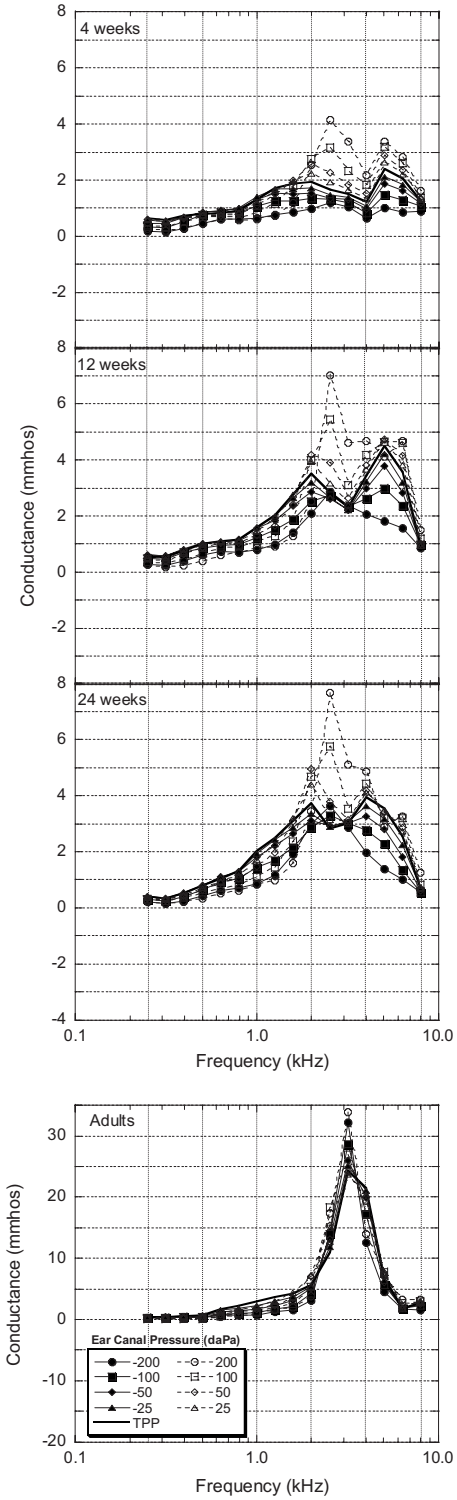


FIG. 6. Group mean one-third-octave conductance values (mmhos) plotted as a function of frequency for all age groups (N=20 for each group). Note the difference in scales on the y axes for infants and adults.

by different methods used to estimate ear canal area. Keefe *et al.* (1993) used an acoustic estimate, while data from the present study were calculated using a set value based on the diameter of the calibration tubes (see Sec. II C). Variation in ER measurements can occur if the ear canal is of a different diameter than the calibration tubes. The diameter of the calibration tubes used for infant measurements was 4.8 mm, close to the ear-canal diameter of 1-month-old infants (4.4 mm) as reported by Keefe *et al.* (1993); this age group had the best ER agreement between studies. The same ear-canal diameter (4.4 mm) used in measurements made in 12- and 24-week-old infants, with larger ear-canal diameters, could contribute to the lower ER results, relative to data from Keefe *et al.* (1993).

2. Static pressure conditions

Results from the present study illustrate maturational effects on acoustic ear-canal measurements made under pres-

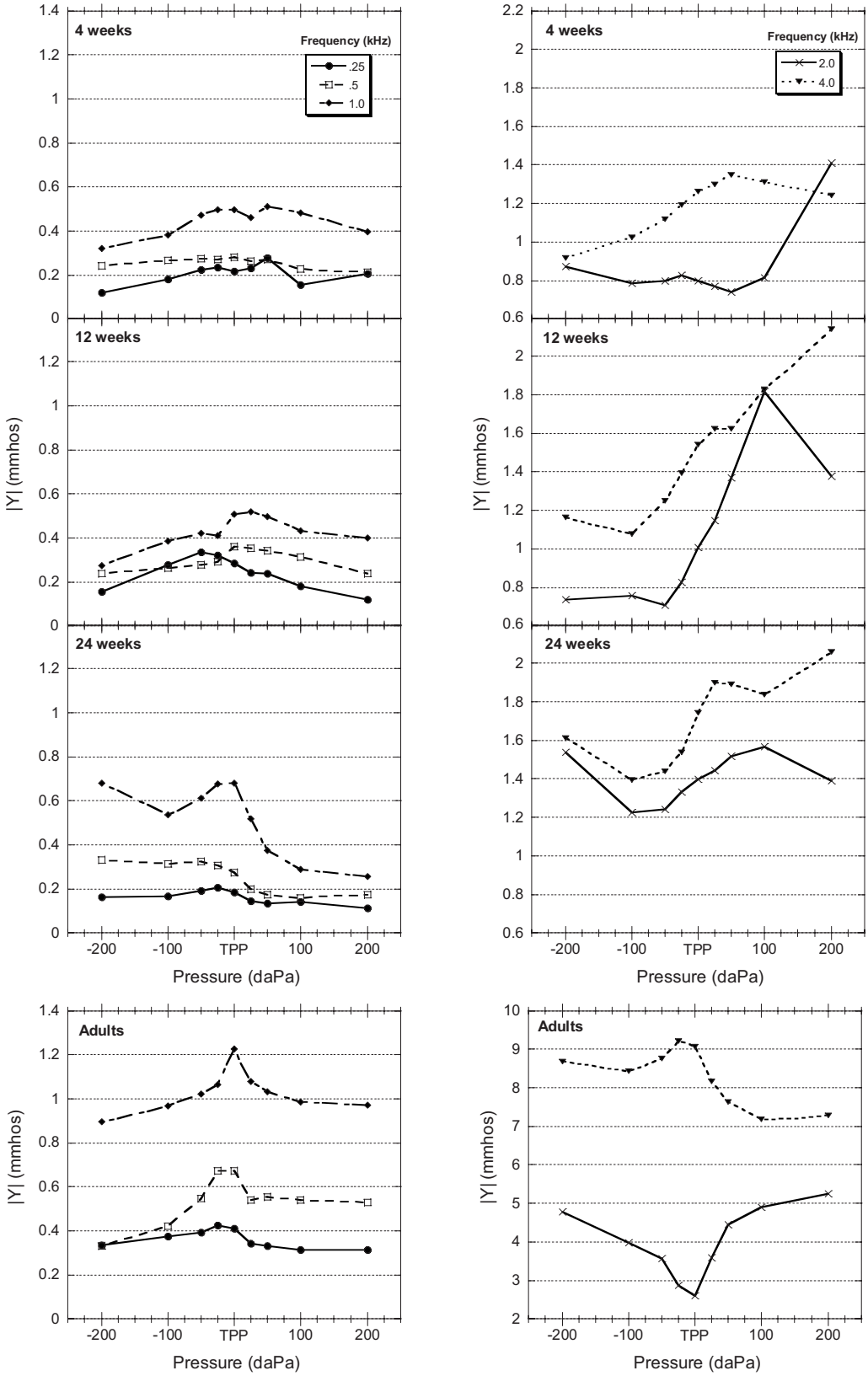


FIG. 7. Group mean one-octave admittance magnitude ($|Y|$), in mmhos, plotted as a function of pressure (daPa) for all age groups (N=20 for each group). Note the difference in scales on the y axes for infants and adults.

surized ear-canal conditions. ER results for 4-week-old infants, especially in the frequency regions below 0.75 kHz and between 2 and 6 kHz, are evidence of a middle-ear system that responds differently to changes in ear-canal pressure than older infants and adults. The magnitude of ER change with positive and negative ear-canal pressures at and below 0.5 kHz was greatest for the 4-week-old infants. This effect may be the result of a more compliant energy absorbing ear-

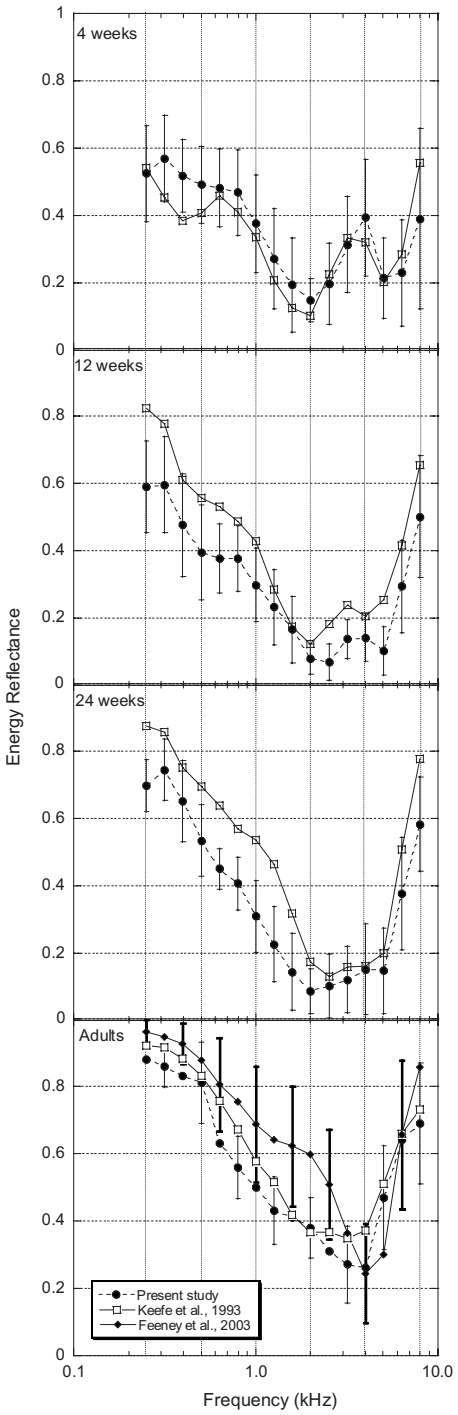


FIG. 8. Group mean one-third-octave ambient energy reflectance ER as a function of frequency for all ages for the present study (N=20 for each group) and for infants and a group of adults from Keefe *et al.* (1993). The numbers of subjects for each age group in Keefe *et al.* (1993) are as follows: 1 month=15, 3 months=18, 6 months=11, and adults=10. The bottom panel also includes young adult ER data (N=40, 75 ears) from Feeney *et al.* (2003). Error bars indicate ±1 SD from the mean. Note that to avoid overlapping error bars in the lower panel, adult SD data are plotted for alternating frequencies for the present study and Feeney *et al.* (2003).

canal wall (Keefe *et al.*, 1993), which becomes stiffer with the introduction of ear-canal pressure and in turn reflects more acoustic energy. The gradual decrease in magnitude of this effect with age may reflect the developmental change from a cartilaginous to ossified infant ear canal (Saunders *et al.*, 1983). The decrease in ER in response to positive pressures between 2 and 6 kHz for the 4-week-old infants is evidence of a more energy-absorbent infant ear under positive-pressure conditions. The effects of pressure on ER for 12- and 24-week-old infants were similar to each other and, while still different than adults, were more adultlike with respect to their overall patterns than those of the 4-week-old group.

In general, the results observed in the adult group were similar to wideband ATF data presented in previous work (Keefe and Simmons, 2003; Margolis *et al.*, 1999), showing increases in ER with pressure up to approximately 3 kHz, with a slight decrease in ER with positive pressures above 3 kHz. In fact, for all age groups, there were only minimal decreases in ER with the introduction of negative ear-canal pressures.

B. Wideband admittance

Changes in $|Y|$ at TPP with age in the present study are consistent with previous work in infants that has shown a general trend of increasing $|Y|$ as a function of age (Holte *et al.*, 1991; Meyer *et al.*, 1997). With regard to changes in $|Y|$ resulting from changes in static ear-canal pressures, positive pressures had the greatest effect on $|Y|$, with similar overall response patterns across infant groups (see Fig. 3). Positive pressures for frequencies from 1 to 4 kHz created more adultlike (peaked) $|Y|$ functions. Negative ear-canal pressures had a different effect for the infant groups when compared to the adult group, especially between 2 and 4 kHz. For example, in the infant groups, negative pressures always caused $|Y|$ to decrease, whereas positive pressures caused both increases and decreases in $|Y|$ depending on frequency. This differential pressure effect suggests that the characteristics of young infant ears are different than the ears of older infants and adults. Possible explanations for these differences are discussed further below.

Admittance phase results at TPP from the present study are in general agreement with data from Keefe *et al.* (1993) and Holte *et al.* (1991). The youngest infants in Keefe *et al.* (1993), Holte *et al.* (1991), and the present study (Fig. 4) present with nonmonotonic changes in admittance phase as a function of frequency. This pattern is suggestive of ear-canal and/or middle-ear resonance effects, possibly due to vibration of the ear-canal wall and/or immature tympanic membrane or middle-ear characteristics. Increased ear-canal wall compliance may partly explain the multiple resonances observed in the infant groups, and the gradual decline of the resonance at 0.5 kHz with increasing age may reflect developmental changes in ear-canal compliance properties. The phase function zero crossing, indicative of equal contributions of outer and middle-ear compliance (+ phase) and mass (− phase) components, increases with age, shifting from 2 kHz for the 4-week-old infants to 4 kHz for the adults. The larger adult ear-canal volume (resulting in increased compliance) is the most likely contributor to the higher-frequency $|Y|$ phase zero crossing in adults. However, increased middle-ear or ear-canal stiffness in adults could also contribute to this upward frequency shift.

Conductance values (G) are of interest inasmuch as they play a part in the power (P) input into the middle ear, as shown through the relationship

$$P = \tfrac{1}{2} G|p|^2,$$

where p represents sound pressure amplitude. Adult conductance values, obtained using a similar calibration technique (Thevenin source parameters), presented by Keefe *et al.* (1993) and Neely and Gorga (1998) are similar to those obtained in the present study. Examination of conductance values at TPP in Fig. 6 reveals increases in conductance as age increases, with the largest increases occurring above 2 kHz. Based on evidence showing behavioral threshold differences between infants and adults (Olsho, 1988), Holmer (2001) examined the relationship between middle-ear conductance measured at ambient ear-canal pressure and behavioral thresholds in 11- and 24-week-old infants and adults. While Holmer found large changes in behavior thresholds (up to 30 dB SPL) she reported only small changes in conductance between infants and adults at 0.5, 1, and 2 kHz. At 4 kHz, however, Holmer observed that as age increased, conductance increased, but behavioral thresholds decreased, suggesting a relationship between improved behavioral thresholds and increases in middle-ear conductance. Specifically, Holmer reported that between 11-week-old infants and adults, conductance increased by a mean of 8 dB (ranging from approximately 2 to 10 dB re: 1 mmho) and thresholds decreased by 27 dB. Holmer noted that while the developmental changes in conductance do not account for all of the infant/adult differences in behavioral thresholds at 4 kHz, they appear to be responsible for part of them. Conductance values at TPP for 4 kHz in the present study also showed age-related increases, in particular, conductance values increased from 4 mmho at 12 weeks to 20 mmho for adults. These conductance values at 4 kHz, converted to decibels in the same manner as the data from Holmer ($10 \log_{10} g$ re: 1 mmho), were 6 and 13 dB for 12-week-old infants and adults, respectively. This 7 dB difference in conductance at 4 kHz between 12-week-old infants and adults for the present study is within 1 dB of the mean differences reported by Holmer, suggesting consistency in results across studies for these ages.

ATF for G and $|Y|$ each exhibited multipeaked structures. Similar multipeaked structures were obtained by Keefe *et al.* (1994) for diffuse-field pressure transfer functions from a reverberant field to the infant ear canal and for diffuse-field absorption cross-section levels (DACLs). Keefe *et al.* (1994) calculated the DACLs using their diffuse-field pressure transfer function data along with G data from Keefe *et al.* (1993). The DACLs, defined as a quantitative measure of the acoustic response of the external ear and power transfer into the middle ear, were compared across infants and adults (see Keefe *et al.*, 1994, Fig. 14). Conductance ATFs from the present study (Fig. 6) are similar in shape to DACL functions shown by Keefe *et al.* (1994) and also follow similar transitions with age. Functions from both studies for 4-week-old infants had two peaks (at 2 and 5 kHz), and as age increased, the higher-frequency peak gradually shifted downward in frequency, possibly representing a downward shift in the ear-canal or middle-ear resonant frequency. In addition, the overall level and magnitude of the functions increased as age increased. For the infant groups, with few exceptions, only positive pressures created increased G or increased power transfer into the middle ear. However, both positive and negative pressures resulted in increased G for the adult group, mainly around the peak of the G transfer function.

1. Wideband admittance tympanograms

Wideband $|Y|$ tympanograms constructed from the present data showed increases in absolute $|Y|$ with age (see Fig. 7). Holte *et al.* (1991) reported similar developmental changes in multifrequency admittance tympanograms for infants less than 5 month old. Holte *et al.* (1991) investigated changes from 0.226 to 0.9 kHz and found increasing $|Y|$ as a function of age for frequencies above 0.226 kHz. Examination of tympanometric shape in the present study reveals a broad notch in the 4-week-old admittance tympanograms at 0.5 kHz. This is consistent with infant admittance tympanometry patterns presented by Holte *et al.* (1991), which were suggestive of canal-wall resonance in the low-frequency region hypothesized to be related to the more compliant nature of the infant ear-canal wall. While tympanogram notching at low frequencies is not observed for the 12- and 24-week-old infants (see Fig. 7), there is some evidence of notching at 2 kHz. The adult group also exhibited a notch in the tympanogram at 2 kHz. The frequency region where admittance notching is observed in adults is consistent with previous research, showing the 10th–90th percentile range for adult middle-ear resonant frequency of 0.8–2 kHz (Margolis and Goycoolea, 1993).

Differences between $|Y|$ at extreme positive and negative ear-canal pressures ($\pm$200 daPa) have been reported in previous studies in young infants for 226 Hz and 1 kHz tympanograms (Kei *et al.*, 2003; Margolis *et al.*, 2003). Results from the present study show similar results and present new data that show a continuation of this pattern up to 4 kHz. If the negative tails of the 12- and 24-week-old infants' 2 kHz tympanograms were adjusted upward to be equal in magnitude with the positive tails, they would appear fairly similar in shape to the adult 2 kHz tympanogram, including the broad notching around TPP (see Fig. 7). The cause of this differential pressure effect is unclear, but Margolis *et al.* (2003) suggested that decreased admittance at the negative tympanogram tail could be due to a partial "collapse" of the ear canal under negative pressure, resulting in reduced ear-canal volume, while positive ear-canal pressure could effectively increase the canal volume, resulting in increased admittance. Supporting this assumption is data from Holte *et al.* (1991) who obtained video recordings of the effects of pneumatic otoscopy pressure pulses (approximating $\pm$250–300 daPa) presented to infant ears. They reported mean changes in ear-canal diameter (relative to resting diameter) of 18.3% and −28.2% in 1–7 day old infants for positive- and negative-pressure pulses, respectively. These effects decreased with age and were negligible in infants of 51–66 days old. Holte *et al.* (1991) did report asymmetrical monotonically increasing tympanograms (from negative to positive decapascals) in some infants but did not specifically

investigate any relationships between tympanometric symmetry and degree of change in ear-canal diameter. It is also possible that the mass characteristics of the young infant middle ear (e.g., tympanic membrane and ossicles) are more easily manipulated with changes in static ear-canal pressure compared to older children and adults and that shifts in mass with negative pressure could possibly explain this sloping pattern in admittance tympanograms, especially since the effects of mass are manifested at higher frequencies and the positive/negative admittance differential seen in these young infants' becomes more pronounced as frequency increases.

C. Hypotheses of anatomical effects

Based on anatomical changes in the infant ear and knowledge regarding the effects of mass and stiffness on acoustic ear-canal measurements, inferences can be made as to what anatomical changes might account for some of the developmental differences observed in the present study. The large mean ER increase with age below 0.7 kHz seen in the infant groups, especially for the extreme pressure conditions, may be due to the stiffening of the compliant infant ear-canal wall. As age increased, the effect of static pressures for frequencies below 0.7 kHz decreased and was small in adults (see Fig. 1). While specific points in the timeline of infant ear-canal development have not been fully described, the decreasing effect of pressure with age could reflect the effects of a maturing (ossifying) ear canal, which becomes more resistant to changes in pressure with age (Saunders *et al.*, 1983). Qi *et al.* (2006) developed a nonlinear finite-element model of the newborn (22 days old) ear-canal wall to investigate the effects of static pressures on infant ear-canal volume. Their model predicts a nonlinear elastic behavior of ear-canal wall displacement in response to high static pressures. Based on their model, maximal displacement plotted as a function of positive and negative ear-canal pressures would result in an asymmetrical sigmoid shaped tympanometric function. $|Y|$ tympanogram data for frequencies above 2 kHz from the present study for all infants generally agree with the model prediction (Fig. 7). However, at frequencies below 2 kHz, the result would be a symmetrically shaped tympanometric function. It is possible that pressure induced changes in infant ear-canal volume, as modeled by Qi *et al.* (2006), could account for the sigmoid shaped $|Y|$ tympanograms seen at frequencies above 2 kHz in 4-week-old infants. A strict comparison between the model from Qi *et al.* (2006) and data from the present study is limited since the experimental data are not measurements of ear-canal wall effects alone and include tympanic membrane and middle-ear effects. Qi *et al.* (2006) acknowledge their model as a beginning step and that further work is needed to include the tympanic membrane and middle ear to better understand development of the complete infant conductive system.

Because of the substantial contribution of the ossicles to the mass effects in middle-ear function, it would follow that acoustic effects at high frequencies could be due to changes in the properties of the ossicular chain (Saunders *et al.*, 1983). Eby and Nadol (1986) reported that while the neonatal ossicles have obtained adult dimensions, they are histologically immature and composed mostly of cartilage. In addition, temporal bone data suggest that mass loading of the ossicles could be due to residual mesenchyme adhering to the ossicular chain. This residual embryonic tissue has been observed in infants as late as 5 months of age (Spector and Ge, 1981) and could influence the high-frequency ATF responses observed in young infants. This developmental timeline is consistent with data from the present study that show a differential pressure effect at frequencies above 2 kHz for 4-week-old infants, which lessens in magnitude for 12- and 24-week-old infants.

Age-related changes in the distance from the tympanic membrane to the footplate of the stapes have also been reported and are thought to be due to a change in orientation of the ossicular chain as the tympanic membrane rises from a nearly horizontal position, relative to the axis of the ear canal, to a more vertical adultlike position (Eby and Nadol, 1986). This change in ossicular chain orientation may also contribute to the markedly different high-frequency ER responses (above 2 kHz) in 4-week-old infants from the present study.

It is possible that the result of negative ear-canal pressures in 4-week-old infants may be to lessen the efficiency of the middle-ear system by functionally disarticulating the ossicular chain as the tympanic membrane is pulled outward into the ear canal. This effect could account for the increased ER with increasing negative pressure. The coupling of the ossicular chain is thought to be less rigid in young infants and may be affected more by the outward or perhaps more downward pull of negative ear-canal pressure on the infant tympanic membrane than in the adult (Saunders *et al.*, 1983). Conversely, positive ear-canal pressure may enhance ossicular coupling due to the ossicles being "pushed" more closely together, creating a more acoustically efficient middle ear, resulting in decreased ER for frequencies from 2 and 6 kHz (see Fig. 1).

V. CONCLUSIONS AND CLINICAL IMPLICATIONS

Data from the present study provide insight into the effects of ear-canal pressure on acoustic measurements in adult and infant ears. Both infant and adult ATF measurements were influenced by the introduction of ear-canal pressure. While ATF results for all infants at ambient and pressurized conditions were different from adult measurements, data from 4-week-old infants were the least adultlike, showing the greatest changes in ER across frequency. The different effects of positive and negative ear-canal pressures on wideband ATFs may be a reflection of functional differences in the infant ear canal and middle-ear system and could potentially serve as a marker for maturation of the infant ear canal and middle ear.

The range of frequencies from 0.8 to 2 kHz, where few significant differences for ATF tympanometry were found for the three infant age groups, may be a frequency region of interest in future studies due to its apparent maturational stability. In addition, Feeney and Sanford (2005) reported that robust acoustic reflex responses and good signal-to-noise ratios were found for young infants in this same general fre-

quency region. For frequencies between 2 and 4 kHz, positive pressures usually resulted in decreased ER and conversely increased $|Y|$, effectively creating a more efficient middle-ear transmission system perhaps due to a better coupling of the ossicles in the pressurized infant middle ear and thus more closely matching the impedance of the ear canal. This hypothesis could be tested by examining changes in otoacoustic emission (OAE) levels or auditory brainstem response (ABR) thresholds with changes in ear-canal pressures as a function of infant age.

Because of the large differences in wideband ER and $|Y|$ between 4- and 12-week-old infants, measurements taken at additional ages could be used to more closely define middle-ear maturation between these two ages. Results from the present study demonstrate that the effects of ear-canal pressure on ATF measurements are different for infants than adults. The maturational changes in middle-ear characteristics observed in the present study may explain much of the previously reported variability in 226 Hz immittance characteristics in infants. (Hunter and Margolis, 1992; Paradise *et al.*, 1976; Sprague *et al.*, 1985). A wideband approach to infant middle-ear assessment would provide a broader "window" with which to observe differences due to middle-ear development and possibly differences due to middle-ear pathology. While ambient ATF data from Keefe *et al.* (1993) show increasingly similar ER for infants above 6 months of age and adults, differences, especially between approximately 1.5 and 6 kHz, still exist, and ATF tympanometry studies for infants above 6 months of age may reveal additional developmental information. Ear-canal and middle-ear development might also be studied by conducting a longitudinal study to examine individual differences in the developmental time course.

Additional work is also needed to determine normative data sets for disordered ears of young infants since age-related differences in ambient and pressurized ER and admittance have been observed in a population with normal middle-ear status. Based on data from Keefe and Simmons (2003), which demonstrated that pressurized transmittance accurately predicts conductive hearing loss, investigations of pressurized transmittance in disordered infant ears would be an appropriate next step.

Results from the present study confirmed the findings of Holmer (2001[1]), showing increases in conductance with age. This information is beneficial in terms of improving our understanding of infant auditory development and for future studies of auditory development. Investigations in young infants employing objective measures of auditory sensitivity, such as ABR or OAEs should take into account the possible effects that external- and middle-ear developments may have on the results. Since universal newborn hearing screening programs employ one or both of these procedures (ABR/OAE), potential middle-ear influences are important factors to consider (Keefe *et al.*, 2003). In addition, a valid middle-ear test for neonates and young infants could potentially help identify false-positive hearing screening results, which are often attributed to middle-ear or ear-canal factors (Keefe *et al.*, 2000; Vander Werff *et al.*, 2007). It would also be of interest to obtain wideband tympanometric ATF measurements in a neonatal population to identify possible developmental differences in these very young infants. Since the age at which infants are seen in early hearing detection and identification programs varies, having an accurate age appropriate tool for identification of middle-ear status would prove useful over the course of rescreening and follow-up as well as provide more accurate interpretation of OAE and ABR test results.

Based on results from the present study, which showed developmental effects across the frequency range from 250 to 8000 Hz in infants, a wideband ATF approach to infant middle-ear assessment is advantageous. However, there are several challenges facing new and evolving wideband ATF measurements. First, since wideband ATF technology is still evolving, new equipment and techniques are being updated and changed. While it appears that results from different studies using different equipment/techniques have been consistent, methodological variables need to be considered as future investigations reference past experimental evidence.

From a clinical standpoint, if wideband ATF measurements are to be useful, data analysis and interpretation must be relatively straightforward. Traditional multifrequency tympanometry, which provides much more information than single-frequency admittance tympanometry, has not enjoyed widespread use. This possibly stems from an inability to quickly and efficiently interpret the results (Fowler and Shanks, 2002). While research findings suggest that valuable diagnostic information can be obtained using ambient and pressurized wideband ATF measurements, quick and efficient interpretation is important if it is to be used as an effective clinical tool. From the large amount of data that can be obtained with wideband ATF measurements, predictive variables should be identified in an effort to develop middle-ear tests with appropriate sensitivity and specificity (Keefe and Simmons, 2003). Future studies of normal and disordered middle-ears could address issues of improving sensitivity and specificity of these emerging middle-ear tests, while also achieving objectivity in test interpretation.

ACKNOWLEDGMENTS

The authors thank Douglas H. Keefe for providing software for the measurement of wideband acoustic transfer functions and for helpful discussions throughout the processes of data collection, analyses, and manuscript preparation. Data collection for this study took place at the Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, WA. This research and manuscript preparation was supported by grants from the National Institutes of Health (NIH), National Institute on Deafness and other Communication Disorders (NIDCD) to the University of Washington (F31-DC007296 and P30-DC004661), and Boys Town National Research Hospital (T32-DC000013).

[1]Holmer, N. M. (**2001**). "The effect of middle ear conductance on behavioral pure-tone thresholds of 11 and 24 week old infants and adults," Unpublished Master's thesis, University of Washington.

Alaerts, J., Luts, H., and Wouters, J. (**2007**). "Evaluation of middle ear

function in young children: Clinical guidelines for the use of 226- and 1,000-Hz tympanometry," Otol. Neurotol. **28**, 727–732.
Allen, J. A., Jeng, P. S., and Levitt, H. (**2005**). "Evaluation of human middle ear function via an acoustic power assessment," J. Rehabil. Res. Dev. **42**, 63–78.
Baldwin, M. (**2006**). "Choice of probe tone and classification of trace patterns in tympanometry undertaken in early infancy," Int. J. Audiol. **45**, 417–427.
Calandruccio, L., Fitzgerald, T. S., and Prieve, B. A. (**2006**). "Normative multifrequency tympanometry in infants and toddlers," J. Am. Acad. Audiol **17**, 470–480.
Eby, T. L., and Nadol, J. B., Jr. (**1986**). "Postnatal growth of the human temporal bone. Implications for cochlear implants in children," Ann. Otol. Rhinol. Laryngol. **95**, 356–364.
Feeney, M. P., Grant, I. L., and Marryott, L. P. (**2003**). "Wideband energy reflectance measurements in adults with middle-ear disorders," J. Speech Lang. Hear. Res. **46**, 901–911.
Feeney, M. P., and Keefe, D. H. (**1999**). "Acoustic reflex detection using wide-band acoustic reflectance, admittance, and power measurements," J. Speech Lang. Hear. Res. **42**, 1029–1041.
Feeney, M. P., and Keefe, D. H. (**2001**). "Estimating the acoustic reflex threshold from wideband measures of reflectance, admittance and power," Ear Hear. **22**, 316–332.
Feeney, M. P., Keefe, D. H., and Sanford, C. A. (**2004**). "Wideband reflectance measures of the ipsilateral acoustic stapedius reflex threshold," Ear Hear. **25**, 421–430.
Feeney, M. P., and Sanford, C. A. (**2005**). "Detection of the acoustic stapedius reflex in infants using wideband energy reflectance and admittance," J. Am. Acad. Audiol **16**, 278–290.
Fowler, C. G., and Shanks, J. E. (**2002**). "Tympanometry," in *Handbook of Clinical Audiology*, edited by J. Katz (Lippincott, New York), pp. 175–204.
Holte, L., Margolis, R. H., and Cavanaugh, R. M., Jr. (**1991**). "Developmental changes in multifrequency tympanograms," Audiology **30**, 1–24.
Hunter, L. L., and Margolis, R. H. (**1992**). "Multifrequency tympanometry: Current clinical application," Am. J. Audiol. **1**, 33–43.
Ikui, A., Sando, I., and Fujita, S. (**1997**). "Postnatal change in angle between the tympanic annulus and surrounding structures: Computer-aided three-dimensional reconstruction study," Ann. Otol. Rhinol. Laryngol. **106**, 33–36.
Keefe, D. H., and Abdala, C. (**2007**). "Theory of forward and reverse middle-ear transmission applied to otoacoustic emissions in infant and adult ears," J. Acoust. Soc. Am. **121**, 978–993.
Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2638.
Keefe, D. H., Bulen, J. C., Campbell, S. L., and Burns, E. M. (**1994**). "Pressure transfer function and absorption cross section from the diffuse field to human infant ear canal," J. Acoust. Soc. Am. **95**, 355–371.
Keefe, D. H., Folsom, R. C., Gorga, M. P., Vohr, B. R., Bulen, J. C., and Norton, S. J. (**2000**). "Identification of neonatal hearing impairment: ear-canal measurements of acoustic admittance and reflectance in neonates," Ear Hear. **21**, 443–461.
Keefe, D. H., and Levi, E. (**1996**). "Maturation of the middle and external ears: Acoustic power-based responses and reflectance tympanometry," Ear Hear. **17**, 361–373.
Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). "Method to measure acoustic impedance and reflection coefficient," J. Acoust. Soc. Am. **91**, 470–485.
Keefe, D. H., and Simmons, J. L. (**2003**). "Energy transmittance predicts conductive hearing loss in older children and adults," J. Acoust. Soc. Am. **114**, 3217–3238.
Keefe, D. H., Zhao, F., Neely, S. T., Gorga, M. P., and Vohr, B. R. (**2003**). "Ear-canal acoustic admittance and reflectance effects in human neonates. I. Predictions of otoacoustic emission and auditory brainstem responses," J. Acoust. Soc. Am. **113**, 389–406.
Kei, J., Allison-Levick, J., Dockray, J., Harrys, R., Kirkegard, C., Wong, J., Maurer, M., Hegarty, J., Young, J., and Tudehope, D. (**2003**). "High-frequency (1000 Hz) tympanometry in normal neonates," J. Am. Acad. Audiol **14**, 20–28.
Margolis, R. H., Bass-Ringdahl, S., Hanks, W. D., Holte, L., and Zapala, D. A. (**2003**). "Tympanometry in newborn infants—1 kHz norms," J. Am. Acad. Audiol **14**, 383–392.
Margolis, R. H., and Goycoolea, L. M. (**1993**). "Multifrequency tympanometry," Ear Hear. **14**, 408–413.
Margolis, R. H., and Heller, J. W. (**1987**). "Screening tympanometry: Criteria for medical referral," Audiology **26**, 197–208.
Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). "Wideband reflectance tympanometry in normal adults," J. Acoust. Soc. Am. **106**, 265–280.
Meyer, S. E., Jardine, C. A., and Deverson, W. (**1997**). "Developmental changes in tympanometry: A case study," Br. J. Audiol. **31**, 189–195.
Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.
Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., and Spetner, N. B. (**1988**). "Pure-tone sensitivity of human infants," J. Acoust. Soc. Am. **84**, 1316–1324.
Paradise, J. L., Smith, C. G., and Bluestone, C. D. (**1976**). "Tympanometric detection of middle ear effusion in infants and young children," Pediatrics **58**, 198–210.
Piskorski, P., Keefe, D. H., Simmons, J. L., and Gorga, M. P. (**1999**). "Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory," J. Acoust. Soc. Am. **105**, 1749–1764.
Qi, L., Liu, H., Lutfy, J., Funnell, W. R., and Daniel, S. J. (**2006**). "A nonlinear finite-element model of the newborn ear canal," J. Acoust. Soc. Am. **120**, 3789–3798.
Roush, J., Bryant, K., Mundy, M., Zeisel, S., and Roberts, J. (**1995**). "Developmental changes in static admittance and tympanometric width in infants and toddlers," J. Am. Acad. Audiol **6**, 334–338.
Ruah, C. B., Schachern, P. A., Zelterman, D., Paparella, M. M., and Yoon, T. H. (**1991**). "Age-related morphologic changes in the human tympanic membrane. A light and electron microscopic study," Arch. Otolaryngol. Head Neck Surg. **117**, 627–634.
Saunders, J. C., Kaltenback, J. A., and Relkin, E. M. (**1983**). "The structural and functional development of the outer and middle ear," in *Development of Auditory and Vestibular Systems*, edited by R. Romand and M. R. Romand (Academic, New York), pp. 3–25.
Spector, G. J., and Ge, X. X. (**1981**). "Development of the hypotympanum in the human fetus and neonate," Ann. Otol. Rhinol. Laryngol. Suppl. **90**, 1–20.
Sprague, B. H., Wiley, T. L., and Goldstein, R. (**1985**). "Tympanometric and acoustic-reflex studies in neonates," J. Speech Hear. Res. **28**, 265–272.
Stinson, M. R., Shaw, E. A., and Lawton, W. B. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.
Vander Werff, K. R., Prieve, B. A., and Georgantas, L. M. (**2007**). "Test-retest reliability of wideband reflectance measures in infants under screening and diagnostic test conditions," Ear Hear. **28**, 669–681.
Voss, S. E., and Allen, J. B. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.

Simulated head related transfer function of the phyllostomid bat *Phyllostomus discolor*

F. De Mey,[a] J. Reijniers, and H. Peremans
Department Milieu, Technologie en Technologiemanagement, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium

M. Otani
Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

U. Firzlaff
Department Biologie II, Ludwig-Maximilians-Universität München, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany

(Received 14 June 2007; revised 6 June 2008; accepted 16 July 2008)

This paper presents a calculation of the head related transfer function (HRTF) for the frontal hemisphere of the phyllostomid bat *Phyllostomus discolor* using an acoustic field simulation tool based on the boundary element method. From the calculated HRTF results, binaural interaural intensity differences (IIDs) are derived. The results: Region of highest sensitivity, HRTF patterns, and IID patterns are shown to be in good agreement with earlier experimental measurements on other specimens of the same bat species, i.e., the differences are within the interspecies variability range. Next, it is argued that the proposed simulation method offers distinct advantages over acoustic measurements on real bat specimens. To illustrate this, it is shown how computer manipulation of the virtual morphology model allows a more detailed comprehension of bat spatial hearing by investigating the effects of different head parts on the HRTF. From this analysis it is concluded that for this species the pinna has a significantly larger effect on the HRTF and IID patterns than the head itself. This conclusion argues in favor of a series of recent simulation studies based on pinna morphology only [R. Müller, J. Acoust. Soc. Am. **116**, 3701–3712 (2004); Müller *et al.*, *ibid* **119**, 4083–4092 (2006)].
 [DOI: 10.1121/1.2968703]

PACS number(s): 43.64.Ha, 43.80.Ka, 43.20.Fn [MCH] Pages: 2123–2132

I. INTRODUCTION

Bats rely on echolocation for navigation and foraging. In the search for auditory localization cues available to the bat during echolocation, the monaural head related transfer function (HRTF) and the binaural interaural intensity differences (IIDs) have been extensively studied. The monaural HRTF characterizes the frequency-dependent relation between the sound pressure of sources at various positions in the free field and the pressure received at the eardrum.[1] This relation is determined by absorption, diffraction, and reflection of the sound waves on the head and the external ears (pinnae) of the subject. Experiments on bats,[2–6] humans,[7,8] and other mammals[9–11] indicate that monaural HRTF information is necessary for elevation estimation of the sound source. However, Aytekin *et al.*[12] conclude that while monaural information contains mostly cues for elevation discrimination, especially in the vertical midplane, it contains ambiguous azimuthal information as well. Nevertheless, to acquire azimuthal information, bats are known to rely predominantly on binaural IID cues.[13] In addition, several studies show the importance of binaural IID cues for humans[14] as well as for other animals,[1,9,15] in particular for sound localization at high frequencies.

Because of their importance for spatial hearing, both the HRTFs and IID patterns have been measured for different bat species[2,4,5,12,16] as well as for other animals.[9–11,17] HRTF measurements are done by inserting a microphone in the ear canal at the eardrum's position and recording sound pulses sent from several locations in the free field around the head. These measurements are laborious and great care needs to be taken in the setup: An anechoic chamber is needed, microphone calibration is necessary, and special specimen preparation should be performed. To keep the duration of the measurements within reasonable bounds, the number of azimuth and elevation positions considered is usually limited.[1,8,17]

Recently, acoustic simulations based on detailed morphological models have been proposed to address the methodological problems inherent to the measurement process. Such simulations have been performed to obtain the HRTF for human heads using boundary element methods (BEM)[18–22] and finite difference methods,[23] and for bat pinnae using finite element methods (FEM).[24,25] Since BEM requires only the boundaries of the model to be discretized, while for FEM discretization of the whole computational do-

[a]Electronic mail: fons.demey@ua.ac.be

main is necessary,[26] the experiments in this paper are obtained with BEM. While this is numerically more challenging, computation will consume less memory.

While valid physical arguments can be made for using models of the pinnae only,[25] analyzing the sound field around a complete head, as proposed in this paper, has some distinct advantages: True binaural IID data can be obtained and the influence of the head and various head parts on the HRTF/IID can be studied. Hence, the main contributions of this paper are as follows:

(1) Comparing HRTF and IID data calculated with the simulation software to HRTF and IID data obtained in measurements allows the examination of the viability of the simulation approach as well as the validity of the assumptions made in this approach, for a complete head.
(2) The influence of different head parts is assessed, both on the monaural HRTF and the binaural IID patterns.

An interesting result from this last study, providing further evidence in favor of pinna only studies, is that from all the head parts, the pinnae have the single biggest effect on HRTF/IID.

After an explanation of the methods and techniques used in Sec. II, the comparison between the measured and the simulated bat HRTF and IID patterns is presented in Sec. III. The influence of the different head parts on the bat's spatial hearing is discussed in Sec. IV, followed by a conclusion in Sec. V.

II. METHODS AND TECHNIQUES

The following outlines the methods and techniques used throughout the calculation process: from model acquisition to result visualization.

A. Model acquisition

A model of the *Phyllostomus discolor* head was acquired by scanning a real head on a microCT machine. One of the pinnae could be scanned in its natural position. Unfortunately the other one got slightly bent due to transportation. However, the deformation was not so severe that it would significantly alter the transfer function of the head for a receiver in the good ear.

The head was fixated and preserved in formaldehyde and dried for approximately 20 min before scanning it with a Skyscan[27] 1076 microCT machine, at a resolution of 35.36 μm. The resulting shadow images were processed with cone-beam reconstruction software that accompanies the scanner, using a Feldkamp reconstruction algorithm. This resulted in a tomographic representation of the head, which was then used to create a stl-boundary (StereoLithography) representation, containing over 1 million boundary elements. Limitations to the simulation environment required an additional model simplification step (see Sec. II B).

MicroCT is well suited for capturing high density materials, but suffers from some problems, i.e., contrast problems and drying artifacts, when imaging soft, wet tissue. They can be avoided or at least minimized by limiting oneself to only scanning the ears, as done in Refs. 24 and 25. When scanning a complete head, these problems complicate both the reconstruction and model creation steps. First, the calcium present in the skull of the bat has a high x-ray absorption ratio compared to the pinna and head tissue. Second, the hairs present on the head are not easily captured and processed (see Fig. 1). Third, drying artifacts are more difficult to avoid due to increased scanning times ($\sim$2 h for the head used in this paper). Scanning a single pinna on the same machine would require 45 min. Finally, data sets tend to grow large (2 Gbytes for the head used in this paper).

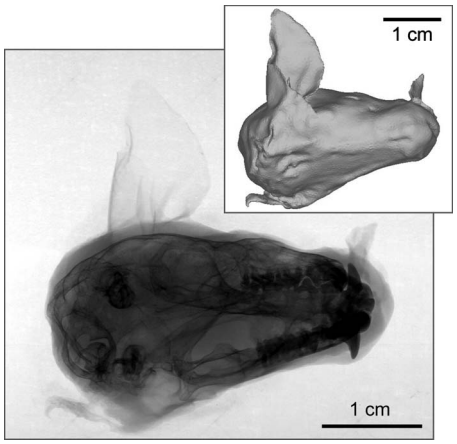


FIG. 1. Shadow image resulting from a microCT scan. Hairs are not visible. Inset: 3D model, positioned corresponding to shadow image.

B. Model simplification

The head model was simplified with standard mesh simplification algorithms to reduce the initial model's complexity ($\approx 10^6$ triangles) to a manageable level ($\approx$30 000 triangles). A maximal error of 0.5 mm was introduced, which was the size at which the simplification process produced a model with around 30 000 surface elements. Moreover, the maximal edge length was varied adaptively over the head: It was 0.85 mm on the head and on the contralateral pinna (the pinna in which we did not position a receiver), 0.68 mm in the vicinity of the ipsilateral pinna (the pinna in which we placed the receiver) and on the back side of that pinna, and 0.56 mm on the sound exposed side of the ipsilateral pinna. This way it was guaranteed that at the highest calculated frequency the simulation would sample the wave on the boundary with at least four points per wavelength away from the receiver, going to six points per wavelength when approaching the receiver. Numerical consistency checks show no significant discretization effects on the results when staying within these bounds.

C. Simulation

For the purposes of this paper it is required that a computer model of a complete bat head is analyzed. The size of a complete head ($35 \times 35 \times 35$ mm^3) is large compared to the wavelength at the maximum simulated frequency (f =95 kHz, λ=3.6 mm). Hence BEM was chosen as the most appropriate numerical technique. Nevertheless, the head models created, as explained earlier, using standard rules of thumb for BEM maximum element size, i.e., having bound-

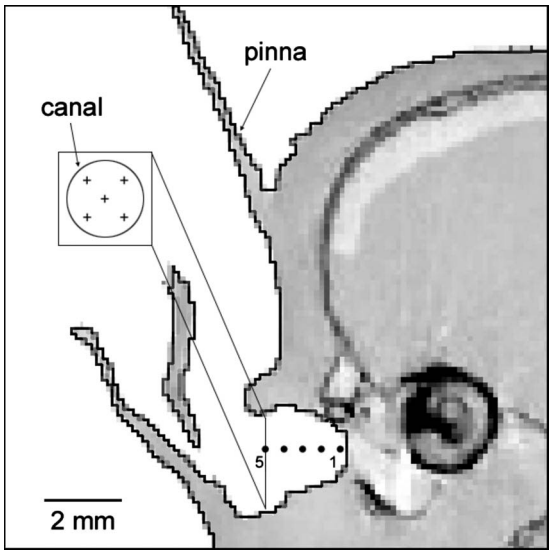


FIG. 2. MicroCT slice on which the left ear of the bat is shown. The outline of the boundary model is superimposed on it (black line). Five receivers at different locations along the ear canal are indicated with a dot. The canal approximately extends to the tympanum. Inset: Five points on a plane perpendicular to the canal composing a receiver.

ary elements with a maximum edge length of 1/4th to 1/6th of the wavelength at the maximum frequency,[18,20] still challenge present computing capabilities.

The simulation of the HRTF was performed in the frequency domain, using the BEM3D software developed to conduct HRTF calculations on models of the human head, described in Refs. 20 and 21. Given the expected complexity of head models, it was designed to be computationally more efficient than the traditional BEM methods.

The software allows one to introduce different acoustic impedance properties per boundary element. However, since, to the best of our knowledge, no impedance data exist for fur covered skin in the relevant frequency band, a rigid boundary model was used. This will affect the HRTF simulation results[19] and it is suggested in Sec. III that this approximation might be responsible for the reduced dynamic range of the calculated HRTFs when compared to the measured ones.

The simulation itself was performed on the CalcUA-computer, a cluster containing 256 AMD dual opteron nodes. Only the nodes with 8 Gbytes of RAM memory at their disposal, however, could be used. On those nodes, it was possible to simulate models with up to ~32 000 triangles.

Since the simulation results are compared to measured HRTFs of the *Phyllostomus discolor*,[2] the HRTF is simulated between 25 and 95 kHz, the same range used in these measurements. In Ref. 28 it is shown that this range contains the frequencies for which the *Phyllostomus discolor* is most sensitive. Moreover, most of the echolocation call's energy is located between 45 and 100 kHz,[2,28] indicating that the main echolocation frequency range is captured by the simulations.

The point receiver used in the simulations is actually a *virtual point receiver*. The sound pressure at this virtual receiver is calculated by averaging the sound pressure at five points on a plane through this receiver perpendicular to the canal (see Fig. 2, inset). In the remainder of this paper, every reference to receiver is a reference to this virtual receiver.

TABLE I. Correlation coefficients ($c_{A,B}$) between all pairwise combinations of the directional component of HRTFs obtained in different locations along the length of the ear canal, starting at the approximate position of the tympanum (point 1; see Fig. 2) and ending at the ear canal entrance (point 5).

Location	1 (tympanum)	2	3	4	5 (entrance)
1	1.0000	0.9843	0.9672	0.9596	0.9127
2	0.9843	1.0000	0.9827	0.9646	0.9176
3	0.9672	0.9827	1.0000	0.9820	0.9267
4	0.9596	0.9646	0.9820	1.0000	0.9520
5	0.9127	0.9176	0.9267	0.9520	1.0000

Contrary to the measurements with which the simulation results are compared, the receiver was not located at the tympanum itself. This was due to imaging problems: The tympanum can only be determined approximately on the data obtained from the microCT machine. Instead, the receiver was located at the entrance of the ear canal, because, as shown by several authors,[17,29] the directional properties of the human HRTF are not affected by such a choice. Since the maximum radius of the ear canal of the bat model has a diameter of 2.8 mm, which is smaller than the wavelength at the highest simulated frequency (3.6 mm), we assume that the dominant propagating modes in the canal will consist of plane waves as well. Hence, the directional properties of the bat HRTF are not expected to change in our simulation either. The correlation coefficients displayed in Table I do indeed validate this assumption by showing that the directional component of the HRTFs, as defined in Ref. 12, obtained from different locations in the ear canal (see Fig. 2) are all very similar. Note that by placing the receivers at the entrance of the ear canal, the frequency transfer characteristics of the ear canal are not modeled. However, those will only introduce interfrequency amplitude differences, not intrafrequency differences and thus only affect the direction-independent component of the HRTF.

The source positions for which the HRTF is calculated are located on the frontal hemisphere with center the receiver location and radius 1 m. The azimuth (elevation) samples for which data are obtained lay between −90° and 90° (between −82.5° and 82.5°), spaced at 2.5°. Points at the right of the midsagittal plane are indicated with negative azimuth values and points below the horizontal plane with negative elevation values. The head is tilted so that the plane in which the nostrils are located is perpendicular to the horizontal plane (cf. Refs. 2 and 16). A schematic of the setup is shown in Fig. 3.

D. Data processing

Unless stated otherwise, the HRTFs are normalized with regard to the highest amplitude over the range of frequencies and simulated source points, not over the simulated source points per frequency. In this way, the relative strength and thus the importance of different frequencies is preserved. Data calculated in the simulations are smoothed by averaging samples in a cube with edge length of five samples, after which it is converted to decibel values, assuming a noise floor below −80 dB.

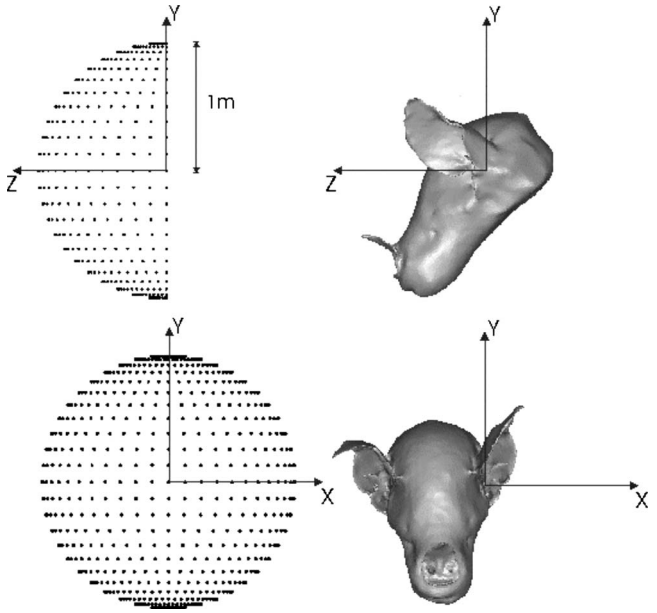


FIG. 3. Calculation setup: The bat is oriented so that the plane containing the nostrils is perpendicular to the horizon; the receiver is located in the origin of the coordinate system at the entrance of the ear canal. Sound sources are placed on the frontal hemisphere at a distance of 1 m, spaced at 2.5° for both azimuth and elevation. The head model is not shown to scale (see Fig. 1).

To visualize the HRTF for a specific frequency, Lambert's azimuthal equal area projection[9] is used. As noted by Müller,[24] the use of an equal area projection is desirable since one can discern the relative cross-sectional area of the lobes of the beam pattern on those projections. The trade-off is that angular distortions are introduced and lobe shape will not be representative. Lambert's projection, however, keeps those distortions to a theoretical minimum.[24]

When comparing different HRTFs, correlation coefficients are taken between the spatial information per frequency or between the entire frequency–space range. They are calculated with

$$c_{A,B,f} = \frac{\Sigma_{\theta,\varphi}(A_{\theta,\varphi,f} - \bar{A}_f)(B_{\theta,\varphi,f} - \bar{B}_f)}{\sqrt{\Sigma_{\theta,\varphi}(A_{\theta,\varphi,f} - \bar{A}_f)^2 \Sigma_{\theta,\varphi}(B_{\theta,\varphi,f} - \bar{B}_f)^2}}, \quad (1)$$

$$c_{A,B} = \frac{\Sigma_{\theta,\varphi,f}(A_{\theta,\varphi,f} - \bar{A})(B_{\theta,\varphi,f} - \bar{B})}{\sqrt{\Sigma_{\theta,\varphi,f}(A_{\theta,\varphi,f} - \bar{A})^2 \Sigma_{\theta,\varphi,f}(B_{\theta,\varphi,f} - \bar{B})^2}}, \quad (2)$$

in which A and B are HRTFs, θ designates azimuth, φ elevation, and f frequency. $\bar{A}_f$ represents the spatial average of the HRTF A at frequency f and $\bar{A}$ represents the spatial and frequency average of the HRTF A.

As mentioned earlier, the frequency transfer characteristics of the ear canal are not modeled in simulation. Moreover, they are different for each measured bat. These characteristics result in an unknown scaling of the HRTF patterns at different frequencies.

Comparing the HRTFs resulting from simulations or measurements on a per frequency basis, i.e., using $c_{A,B,f}$, is not affected by this phenomenon, since this correlation operation is invariant for scaling. Also, since the ear canal frequency response will be subtracted out by the method to construct IIDs proposed in Ref. 2 and followed in this paper, comparing IIDs using $c_{A,B}$ will not be affected by this scaling either.

However, the ear canal frequency transfer characteristic does affect the comparison of a measured HRTF with a simulated HRTF or another measured HRTF using $c_{A,B}$, since the frequency transfer characteristic of the canal is different for all those HRTFs. Hence, since HRTFs cannot be compared directly, the use of the directional component of the HRTF is proposed instead. In accordance with Ref. 12, this directional component, i.e., the direction dependent transfer function (DTF), is obtained by subtracting the spatial average of the HRTF at each frequency. Since, for a logarithmic representation, the use of DTFs effectively eliminates the canal frequency transfer characteristics, we will be using DTFs when comparing monaural information of bats with the correlation measure $c_{A,B}$.

III. COMPARISON WITH MEASURED DATA

To gain insight in the possibilities versus the limitations of using simulations to characterize the HRTF of bats, we compare a simulation of the HRTF of the *Phyllostomus discolor* to measurements done on three bats from the same species by Firzlaff *et al.* in Ref. 2.

In Ref. 2 the spatial hearing properties of three different specimens of the *Phyllostomus discolor* were characterized both monaurally (HRTF) as well as binaurally (IIDs). Throughout Sec. III, the following nomenclature is used: Bat 1, Bat 2, and Bat 3 are the specimens for which the measurement results are given in Ref. 2; Bat BEM is the specimen used to extract the boundary model on which the simulation was done.

For HRTFs, the region of highest sensitivity was characterized by three parameters per frequency: elevation and azimuth of the axis of highest sensitivity and area of the region of highest sensitivity. Whereas the two first parameters locate the region in space, the third gives an indication of the spatial tuning of the bat's hearing at a certain frequency. The axis of highest sensitivity per frequency was extracted by calculating the center of gravity for the −3 dB contour around the maximum gain (region of highest sensitivity) observed at that frequency. When this region consisted of two lobes, that lobe containing the maximum gain for that frequency was considered the region of highest sensitivity.

A. Comparing regions of highest sensitivity

Here, we investigate the extent to which the frequency-dependent behavior of the calculated region of highest sensitivity differs from that of the measured ones.[2]

First, it is checked whether the evolution of the axis of highest sensitivity over frequency for the simulated bat resembles those of the measured bats. To this end, the azimuthal and elevation position of this axis is calculated in accordance with Ref. 2.

As can be seen in Fig. 4(a), for the measured specimens the axis of maximum sensitivity shifts toward the vertical midplane with increasing frequency. Since the receiver was

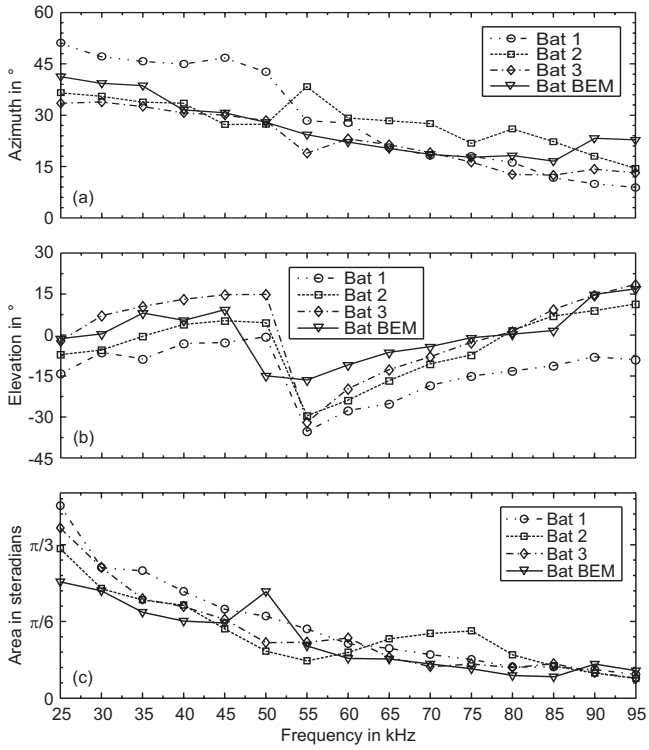


FIG. 4. (a) Azimuth of the axis of highest sensitivity for the measured bats (Bat 1, Bat 2, Bat 3; data from Ref. 2) and the simulated bat; (b) elevation of the axis of highest sensitivity; and (c) area of the region of highest sensitivity, expressed in steradians.

placed in the left pinna, the axis shifts from left to right. The axis of maximum sensitivity of the simulated HRTF exhibits the same behavior.

Figure 4(b) shows the elevation location of the axis of highest sensitivity for all the bats. In both the measured HRTFs as well as the simulated one, the same trends are displayed: an increase in elevation for the axis of highest sensitivity with increasing frequency up to ~50 kHz followed by a sudden drop around 50–55 kHz, in which the former region of maximum sensitivity splits into two lobes, one of which has a lower sensitivity. For higher frequencies a renewed increase can be observed. Thus, the region of highest sensitivity displays an elevational frequency scanning behavior.

The area of the region of highest sensitivity is expressed by its solid angle as done in Refs. 2 and 9. This is a measure for a two-dimensional angular span in three-dimensional space in which a complete hemisphere is indicated by 2π sr. As depicted in Fig. 4(c), all bat regions of highest sensitivity have a decreasing area, i.e., improved spatial tuning with increasing frequency.

From these results it is concluded that one of the main features of the HRTF, i.e., the region of highest sensitivity, shows very similar behavior for both the calculations and the measurements.

B. Comparing HRTF patterns

Although the region of highest sensitivity already captures a significant part of spatial hearing,[16] complete HRTF patterns still contain considerable extra information.[12] Hence, the following analyzes the correspondence between the complete calculated and measured HRTF patterns.

As can be seen from Fig. 5, visualizing the different HRTF patterns again shows good qualitative correspondence for all bats and all frequencies. In particular, because intraspecies variability has to be taken into account, as the specimen used in the simulation is not one of the three bats used to measure the HRTFs in Ref. 2.

A quantitative comparison of the HRTF patterns is performed using the previously defined correlation measures. To this end, correlations between all the specimens for which an HRTF was acquired are calculated. Table II displays the $c_{A,B}$ correlations of the HRTFs. From these results it can be seen that the simulated HRTF denoted by Bat BEM resembles Bat 2 and Bat 3 better than the HRTF of Bat 1 does, positioning the simulated bat well within the population of measured bats.

On a per frequency basis, using the $c_{A,B,f}$ correlation, the results depicted in Fig. 6(a) are produced. Again, Bat BEM shows good resemblances to the other bats, in the sense that

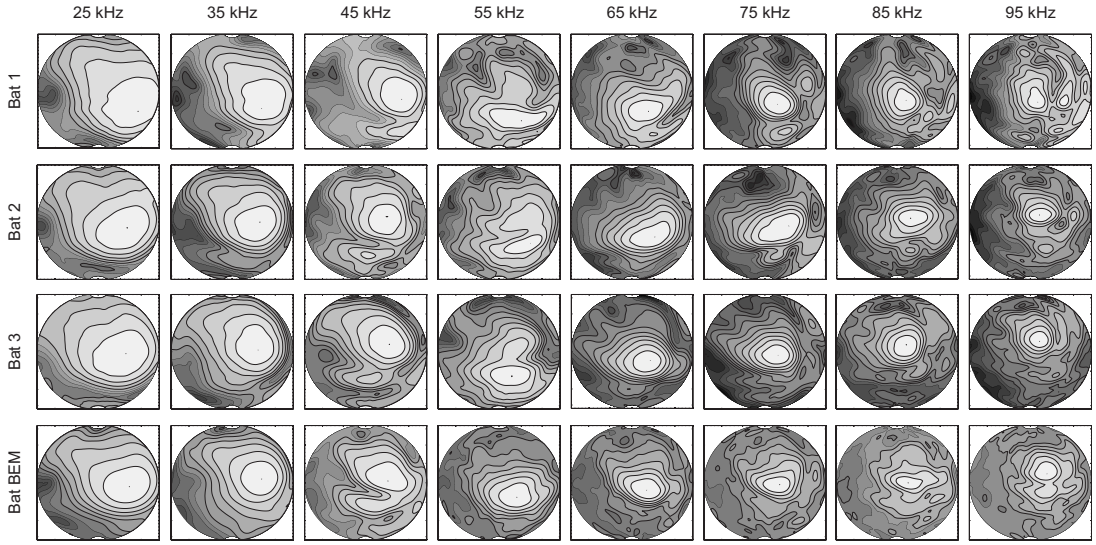


FIG. 5. HRTF patterns of the measured bats (Bat 1, Bat 2, Bat 3; data from Ref. 2) and of the simulated bat (Bat BEM) over the entire frequency range.

TABLE II. Correlation coefficients ($c_{A,B}$) between all pairwise combinations of HRTFs of the bats.

	Bat 1	Bat 2	Bat 3	Bat BEM
Bat 1	1.0000	0.7177	0.6443	0.6701
Bat 2	0.7177	1.0000	0.8083	0.8114
Bat 3	0.6443	0.8083	1.0000	0.7719
Bat BEM	0.6701	0.8114	0.7719	1.0000

it differs from the other bats as much as the other bats differ from each other. The only outlier in the correlation data occurs at 60 kHz: The correlation between Bat 1 and the other ones is especially bad there. Figure 7 displays the HRTF pattern for the bats at this frequency. Despite the quantitative differences between all four bats, we clearly see that Bat 2, Bat 3, and Bat BEM display the same qualitative HRTF characteristics: The region of highest sensitivity has comparable shape and is located in the lower left quadrant (azimuth: 0–90°; elevation 0 to −90°). Notches are located in the perifery. Bat 1, however, has a less pronounced peak. Moreover, it has additional sidelobe features, especially at higher elevation values.

From those correlation values as well as from the displayed HRTFs, it is again concluded that the simulated HRTF is that of a bat among bats, i.e., no significant differences between the simulated and the measured HRTFs are apparent. However, despite the good overall correspondence between the simulations and the measurements, as shown in Fig. 5, some discrepancies remain. In particular, the dynamic range of the HRTF patterns for Bats 1, 2, and 3 (~40 dB) is considerably larger than that of Bat BEM (~24 dB). We conjecture that this is one instance where the absence of realistic impedance information for the boundary affects the results. Indeed, noting that sound coming from the contralateral side is diffracted around the head before it enters the ipsilateral ear, it can be assumed that it will also be absorbed by the fur on the bat's head. For the simulated bat, though, with its acoustically rigid boundary, diffraction without absorption will determine the strength of the contralateral sound. The rigid boundary assumption affects the sound coming from the ipsilateral side much less, as these sound waves are directly captured by the pinna for which this assumption is much more appropriate.

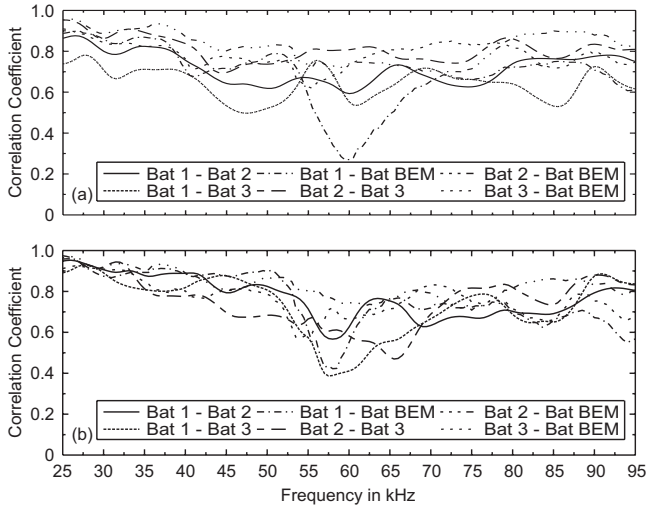


FIG. 6. (a) Correlation coefficients ($c_{A,B,f}$) between all pairwise combinations of the HRTFs of the bats; and (b) correlation coefficients ($c_{A,B,f}$) between all pairwise combinations of the IIDs of the bats.

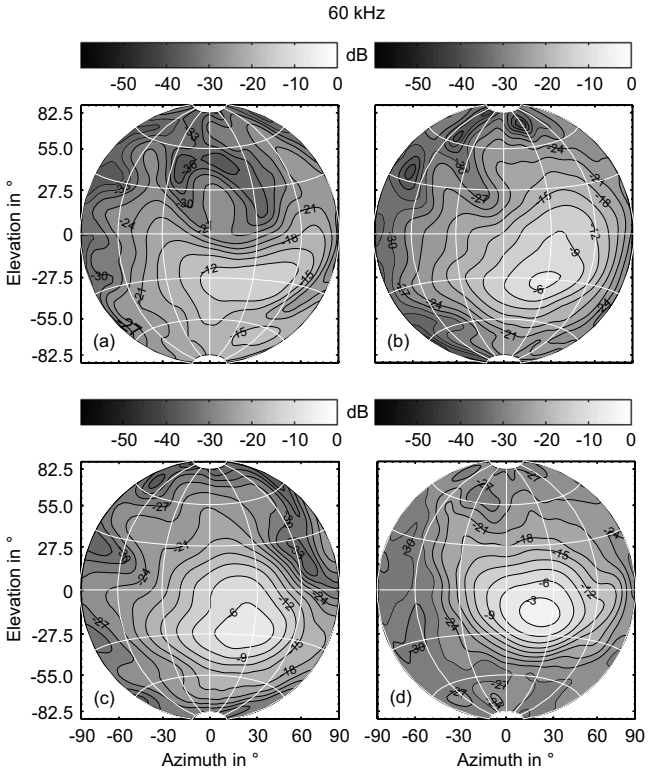


FIG. 7. HRTFs of the four bats at 60 kHz: (a) Bat 1; (b) Bat 2; (c) Bat 3; and (d) Bat BEM.

C. Comparing IID patterns

Most experimental approaches exploit the gross symmetry of the bat's head to obtain IID patterns: The measurements of the monaural HRTF are mirrored over the vertical midplane to obtain the response of the contralateral ear and then they are subtracted from each other.[2,4,16] Such an approach simplifies the experimental setup considerably, especially for small animals like bats, as it allows one to derive binaural information from monaural microphone measurements. For a simulation approach like the one proposed in this paper, such mirroring does not provide any advantages, because calculating the HRTF for both ears simultaneously is just a matter of specifying more receivers, resulting in a negligible extra computational load. However, in this paper, this multiple receiver capability is not taken advantage of, as the same symmetry property is employed to make the results directly comparable with the measurements.[2]

Table III and Fig. 6(b) show similar results as in Sec. III B: Bat BEM resembles Bat 2 and Bat 3 better than Bat 1 does. So, as with the previous two comparison criteria, binaural IID cues also indicate that the simulated bat behaves quite similarly to the measured ones.

TABLE III. Correlation coefficients ($c_{A,B}$) between all pairwise combinations of IIDs of the bats.

	Bat 1	Bat 2	Bat 3	Bat BEM
Bat 1	1.0000	0.7647	0.7555	0.7515
Bat 2	0.7647	1.0000	0.7521	0.8249
Bat 3	0.7555	0.7521	1.0000	0.7631
Bat BEM	0.7515	0.8249	0.7631	1.0000

IV. INFLUENCE OF DIFFERENT PARTS OF THE HEAD

Here we discuss an important application that is made possible by having access to simulation data: the assessment of contributions made by various head structures to the HRTF and the IID.

By using a virtual model instead of a real head, the morphology can be easily altered. Morphology changes that would involve irreversible surgery on a real head can be reversed in virtual reality, making possible the systematic study of the contributions made by various head parts and their functional relevance for the bat's echolocation. In Ref. 25 a similar study was performed to analyze the contributions of pinna features. In analogy, the head of *Phyllostomus discolor* is considered to consist of various acoustically conspicuous subcomponents: the head, the ipsilateral and contralateral pinna, and the noseleaf.

A. Influence of the subcomponents on the HRTF

The results shown in Fig. 8, i.e., the differences between the HRTF of a complete head and the HRTFs of the same head with various subcomponents missing, give a global impression of how the removal of the different subcomponents of the complete head affects the calculated HRTFs. First, it is noted that the differences between the HRTF of the complete head and the HRTFs with missing noseleaf and missing contralateral pinna are very small, indicating that those two subcomponents do not seem to interact much with the acoustic field around the head. On the other hand, the differences between the HRTF of the complete head and the HRTF of

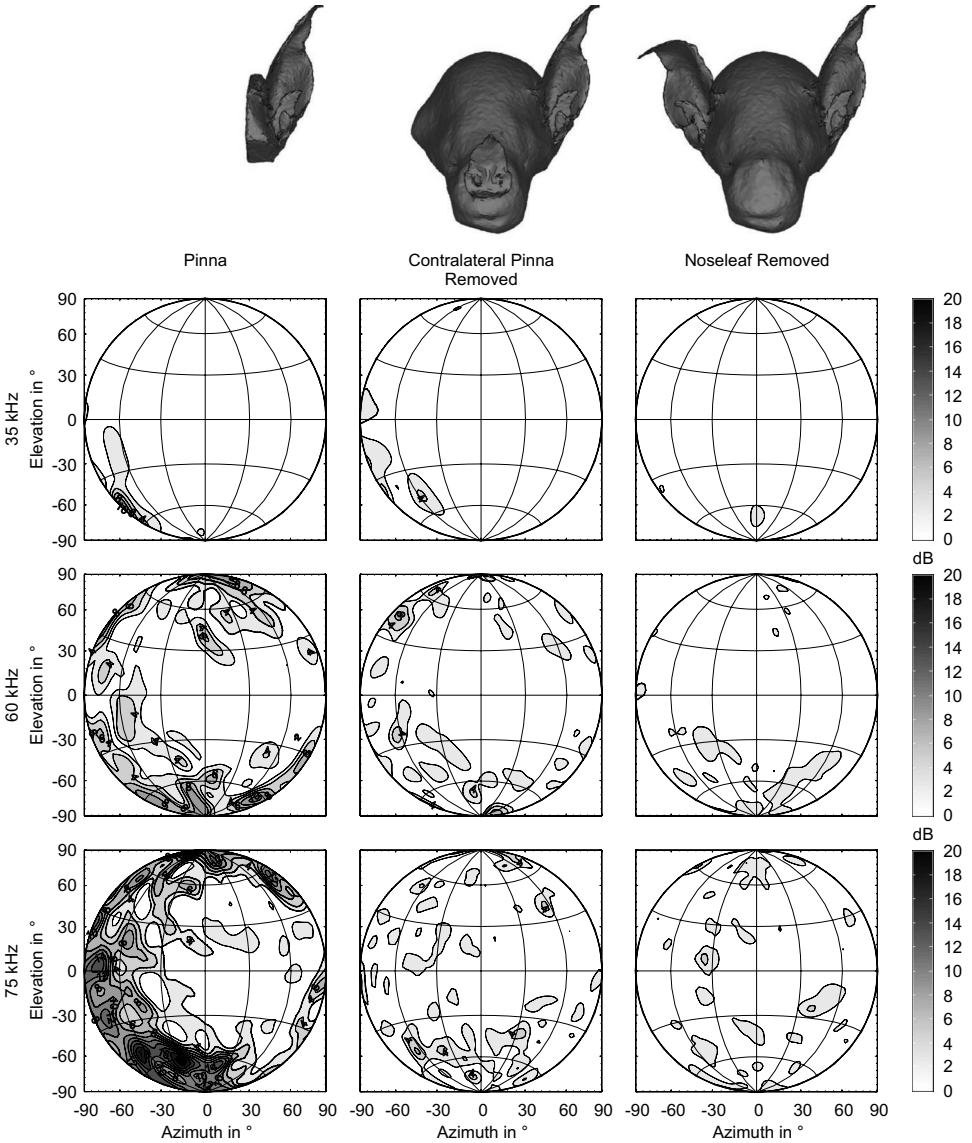


FIG. 8. Differences between the HRTF patterns of a complete head and, from left to right, the ipsilateral pinna only, the head without the contralateral pinna, and the head without the noseleaf.

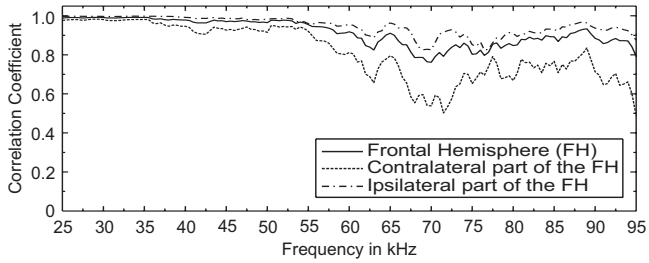


FIG. 9. Correlation coefficients ($c_{A,B,f}$) between the HRTF generated by the ipsilateral pinna only and the HRTF generated by the complete head, respectively, for the frontal hemisphere, its ipsilateral side, and its contralateral side.

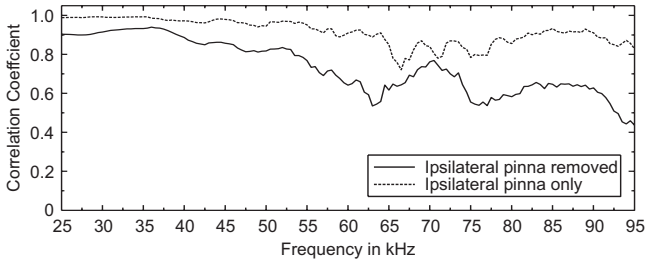


FIG. 10. Correlation coefficients ($c_{A,B,f}$) between the IIDs generated by a complete head, and respectively, a head from which the ipsilateral pinna is removed and a pinna only.

the pinna only model, while very small at low frequencies, grow larger at higher frequencies. In addition, it can be seen that the discrepancies between these two models are biased toward the contralateral half of the frontal hemisphere.

The very minor effect on the HRTF of the presence of the contralateral pinna as well as the noseleaf can be quantified by calculating the correlations between the HRTF of the complete head and the HRTFs after removal of the contralateral pinna ($c_{A,B,f}>0.9$) and removal of the noseleaf ($c_{A,B,f}>0.98$). Similarly, the good correspondence between the HRTF of the complete head and the HRTF of the pinna (positioned as if it were still attached to the head) can also be quantified (see Fig. 9) by calculating the correlation ($c_{A,B,f}>0.8$), providing evidence in favor of the pinna only approach employed in Refs. 24 and 25. However, splitting up the global correlation into separate correlation measures for the ipsilateral and the contralateral part of the frontal hemisphere reveals a consistently high ipsilateral correlation but a considerably lower contralateral correlation from 55 kHz on, indicating, as mentioned before, that the complete head and the pinna only models differ most in predicting the contralateral part of the HRTF for the higher frequencies.

This result can be understood by noting that the effect of the head itself is expected to be most pronounced for this case, i.e., contralateral sound at high frequencies.

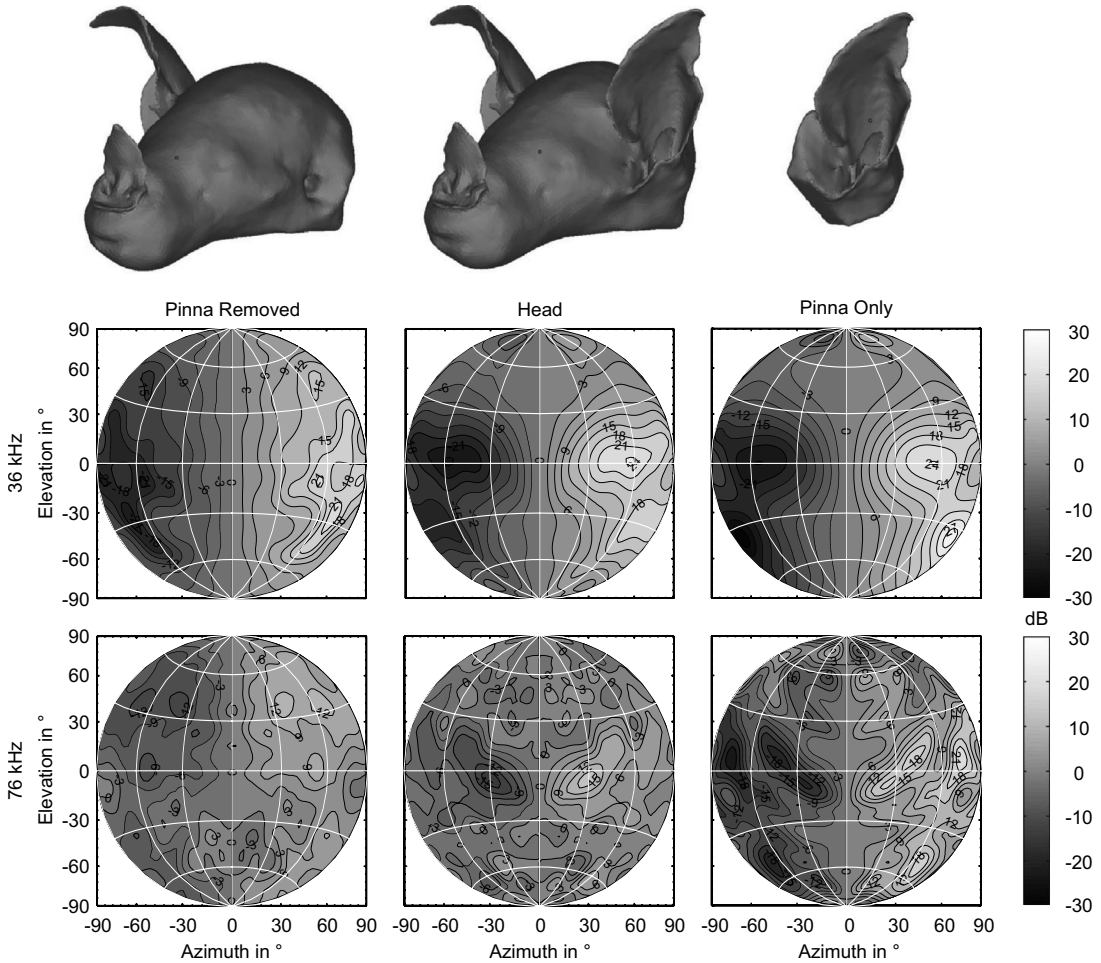


FIG. 11. IID patterns for the depicted models at 36 and 76 kHz.

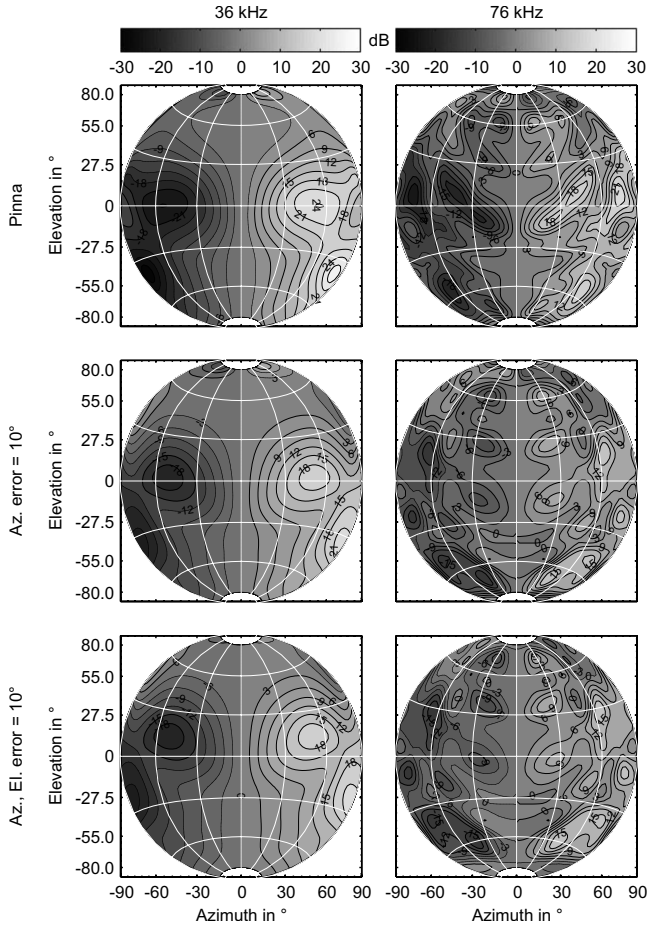


FIG. 12. IID patterns at 36 and 76 kHz for the pinna only model. Correct pinna orientation (top row); pinna orientation with azimuth error 10° (middle row); and pinna orientation with both azimuth and elevation error 10° (bottom row).

B. Influence of the subcomponents on the IID

Finally, the shadowing effect of the head on the IID pattern is appraised by cutting away the ipsilateral pinna in one model and comparing the resulting IID cues to those corresponding with a complete model and a pinna only model.

Inferring from the results in Figs. 10 and 11, it is concluded that the contribution of the ipsilateral pinna is a considerably more important factor in shaping the IID than the head itself is. Indeed, for all frequencies the IID patterns for the pinna only model resemble the complete head model much better than do the IID patterns for the head without pinna model, in accordance with the results in Sec. IV A, which already indicated the importance of the ipsilateral pinna for the HRTF.

As was the case for the HRTF comparison, the results in Fig. 11 also show that, while for lower frequencies the pinna only model and the complete head model result in very similar predictions for the IID patterns, discrepancies appear at the higher frequencies.

It should be noted that when replacing the complete head by a pinna only model, it is still advantageous to start from a model of the entire bat head and virtually cut away the head afterwards. In this way, correct position and orientation of the pinna with respect to the head is guaranteed. Figure 12 indicates that errors in orientation of the pinna will distort the IID patterns. In particular, azimuth errors will have the biggest effect on the IID pattern as they result in nonrigid deformations, whereas elevation errors will, to a first approximation, result in a rigid shift of the IID pattern over the elevation offset.

V. CONCLUSION

This paper calculates the HRTF and the IID patterns on a model of the complete head of a *Phyllostomus discolor*, with a fast BEM tool designed for HRTF calculation. The tools to acquire and to process the 3D model, yielding the HRTF, have only recently become available. The obtained results are in good agreement with acoustic measurements on specimens of the same species. In addition, it is noted that the remaining discrepancies between the model and the measurements are most pronounced in the contralateral half of the frontal hemisphere. It is conjectured that this part of the sound field is determined to a larger extent by diffraction around and absorption on (fur) the bat's head, and hence, that this is one area where knowledge of realistic impedance information would allow further improvement of the results.

Once validated by comparison with real measurements, this simulation approach is shown to have a large potential for analyzing "what-if" scenarios. In particular, simulations are used in this paper to determine the effect of the different subcomponents of the bat's head on the resulting HRTF and IID patterns. It is concluded from this analysis that, for this species, the ipsilateral pinna is the single most important subcomponent of the head. It is also shown that if the complete head model is replaced by a pinna only model for simulation, precise positioning and orienting of this pinna is essential.

The application of the simulation method paves the way for further experiments and research, e.g., the study of the interplay between the directivity of the noseleaf and the HRTF, i.e., sender–receiver interaction, and the calculation of the HRTF for dynamically deforming pinnae. So far, simulation studies have assumed the pinnae-head configuration to be fixed. Several species, however, actively position/deform their pinnae during the echolocation process.

ACKNOWLEDGMENTS

The authors would like to thank N. De Clerck, A. Postnov, and F. Lakiere, from the MCT group at the Universiteit Antwerpen and S. Ise at Kyoto University for his contributions to the BEM3D software used in this paper. This work was funded through the CILIA project (EU-FET).

[1]E. Young, J. Rice, and S. Tong, "Effects of pinna position on head-related transfer functions in the cat," J. Acoust. Soc. Am. **99**, 3064–3076 (1996).
[2]U. Firzlaff and G. Schuller, "Spectral directionality of the external ear of the lesser spear-nosed bat, phyllostomus discolor," Hear. Res. **181**, 27–39 (2003).
[3]B. Lawrence and J. Simmons, "Echolocation in bats: The external ear and perception of the vertical positions of targets," Science **218**, 481–483 (1982).
[4]T. Shimozawa, N. Suga, P. Hendler, and S. Scheutze, "Directional sensi-

tivity of echolocation system in bats producing frequency-modulated signals," J. Exp. Biol. **60**, 53–69 (1974).
[5]J. Wotton, T. Haresign, and J. Simmons, "Spatially dependent acoustic cues generated by the external ear of the big brown bat, eptesicus fuscus," J. Acoust. Soc. Am. **98**, 1423–1445 (1995).
[6]J. Wotton and J. Simmons, "Spectral cues and perception of the vertical position of targets by the big brown bat, eptesicus fuscus," J. Acoust. Soc. Am. **107**, 1034–1041 (2000).
[7]R. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagital plane," J. Acoust. Soc. Am. **61**, 1264–1269 (1977).
[8]R. Humanski and R. Butler, "The contribution of the near ear and far ear toward localization of sound in the sagittal plane," J. Acoust. Soc. Am. **83**, 2300–2310 (1988).
[9]S. Carlile, "The auditory periphery of the ferret. I. Directional response properties and the pattern of interaural level differences," J. Acoust. Soc. Am. **88**, 2180–2195 (1990).
[10]A. Musicant, J. Chan, and J. Hind, "Direction-dependent spectral properties of cat external ear: New data and cross species comparisons," J. Acoust. Soc. Am. **87**, 757–781 (1990).
[11]L. Xu and J. Middlebrooks, "Individual differences in external-ear transfer functions of cats," J. Acoust. Soc. Am. **107**, 1451–1459 (2000).
[12]M. Aytekin, E. Grassi, M. Sahota, and C. Moss, "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," J. Acoust. Soc. Am. **116**, 3594–3605 (2004).
[13]T. Park, B. Grothe, G. Pollak, G. Schuller, and U. Koch, "Neural delays shape selectivity to interaural intensity differences in the lateral superior olive," J. Neurosci. **16**, 6554–6566 (1996).
[14]J. Blauert, *Spatial Hearing* (MIT, Cambridge, 1997).
[15]M. Brainard, E. Knudsen, and S. Esterly, "Neural derivation of sound source location: Resolution of spatial ambiguities in binaural cues," J. Acoust. Soc. Am. **91**, 1015–1027 (1992).
[16]M. Obrist, M. Fenton, J. Eger, and P. Schegel, "What ears do for bats: A comparative study of pinna sound pressure transformation in chiroptera," J. Exp. Biol. **180**, 119–152 (1993).
[17]J. Middlebrooks, J. Makous, and D. Green, "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108 (1989).
[18]B. Katz, "Boundary element method calculation of individual head-related transfer function. I. rigid model calculation," J. Acoust. Soc. Am. **110**, 2240–2448 (2001).
[19]B. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparison to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455 (2001).
[20]M. Otani and S. Ise, "A fast calculation method of the head-related transfer functions for multiple source points based on the boundary element method," Acoust. Sci. & Tech. **24**, 259–266 (2003).
[21]M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," J. Acoust. Soc. Am. **119**, 2589–2598 (2006).
[22]T. Walsh, L. Demkowicz, and R. Charles, "Boundary element modeling of the external human auditory system," J. Acoust. Soc. Am. **115**, 1033–1043 (2004).
[23]T. Xiao and Q. Liu, "Finite difference computation of head-related transfer function for human hearing," J. Acoust. Soc. Am. **113**, 2434–2441 (2003).
[24]R. Müller, "A numerical study of the role of the tragus in the big brown bat," J. Acoust. Soc. Am. **116**, 3701–3712 (2004).
[25]R. Müller, H. Lu, S. Zhang, and H. Peremans, "A helical scanning pattern in the Chinese noctule, nyctalus plancyi," J. Acoust. Soc. Am. **119**, 4083–4092 (2006).
[26]Y. Naka, A. Oberai, and B. Shinn-Cunningham, "The finite element method with the Dirichlet-to-Neumann map for sound-hard rectangular rooms," in *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan, April 2004.
[27]http://www.skyscan.be/products/1076.html; last viewed 24 January 2008.
[28]K. H. Esser and A. Daucher, "Hearing in the fm-bat phyllostomus discolor: A behavioral audiogram," J. Comp. Physiol., A **178**, 779–785 (1996).
[29]D. Hammershoi and H. Moller, "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427 (1996).

Distortion product otoacoustic emission contralateral suppression functions obtained with ramped stimuli

David W. Purcell,[a] Blake E. Butler, Tracy J. Saunders, and Prudence Allen
National Centre for Audiology, University of Western Ontario, 1201 Western Road, London, Ontario, N6G 1H1, Canada

(Received 21 April 2008; revised 17 July 2008; accepted 18 July 2008)

The purpose of this research was to investigate the changes that occur in human distortion product otoacoustic emission (DPOAE) level functions over continuous frequency bands in response to activation of the medial olivocochlear (MOC) efferent system by contralateral broadband noise. DPOAEs were obtained using continuous upward ramps of the lower frequency tone (f_1) while the higher frequency tone (f_2) was fixed. These ramps were designed to change the stimulus frequency ratio f_2/f_1 over a fixed range for each fixed f_2 value of 2, 3, and 4 kHz. Contralateral noise was presented on alternating ramps and the DPOAEs with and without contralateral noise were averaged separately. Stimulus frequency ratios of 1.10 and 1.22, and noise levels of 60 and 50 dB sound pressure level (SPL) were employed. Changes in DPOAE level were generally suppression (a reduction in DPOAE magnitude), but enhancement was also observed. For most participants, changes were evident for much of the frequency ranges tested. Average absolute changes for 60 dB SPL noise were 0.95, 0.81, and 0.42 dB for the wider stimulus frequency ratios and f_2 of 2, 3, and 4 kHz, respectively. For the narrower ratio and 60 dB SPL noise, the changes were larger with average absolute changes of 1.33, 1.09, and 0.87 dB. For the narrower ratio and 50 dB SPL noise, the changes were 1.08, 0.78, and 0.55 dB with f_2 of 2, 3, and 4 kHz, respectively. DPOAE nulls were observed and a common response pattern was a shift of emission morphology to higher frequencies with contralateral acoustic stimulation. The method appears promising for relatively rapid evaluation of the MOC efferent system in humans and offers information complementary to measurement strategies that explore the effects of stimulus level.
 [DOI: 10.1121/1.2973192]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.66.Pn [BLM] Pages: 2133–2148

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) can rapidly and noninvasively deliver frequency specific information about activity of the outer hair cells within the human cochlea. This has found extensive use in infant hearing screening programs around the world. When measured in the presence of a contralateral acoustic stimulus, changes, typically suppression, in the DPOAE response can be observed. These effects are dependent upon activity in the efferent pathways of the lower brainstem, as well as the inner hair cells of the contralateral ear. The ability to measure these changes and thus evaluate not only outer hair cell function, but also efferent system function, would be a great addition to the use of emissions in clinical assessment. The goal of this study was to evaluate contralateral suppression functions over continuous ranges of DPOAE frequencies. These functions may offer information about efferent activity that is both useful and complementary to other techniques reported in the literature for observing DPOAE changes in response to contralateral acoustic stimulation.

The presence of contralateral acoustic stimulation has been shown to activate the medial olivocochlear (MOC) efferent system in human listeners, which elicits changes in outer hair cell activity. These changes can be observed as alterations in DPOAE magnitude in a manner that is typically suppressive (Moulin *et al.*, 1993; Chery-Croze *et al.*, 1993), but enhancement has also been observed (Bassim *et al.*, 2003; Lisowska *et al.*, 2002; Harrison *et al.*, 2008). A reduction of suppression and an increase in the prevalence of enhancement has been observed with aging in both humans and animal models (Kim *et al.*, 2002; Varghese *et al.*, 2005; Frisina *et al.*, 2007). Unfortunately, DPOAE changes with contralateral acoustic stimulation are generally small and a great deal of intersubject variability exists that makes the establishment of normative values challenging. Although it has been shown that DPOAEs resulting from lower level stimulus tones (~45 dB SPL) may be more easily suppressed (Williams and Brown, 1995), it is suggested that testing closer to the DPOAE response plateau of 65 to 70 dB SPL may be the best compromise to elicit OAEs from the majority of individuals (Bassim *et al.*, 2003).

When eliciting a DPOAE, the level of the emission peaks at the onset of the f_1 and f_2 stimulus tones, before falling to its steady-state value in a biphasic manner (Bassim *et al.*, 2003). This initial adaptation involves a steep decrease in emission amplitude, followed by a slower decrease until steady-state values are obtained. The amplitude of the emission can be perturbed from this steady state value by the addition of contralateral acoustic stimulation, and this suppression follows a similar biphasic time course both in ani-

[a]Author to whom correspondence should be addressed. Electronic-mail: purcelld@nca.uwo.ca

mal models and human listeners. In human listeners, the amplitude of this effect has been shown to be smaller than in animal models (Bassim *et al.*, 2003), but occurs on a similar time course to the initial adaptation described above (Kim *et al.*, 2001). Additionally, in the guinea pig, this effect has been shown to be simulated or blocked by intracochlear injection of the neurotransmitter acetylcholine or a cholinergic antagonist, respectively (Kujawa and Liberman, 2001).

Other studies have further evaluated the effects of contralateral stimulation on the time course and latency of the DPOAE. Silva and Ysunza (1998) hypothesized that the introduction of acoustic stimulation would have an effect on the steady-state latency of DPOAEs. However, their experiments using a phase gradient model to calculate latencies did not find significant changes. Some studies have employed a real-time analyzer to evaluate both amplitude and time course of the effect (James *et al.*, 2002; James *et al.*, 2005; Harrison *et al.*, 2008). James *et al.* (2005) determined that contralateral stimuli as short as 5 ms can elicit the suppressive effect in chinchilla models. The suppressive effect increases to a maximum over time and lasts for a minimum of 40 ms, indicating that a temporal integration function may be involved.

Contrary to animal models, work with humans has shown contralateral stimulation and ipsilateral suppressor tones close to the f_2 stimulus tone frequency have similar suppressive effects. This is seemingly due to the equal number of crossed and uncrossed medial olivocochlear fibers in the human MOC bundle that are responsible for the ipsilateral and contralateral reflex arcs, respectively (Guinan, 2006). However, the contralateral signal has been shown to alter the tuning curve of an ipsilateral suppressor, attenuating its peak and elevating its tails (Williams and Brown, 1995).

Although contralateral broadband noise signals have repeatedly been shown to have an overall suppressive effect on DPOAEs (e.g., Williams and Brown, 1997; Bassim *et al.*, 2003; Jacobson *et al.*, 2003; James *et al.*, 2005; Zhang *et al.*, 2007), the same cannot be said for contralateral tone pairs or narrowband noise signals (Lisowska *et al.*, 2002). Therefore, the clinical possibility of observing changes in DPOAE level with contralateral acoustic stimulation (as evidence of an intact efferent system) might be best pursued using broadband noise. Evaluation of the integrity of auditory efferent system function in a manner that was rapid, reliable, and not confounded by the need for behavioral responses is attractive. This would be useful for the assessment of clinical conditions for which lower brainstem function and efferent system function is questioned, such as may be the case in instances of lower brainstem lesions, auditory processing disorders, or peripheral neuropathies.

The purpose of this study was to evaluate the characteristics of the suppression function resulting from contralateral noise across relatively narrow but continuous frequency ranges. Previous studies have found that changes in DPOAE level vary with dips or fine structure of the emission (Kujawa and Liberman, 2001; Müller *et al.*, 2005; Wagner *et al.*, 2007). Fine structure refers to the details of the peaks and valleys in DPOAE level that occur as stimulus parameters are changed. These details are likely the result of the interference of two DPOAE sources and multiple internal reflections (Talmadge *et al.*, 1999; Dhar *et al.*, 2002; Dhar and Shaffer, 2004; Dhar and Abdala, 2007). There are some stimulus frequencies where the efferent system elicits relatively large changes in DPOAE level and other frequencies where the changes are so small that they approach the limit of our ability to detect them, given the perpetual presence of background noise in the ear canal. In a clinical setting, measurement at a single, or small set of frequencies, could potentially under-represent the presence of efferent activity in any given individual. Strategies have been developed to search for the best measurement frequencies and stimulus levels (Müller *et al.*, 2005; Wagner *et al.*, 2007; Zhang *et al.*, 2007), which require time and knowledgeable operators (e.g., a 4 min search plus 10 min of input-output level measurements; Wagner *et al.*, 2008). The approach pursued here was to evaluate the continuous contralateral suppression function (or more inclusively, change function) by ramping the DPOAE stimulus to produce a continuously ramping emission frequency. This approach has been applied by Williams and Brown (1997) with four participants and without averaging the DPOAE (when contralateral stimulation was present) to reduce background noise in the emission estimate. Extensive averaging to achieve excellent signal-to-noise ratios and a larger participant set were employed in the experiments described here. With continuous ranges of DPOAE frequencies, it is possible to observe changes with contralateral noise that can be evaluated across the band graphically or by measuring the area of the change function, the mean absolute value, and the maximum absolute value. This approach may offer complementary information about efferent activity relative to search strategies (i.e. wide rather than deep) and may help to avoid missing activity in individuals with unfortunate efferent response minima that happen to occur near standard stimulus frequencies.

Since fine structure has been identified as potentially rewarding for eliciting efferent effects, it was hypothesized that the change functions elicited here might be more robust if fine structure were encouraged in the DPOAE ranges being tested. The DPOAE is initiated near the characteristic place of the f_2 stimulus tone on the basilar membrane where it interacts most prevalently with the f_1 stimulus (Talmadge *et al.*, 1999; Shera and Guinan, 1999; Knight and Kemp, 2000). For the $f_{dp}=2f_1-f_2$ distortion product employed here, energy propagates both apically towards the characteristic place of f_{dp} and basally towards the middle ear and ear canal. A second reflection source at the f_{dp} characteristic place can also return energy basally (Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Shera and Guinan, 1999). The measured emission in the ear canal is the sum of the two sources and the relative mix can vary with stimulus parameters (Dhar *et al.*, 2005). Much of the observed fine structure is believed to be the result of interference between the two sources when they are of similar magnitude but near opposite phase. At very narrow stimulus frequency ratios, the measured DPOAEs are dominated by the f_{dp} place source, and the ratio $f_2/f_1=1.10$ has been identified as a ratio where transition can occur between dominant sources, although it varies by individual (Shera and Guinan, 1999; Knight and Kemp 2000; 2001;

Dhar *et al.*, 2005). Clinically, high emission levels are valued to achieve acceptable signal-to-noise ratios in the minimum measurement time. A ratio of $f_2/f_1=1.22$ or 1.20 has therefore been favored clinically because it has been shown to elicit near maximal DPOAEs for stimulus levels that are clinically useful (Harris *et al.*, 1989; Abdala, 1996). If good signal-to-noise ratios can be achieved at narrower ratios that facilitate fine structure, then the effects of efferent activity on DPOAE level might be enhanced. This might be because efferent effects on the delicate balance between sources near fine structure nulls are more observable in DPOAE level than modulation of a single dominant source or dual sources adding constructively (Wagner *et al.*, 2007). To encourage source interference, the approach chosen here was to ramp the stimulus f_1 while holding f_2 fixed such that a continuous range of f_2/f_1 near 1.10 could be used to elicit DPOAEs.

II. METHODS

A. Participants

Twenty-nine adults (23 female) varying in age from 19 to 33 years took part in three main experiments. Not all participated in every experiment and numbers for each experiment are given with the results below. Those who took part in multiple experiments generally returned on different dates (two people performed Experiments 1 and 2 on the same date, and one performed Experiments 2 and 3 on the same date). All participants showed normal otoscopic findings, hearing thresholds, distortion product otoacoustic emission (DPOAE) amplitude, and peak compensated static acoustic admittance (Peak Ytm). Normal hearing thresholds were defined as being $\leqslant$20 dB HL at 500, 1000, 2000, and 4000 Hz. DPOAE measurements at the frequencies used in the main experiments (2, 3, and 4 kHz) were recorded using a Starkey DP2000 measurement system with stimulus levels of $L_1=65$ and $L_2=55$ dB SPL for the lower and higher frequency tones, respectively. Emissions were considered to be normal when their levels exceeded the lower 5th percentile of normal hearing ears, as defined by Gorga *et al.* (1997). Only one additional participant was excluded from the study based on this criterion (all other participants had levels that exceeded the lower 10th percentile of normal hearing ears). Peak Ytm measurements were recorded using a GSI TympStar middle-ear analyzer with a 226 Hz probe tone. Peak Ytm values within the range of 0.30–1.50 mmho were considered to be normal (Roup *et al.*, 1998). Using this criterion, one additional participant was excluded from the study. Contralateral acoustic reflex thresholds (ARTs) using steady-state broadband noise were also recorded with the GSI TympStar. Participants with contralateral ARTs below 70 dB HL (eight beyond those mentioned above) were excluded in an effort to exclude middle ear muscle activity as a potential cause of DPOAE level changes.

The experiments were approved by the Health Sciences Research Ethics Board of the University of Western Ontario and written informed consent was obtained from each participant after the nature of the study was explained in accordance with the principles of the Declaration of Helsinki.

B. Stimulus generation and recording

Stimuli were generated and responses were recorded using a program developed in National Instrument's LabView language. Digital-to-analog and analog-to-digital conversions were performed at the rate of 32 000 samples/s with input resolution of 18 bits and output resolution of 16 bits by a National Instruments 6289 m-series acquisition card. Three output channels were used in total: two for the DPOAE stimulus tones f_2 and f_1, and a third for contralateral broadband noise. The levels of the analogue output signals were set using Tucker-Davis Technologies PA4 attenuators. Power amplification was provided by three channels on two Amcron D-75 amplifiers. The stimulus tones were generated by a pair of Etymotic ER2 earphones connected to the input ports of an Etymotic ER10B+ otoacoustic emission probe system. The contralateral noise was generated by a single Etymotic ER3A insert. A suitably sized probe tip and contralateral foam ear insert were chosen for each participant. The ER10B+ amplifier was set to the +40 dB gain condition, and the amplified signal was low-pass filtered prior to digitization with a cutoff of 10 kHz by a Tucker-Davis Technologies FT6-2 filter. Stimuli were calibrated using a Larson Davis system 824 sound level meter and a G.R.A.S. sound & Vibration IEC 126 2cc coupler. A single calibration was used for all participants and no in-the-canal adjustments were performed (Siegel, 1994; Whitehead *et al.*, 1995a).

Stimuli were presented as shown schematically in Fig. 1. A single stimulus sweep was composed of 16 epochs of 1.024 s. In the first eight epochs, the lower frequency tone f_1 was ramped from minimum to maximum value. This was then repeated in the second eight epochs with the addition of a broadband noise signal presented to the contralateral ear. Since the DPOAE occurs at $f_{dp}=2f_1-f_2$, the emission ramped at twice the rate of the f_1 tone. These rates were, however, slow relative to the capabilities of the ear, and no effect of up versus down ramps has been found in pilots for previous experiments (Purcell *et al.*, 2006). Multiple sweeps were presented sequentially in a given measurement. The f_2 stimulus tones were designed so that multiple sweeps could be concatenated without discontinuity. The f_1 tones and contralateral noise had linear fade-in and fade-out envelopes lasting 50 ms at their onset and offset. As described further below, the change in DPOAE level across frequency that is expected from efferent activity was obtained by subtracting the emission estimate obtained in the second half of the sweep from that obtained in the first. Since the absence and presence of contralateral noise alternated during each sweep, this helped to reduce potential order-of-test effects that might occur with slow changes in background noise or probe position (the DPOAE generators themselves were assumed not to change over the relatively short measurement time).

Participants reclined in a comfortable chair in a sound attenuated booth throughout the measurements and were encouraged to relax quietly or sleep. Emissions were recorded from a single ear, which was chosen at random in a balanced fashion for each experiment. The order of experimental conditions described below was also chosen at random and counter-balanced across participants and ears.

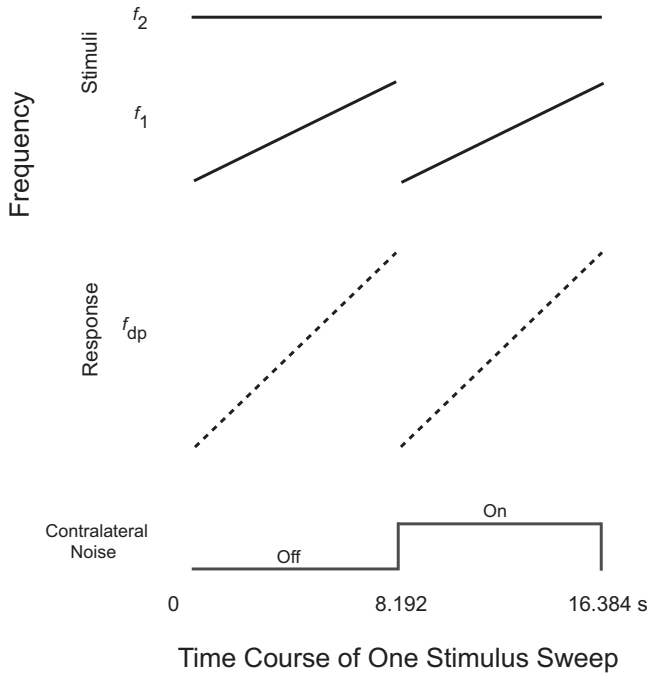


FIG. 1. A single sweep of the stimulus used to obtain an estimate of changes in DPOAE level with efferent activity. A sweep is the smallest unit of the stimulus that is repeated identically throughout a measurement for the purpose of averaging. In the experiments reported here, the DPOAE stimulus tone f_1 is also ramped upwards identically in each half of the sweep. The higher tone f_2 is always held constant and plays continuously without discontinuity from one sweep to the next. The DPOAE f_{dp} is determined by $2f_1-f_2$ and therefore also ramps upwards in frequency. At the bottom of the figure is shown the time course of the broadband noise that is provided to the ear contralateral to the emission measurement. It is off for the first half of the sweep and on for the second half of the sweep. Subtracting the DPOAE level estimated in the second half of the sweep from that estimated in the first obtains a difference function that shows the effect of efferent activity and low level background noise in the ipsilateral ear canal.

C. Experimental conditions

Three experiments were designed to investigate DPOAE contralateral suppression functions across small but continuous DPOAE frequency ranges. In all experiments, measurements were completed with the f_2 tone fixed at 4, 3, and 2 kHz. Measurements proceeded in that order for each of two conditions since no order confound was expected, but it was desirable to have participants as relaxed as possible for the 2 kHz measurement. The ear canal has more background noise at lower frequencies, which could contaminate lower frequency DPOAEs. Testing the highest frequency as the first measurement gave more opportunity for the participant to quiet before making the lower frequency measurements. Each DPOAE measurement consisted of 21 sweeps, lasting almost 6 min. With three f_2 frequencies and two conditions per experiment, the total DPOAE measurement time was approximately 35 min in each experiment. Stimulus frequency parameters used to elicit the DPOAE are shown in Table I. Stimulus levels were fixed at $L_1=60$ dB SPL for f_1 and $L_2=55$ dB SPL for f_2. These are closer in level than optimal for clinical measurements with the stimulus frequency ratio $f_2/f_1=1.22$ (Gaskill and Brown, 1990; Kummer *et al.*, 2000). The narrower ratio $f_2/f_1=1.10$ was used in all three experiments reported here with the exception of one condition in Experiment 3. Since the f_1 characteristic place was closer to the f_2 place when using the narrower ratio, L_1 was therefore set to be only 5 dB higher than L_2 for all measurements.

The purpose of the first experiment was to determine whether the continuous ramp technique could elicit changes in DPOAE level that were significantly different from a control condition. In the test condition, contralateral noise of 60 dB SPL was presented for the latter half of each sweep. For the control condition, no contralateral noise was presented. It was hypothesized that changes in DPOAE level would be significantly greater in the test condition due to the presence of contralateral noise and the activity of the efferent system. Small changes were expected in the control condition due to the influence of background noise.

The second experiment was designed to investigate the effect of contralateral noise level on changes in the DPOAE suppression function. One condition was identical to the test condition of Experiment 1, where the contralateral noise was 60 dB SPL. A second condition employed noise of 50 dB SPL. It was expected that DPOAE changes would be smaller for the lower level noise.

The purpose of the third experiment was to investigate the effect of stimulus frequency ratio f_2/f_1 on the DPOAE suppression function. Both conditions used 60 dB SPL contralateral noise as previously. One condition was identical to the test condition of Experiment 1, where the f_2/f_1 ratio was ramped around 1.10 (see Table I for values). A second condition employed a ratio ramped around the typical clinical value of 1.22. It was hypothesized that DPOAE changes would be greater with contralateral noise for the narrower stimulus frequency ratio. It was also expected that overall DPOAE levels would be smaller compared to the wider ratio.

D. Response analysis

Since the goal of these experiments was to observe changes in DPOAE level with efferent activity, it was desir-

TABLE I. Stimulus parameters used to elicit the DPOAE. The first row was used in all experiments, the second row was for the wider ratio condition of Experiment 3. Note that for higher f_2, that f_1 and f_{dp} must vary over a wider range to cover the same range of stimulus frequency ratios. Ranges were swept linearly in ascending fashion in 8.192 s.

Center ratio	Ratio range	$f_2=2000$ Hz		$f_2=3000$ Hz		$f_2=4000$ Hz	
f_2/f_1	f_2/f_1	f_1 range (Hz)	f_{dp} range (Hz)	f_1 range (Hz)	f_{dp} range (Hz)	f_1 range (Hz)	f_{dp} range (Hz)
1.10	1.08–1.12	1786–1852	1571–1703	2680–2781	2360–2562	3575–3707	3149–3413
1.22	1.20–1.25	1606–1672	1211–1343	2411–2512	1821–2023	3216–3347	2431–2694

able to minimize the impact of ipsilateral ear canal noise on the measured emissions. Therefore, prior to averaging, the epochs of microphone data from the ear canal were evaluated for the presence of noise events with higher than average sound pressure level. All epochs from a given measurement were first evaluated using a Fourier transform to obtain a noise metric based on the sound pressure level in a 4 Hz band located approximately 9 Hz below the lowest DPOAE frequency (for that measurement). Since the DPOAE frequency was changing, it was not possible to determine a noise estimate at immediately adjacent frequencies, as is commonly done with fixed frequency DPOAEs. Choosing a region below the emission frequencies is conservative because ear canal noise rises with decreasing frequency. Care was taken to avoid other DPOAE components such as $3f_1 - 2f_2$. The epoch noise metrics were then used to obtain an average sweep. Individual epochs were rejected from the average sweep if their noise metric exceeded the mean plus 2 standard deviations (s.d.), or if they exceeded −5 dB SPL. The first four epochs of a measurement were also rejected to allow time for the participant to settle. Typically, less than 5% of the available epochs were removed from the average sweep.

A software implemented Fourier analyzer was employed to extract an estimate of the DPOAE as it ramped in response frequency (Regan, 1989; Purcell *et al.*, 2004). Orthogonal reference sinusoids at the instantaneous DPOAE frequency were multiplied with the average sweep, and the complex analyzer outputs were low-pass filtered twice each using 400 ms rectangular filters. Due to the relatively slow sweep rate, no correction was applied for the small physiological delay between stimulus and response (a typical maximum of about 6 ms for the lowest f_2). After conversion to a magnitude estimate, a final rectangular smoothing filter of 100 ms was applied. To avoid end effects where the f_1 and f_{dp} ramps started and stopped, small frequency ranges of DPOAE data were excluded from further analysis at each end of the ramp. The excluded range was based on the analyzer filter lengths and ramp rates (f_2=2 kHz, 7.3 Hz excluded; f_2=3 kHz, 11.1 Hz excluded; f_2=4 kHz, 14.5 Hz excluded).

An estimate of the final noise in a measurement was obtained using a Fourier transform after averaging the two halves of the sweep. The noise estimate used mean amplitudes from 40 bins centered at approximately 9 Hz below the lowest DPOAE frequency as introduced above (here each bin is 0.122 Hz, for a range of 4.9 Hz). Since the Fourier analyzer passes more noise than the Fourier transform, the noise estimate was scaled by a factor of 5.25 (determined using simulated noise; Purcell *et al.* 2004; Purcell *et al.* 2006). At each frequency, DPOAE magnitude was evaluated to determine whether it was statistically different from the noise using an F-ratio at $p<0.01$ [$F(2,80)=4.88$].

With the goal of observing true changes in DPOAE level with efferent activity, further restrictions were applied to include only data expected to be least contaminated by background noise. This was implemented using the following additional criteria. For all measurements, except the condition using the wider clinical ratio $f_2/f_1=1.22$ in Experiment 3, the estimated noise could be no larger than −20 dB SPL. Since wider ratio DPOAEs tended to be larger (as well as the noise present at the concomitant lower DPOAE frequencies), this criterion was set at −15 dB SPL for that condition to prevent unnecessary data loss and yet maintain an excellent signal-to-noise ratio. A second criterion was applied for all measurements in Experiments 1 and 2, at f_2 of 3 and 4 kHz where noise tended to be lower. Data were excluded if the noise estimates between conditions (for a given frequency) differed by more than 5 dB and one of the estimates exceeded −25 dB SPL. These cross-condition restrictions were not applied in Experiment 3 since the noise estimates were from different frequency bands because of the ratio change.

To simply describe changes in DPOAE level with efferent activity across the measured frequency ranges, several metrics were developed to convey the observed differences. The average change across a frequency range is appropriate when only suppression or only enhancement is observed. However, many participants demonstrated both suppression and enhancement within the measured ranges. Therefore, the average absolute change is a more appropriate metric. The drawback of this choice is, however, that noise will lead to small positive values in the absence of true DPOAE changes. Related to the average absolute change is the area under the curve that results from subtracting DPOAE levels with the suppressor from those obtained without. The units of this area are unusual: dB s or dB Hz (change in dB times ramp duration or ramp range). Since the DPOAE frequency ranges vary for different values of f_2, using dB s allows comparisons between f_2 measurements. The single factor relating average absolute change in dB and area in dB s is 8.192 for the stimulus parameters employed in the experiments reported here. In the tables presented below, both will be given. Analyses of variance (ANOVAs) were performed for each experiment with data from all individuals who passed the noise criteria for every f_2 frequency. When assumptions regarding homogeneity of variance were violated, the reported statistics were adjusted using Greenhouse-Geisser corrections. A value of $p<0.05$ was used to indicate a significant effect. Due to the linear relationship between area and mean absolute change, the statistical results were identical and will be presented together for these two metrics.

If there are prominent features in the difference curve across a DPOAE frequency range, then there may be points with absolute differences substantially larger than given by the mean. The maximum absolute change is therefore also reported. All of these metrics only employ data from frequencies where the DPOAEs were statistically larger than the noise at $p<0.01$ for both halves of the sweep (e.g., with and without contralateral noise). This is a conservative approach and will not capture large emission changes where the DPOAE is significant without contralateral noise and then plunges into the noise floor with efferent activity (or vice versa).

III. RESULTS

A. Experiment 1

Eighteen people participated in Experiment 1. Some data were excluded from further analysis because they failed

TABLE II. Average results for Experiment 1 with standard deviations in parentheses. These values represent the difference in DPOAE level between the two halves of the sweep. In the test condition, contralateral noise was present for the second half of the sweep. The given metrics are the area of the difference function in dB s, the average absolute difference in dB, and the maximum absolute difference in dB. The number of participant's data included at each frequency is also given in the last row; data for some participants were excluded at one frequency due to failing the noise floor criteria given in the methods section for the average sweep.

	f_2=2000 Hz			f_2=3000 Hz			f_2=4000 Hz		
Condition	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)
Test: alternating noise	11.67 (6.88)	1.42 (0.84)	4.32 (2.21)	8.70 (4.19)	1.06 (0.51)	3.03 (1.62)	6.85 (3.28)	0.84 (0.40)	3.13 (2.83)
Control: no noise	2.86 (1.35)	0.35 (0.16)	1.53 (1.40)	3.52 (2.87)	0.43 (0.35)	2.23 (2.54)	2.41 (1.72)	0.29 (0.21)	1.96 (1.68)
N	15			17			17		

the noise criteria outlined above. For f_2=2 kHz, measurements from three people were too noisy to be included. At each f_2 of 3 and 4 kHz, data were also removed for one person. Average results for the remaining individuals at each f_2 frequency are given in Table II. The differences in DPOAE level between the first and second halves of the sweep are indicated with the area under the difference function, the mean absolute difference, and the maximum absolute difference as introduced above. Figure 2 also shows the mean absolute differences with ±1 s.d. error bars. Table II and Fig. 2 include data for 15, 17, and 17 people at f_2 values of 2, 3, and 4 kHz, respectively. Individual results for the test and control conditions (i.e., with and without contralateral noise in the second half of each sweep) are shown in Fig. 3 for two participants at f_2=2 kHz. The first row plots DPOAE level, and the second row shows the difference curve between the two halves of the average sweep. While the first half of the average test sweep (no contralateral noise) and both halves of the control sweeps are coincident in Fig. 3, there were individuals whose test and control curves were

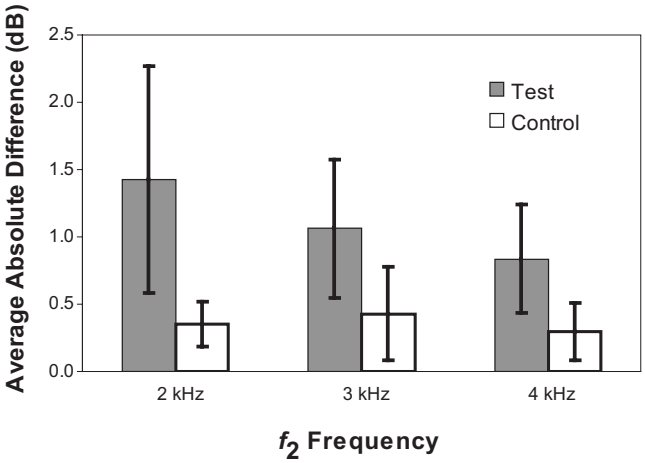


FIG. 2. This bar chart shows the average absolute difference between the two sweep halves in dB for a test condition and a control condition. In the test condition, contralateral noise was presented for half the sweep at 60 dB SPL and efferent activity would be expected to change DPOAE level in the measurement ear. For the control condition, the contralateral ear never received noise, so changes in the DPOAE would be expected to be small and due to ear canal noise in the measurement ear. Error bars show plus/minus one standard deviation.

offset a small amount (typically less than 1 or 2 dB; this occurred between conditions for some participants in Experiments 1 and 2). This was likely due to minor changes in probe position during the almost 20 min delay between test and control measurements at a given f_2 frequency (condition order was counter balanced). This would not affect the calculation of difference functions since they are obtained within condition between sweep halves (i.e., slow changes over time are likely to affect both halves of the sweep equally). Figure 4 shows individual test condition results for a further three participants across f_2 frequencies. These plots show examples of some of the configurations that were observed in response to efferent activity.

ANOVAs were performed with data from the 13 participants who passed the noise criteria at every f_2 frequency. The factors were condition (test with alternating contralateral noise versus control with no contralateral noise) and f_2 frequency (2, 3, and 4 kHz). For area under the difference curve and mean absolute difference metrics, there was a significant main effect of condition [$F(1,12)=31.5$, $p<0.001$], where efferent activity due to contralateral noise caused a larger change in the DPOAE than did background noise in the ipsilateral test ear (a single ANOVA result is reported for these two metrics because of their linear relation, which leads to identical statistical results). This is visually apparent in Fig. 2 for mean absolute difference where the bars are higher in the test condition at each f_2 frequency. There was no main effect of f_2, but there was a significant interaction between condition and f_2 [$F(1.7,20.4)=6.9$, $p<0.01$]. Post-hoc testing showed significant differences for the test condition between f_2 of 2 and 4 kHz [$F(1,12)=9.4$, $p<0.01$], and between 3 and 4 kHz [$F(1,12)=5.2$, $p<0.05$]. In each case, DPOAEs elicited with the lower f_2 experienced larger changes. There were no significant differences between f_2 of 2 and 3 kHz in the test condition, and no significant differences between any frequencies for the control condition. Similar results were obtained for the ANOVA performed on the maximum absolute difference metric. There was again a significant main effect of condition [$F(1,12)=11.1$, $p<0.01$] where values were larger in the test condition which employed contralateral noise than in the control condition that did not. There was also no main effect of f_2, but there was an interaction of condition with f_2 frequency [$F(1.9,23)=5.0$, $p<0.05$]. Post-

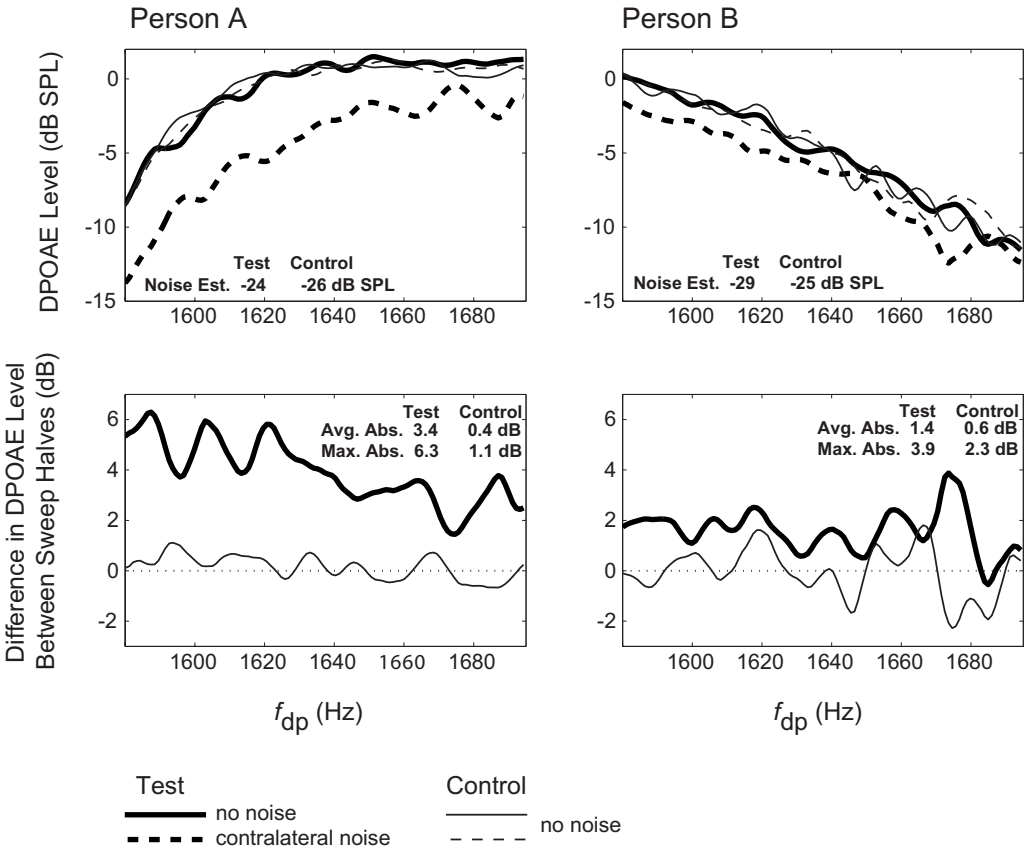


FIG. 3. Example data from two individuals are shown in each column for the f_2 frequency of 2 kHz. The top row shows DPOAE level in dB SPL vs emission frequency, where the test condition uses thicker lines, and the control condition uses thinner lines. The solid lines are for the first half of the sweep, where contralateral noise was never presented. The dashed lines are for the second half of the sweep, where contralateral noise was presented in the test condition. The values given in the lower left corners of the top row are the estimated noise floors in dB SPL for the test and control conditions. The lower row shows the difference in DPOAE level obtained when values from the second half of the sweep were subtracted from those obtained in the first half. The test condition is shown with thicker lines, and the control condition is displayed with thinner lines. Positive values indicate suppression of the DPOAE, whereas negative values show enhancement. The values given in the top right corners of the bottom row are metrics describing the changes observed between sweep halves. They are the average absolute change and the maximum absolute change in dB. Vertical scales are consistent within rows so that data can be compared across the two individuals. Person A showed more robust suppression and person B had a more variable control condition.

hoc testing performed to probe this interaction did not, however, show any significant differences between f_2 frequencies for the maximum absolute difference metric.

B. Experiment 2

Eighteen individuals participated in Experiment 2, 11 of whom had also taken part in Experiment 1. Data from some individuals at a single frequency were removed as in Experiment 1 due to failing the noise criteria. Data from three individuals were excluded for f_2=2 kHz (one of these individuals was previously excluded at 2 kHz in Experiment 1, and one was previously excluded at 4 kHz in Experiment 1), and one each was excluded from 3 and 4 kHz. Group average results for the difference curves at each f_2 frequency and experimental condition are given in Table III. Figure 5 shows example individual difference curves for one participant.

Repeated measures ANOVAs were calculated with data from the 13 individuals who had results for every f_2 frequency. Factors were level (60 versus 50 dB SPL contralateral noise) and f_2 frequency (2, 3, and 4 kHz). A significant main effect of level was found for area under the difference curve and mean absolute difference metrics [$F(1,12)=44.4$, $p<0.001$], where changes were larger for the 60 dB SPL contralateral noise. There was also a significant main effect of f_2 frequency [$F(1.8,21.9)=5.5$, $p<0.05$] and post-hoc testing showed efferent effects on DPOAE level were larger at f_2=2 kHz than at 4 kHz [$F(1,12)=13.7$, $p<0.01$]. There were no significant differences at $p<0.05$ between f_2 of 2 and 3 kHz, or 3 and 4 kHz. For the maximum absolute change metric, a significant effect was found for level [$F(1,12)=22.6$, $p<0.001$], where again 60 dB SPL produced larger changes in DPOAE level. There was no significant effect of f_2 or interactions.

C. Experiment 3

The final experiment included 22 participants, 10 of whom took part in both prior experiments, and eight who only had previously been in one experiment (there were four new participants with no prior measurements). Again there were some individuals whose data at single f_2 frequencies did not meet the noise criteria. Three sets of data were removed from analysis at f_2=2 kHz (one of these individuals was previously excluded at 2 kHz in Experiment 1), and two sets were removed at 3 kHz. Average results for the remaining participants are given in Table IV. Figure 6 shows example results for a single individual across f_2 frequencies. In

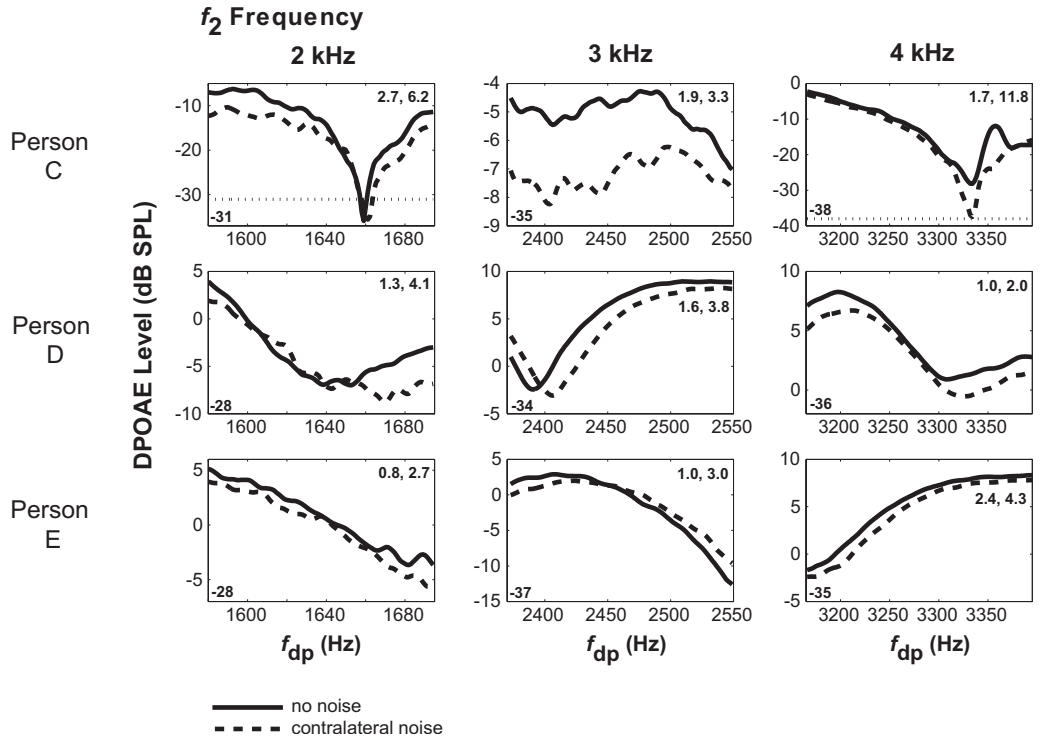


FIG. 4. Example DPOAE level data from the test condition for another three participants at every f_2 frequency tested. Each row shows data for a single person across f_2 values. The vertical axes are DPOAE level in dB SPL, and the horizontal axes are emission frequency in Hz. Note that the vertical scales vary between plots to show the most detail by using the smallest reasonable range. Solid lines are for the first half of the sweep with no contralateral noise, and dashed lines are for the second half of the sweep with contralateral noise. The majority of the curves show a reduction of DPOAE level with efferent activity, however, enhancement was also observed (e.g., Person E, f_2=3 kHz). Values given in the lower left corners of each plot are the estimated noise floors in dB SPL. For the two plots in the top row with nulls reaching the noise floor, the noise floor is shown as a horizontal dotted line. The two values given in the top right corner of each plot are the mean absolute difference and maximum absolute difference in dB, which were obtained by subtracting the two halves of the sweep.

this figure, changes in DPOAE level are particularly evident near the minimum, or dip, in emission level, for f_2=3 kHz with the narrower stimulus frequency ratio (middle row, right hand side panel). There were individual examples where large changes occurred near prominent emission features for the wider ratio as well. The same (right) ear of this person was tested on a different date in Experiment 1 and those data are shown in Fig. 4 (person D, middle row). Despite the probe placement being necessarily slightly different, the results for the narrower ratio condition shown in Fig. 6 are similar to those in Fig. 4, where the stimulus was identical for the test condition of Experiment 1.

Repeated measures ANOVAs were calculated for the 17 individuals whose data passed the noise criteria at all three f_2 frequencies. Factors were stimulus frequency ratio (narrower f_2/f_1=1.10 ratio versus wider ratio 1.22) and f_2 frequency (2, 3, and 4 kHz). A significant main effect of ratio was found $[F(1,16)=12.7,\ p<0.005]$ for the area and average

TABLE III. Average results for Experiment 2 with standard deviations in parentheses. These values represent the difference in DPOAE level between the two halves of the sweep, where contralateral noise was present for the second half of the sweep. The given metrics are the area of the difference function in dB s, the average absolute difference in dB, and the maximum absolute difference in dB. The number of participant's data included at each frequency is also given in the last row; data for some participants was excluded at one frequency due to failing the noise floor criteria given in the methods section for the average sweep.

	f_2=2000 Hz			f_2=3000 Hz			f_2=4000 Hz		
Condition	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)
Noise 60 dB SPL	11.60 (7.02)	1.42 (0.86)	3.85 (2.33)	8.32 (3.46)	1.02 (0.42)	3.41 (2.34)	5.64 (2.34)	0.69 (0.29)	2.25 (1.30)
Noise 50 dB SPL	8.86 (5.03)	1.08 (0.61)	3.15 (1.84)	6.37 (2.62)	0.78 (0.32)	2.84 (2.05)	4.47 (2.80)	0.55 (0.34)	1.98 (1.40)
N	15			17			17		

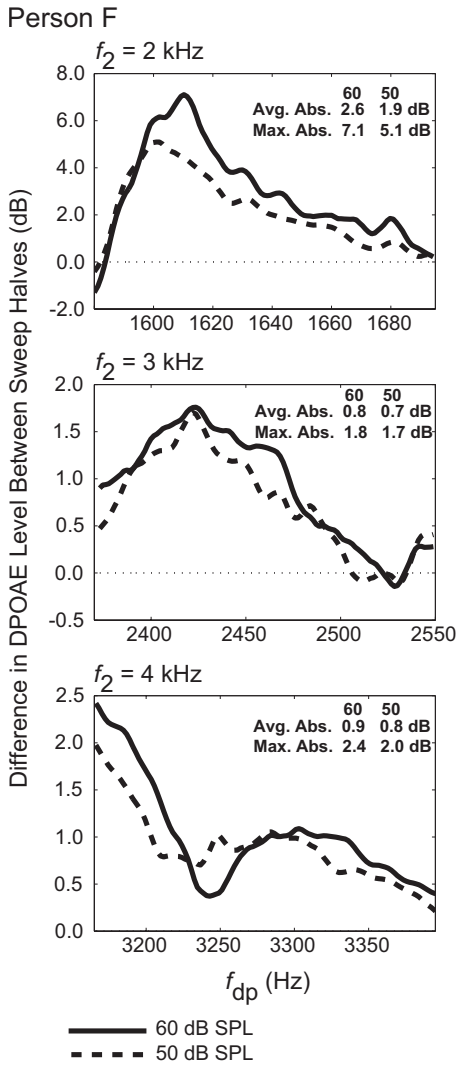


FIG. 5. Example data from Experiment 2 for a single individual (different person from previous figures) showing the effect of contralateral suppressor level. The vertical axes are the change in DPOAE level between the first half of the sweep with no contralateral noise and the second half with contralateral noise. The horizontal axes are emission frequency. Stimulus frequency f_2 increases from the top to the bottom plots. Positive values indicate DPOAE suppression and negative values show enhancement. The solid line is for contralateral noise of 60 dB SPL and the dashed line is for 50 dB SPL. The change metrics, mean absolute difference and maximum absolute difference, are given in the top right of the plots for each condition. The signal-to-noise ratio in these measurements was excellent: approximately 29, 37, and 43 dB for f_2 of 2, 3, and 4 kHz, respectively.

absolute change metrics. These were larger for the narrower stimulus ratio condition $f_2/f_1=1.10$. A significant main effect was also present for f_2 frequency [$F(1.8,29.6)=9.2$, $p<0.001$], and there was no significant interaction. Post-hoc testing on f_2 showed that these metrics were smaller at $f_2=4$ kHz compared to 3 kHz [$F(1,16)=19.0$, $p<0.001$] and at 4 kHz compared to 2 kHz [$F(1,16)=13.2$, $p<0.005$]. There was no significant difference between $f_2=2$ and 3 kHz. For the maximum absolute change metric there was also a significant main effect of ratio [$F(1,16)=12.9$, $p<0.005$], where this metric was larger for the narrower ratio 1.10. A significant main effect was also present for f_2 frequency [$F(1.7,26.8)=8.6$, $p<0.005$], and there was no significant interaction. Post-hoc testing on f_2 showed $f_2=4$ kHz had smaller changes than 3 kHz [$F(1,16)=15.6$, $p<0.001$] and 2 kHz [$F(1,16)=13.7$, $p<0.005$]. There was no significant difference between $f_2=2$ and 3 kHz.

An additional ANOVA was performed to evaluate DPOAE level (in dB SPL) because it was expected that emissions would be larger at the clinical stimulus frequency ratio 1.22 compared to the narrower ratio used in all three experiments. Mean DPOAE level was determined for the first half of each average sweep, where no contralateral noise had been presented. Factors were again stimulus frequency ratio (narrower ratio versus wider) and f_2 frequency (2, 3, and 4 kHz). There was a significant main effect of ratio [$F(1,16)=116.6$, $p<0.001$], where the wider clinical ratio produced larger DPOAE magnitudes. There was also a significant main effect of f_2 frequency [$F(1.7,26.6)=17.4$, $p<0.001$] and no significant interaction. Post-hoc tests showed that DPOAEs at $f_2=3$ kHz were smaller than at 2 kHz [$F(1,16)=22.3$, $p<0.001$] and smaller than at 4 kHz [$F(1,16)=21.5$, $p<0.001$]. There was no significant difference in DPOAE level between $f_2=2$ and 4 kHz.

For the wider ratio, changes in DPOAE level were also determined at the single DPOAE frequency that might be used clinically for fixed frequency stimuli and the stimulus frequency ratio 1.22 (i.e., near the center point of the DPOAE range measured here). For $f_2=2$ to 4 kHz, these emission frequencies were 1279, 1918, and 2557 Hz, respectively. An average change in absolute DPOAE level at these frequencies was calculated for the participants in Table IV and the values were 1.08 (0.88), 0.98 (0.82), and 0.57 (0.50) dB with s.d. shown in parentheses. At the three ascending f_2 frequencies, 74%, 85%, and 73% of individuals showed suppression instead of enhancement. A repeated measures ANOVA was calculated for the 17 individuals with valid data at all f_2 with factors of metric (single point absolute change, average absolute change across ramped DPOAE range, and maximum absolute change across range) and f_2 (2, 3, and 4 kHz). There was a significant main effect of metric [$F(1.3,20.1)=51.4$, $p<0.001$] and similar results as reported above involving f_2, but these will not be discussed further for brevity. Post-hoc testing on metric showed no significant difference between the single point absolute change and average absolute change. Maximum absolute change was however larger than either the single point [$F(1,16)=52.8$, $p<0.001$] or average absolute change [$F(1,16)=58.0$, $p<0.001$].

D. Repeatability on different dates

The narrower stimulus frequency ratio $f_2/f_1=1.10$ and the 60 dB SPL contralateral noise were used as one condition in all experiments. There were ten participants who had this measurement performed in one ear on at least two different dates. Five of these individuals had three measurements in the same ear on different dates. Repeated measures ANOVAs were calculated for the ten participants with factors of date (A and B) and f_2 frequency (2, 3, and 4 kHz). The first measurement was discarded for those who had three available. There were no significant effects of date for either mean absolute change or maximum absolute change. Additionally,

TABLE IV. Average results for Experiment 3 with standard deviations in parentheses. The values in the first three columns at each f_2 frequency represent the difference in DPOAE level between the two halves of the sweep, where contralateral noise was present for the second half of the sweep. The given metrics are the area of the difference function in dB s, the average absolute difference in dB, and the maximum absolute difference in dB. The fourth column at each f_2 gives the mean DPOAE level across that measurement range for the first half of the average sweep where no contralateral noise was present. The number of participant's data included at each frequency is given in the last row; data for some participants was excluded at one frequency due to failing the noise floor criteria given in the methods section for the average sweep.

	f_2=2000 Hz				f_2=3000 Hz				f_2=4000 Hz			
Condition	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Avg. DPOAE (dB SPL)	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Avg. DPOAE (dB SPL)	Area (dB s)	Avg. abs. (dB)	Max. abs. (dB)	Avg. DPOAE (dB SPL)
Narrower ratio 1.10	8.59 (4.41)	1.05 (0.54)	3.22 (1.39)	−4.27 (5.13)	9.11 (4.26)	1.11 (0.52)	3.92 (1.93)	−7.85 (6.25)	7.64 (3.74)	0.93 (0.46)	2.83 (1.76)	−3.83 6.26
Wider ratio 1.22	7.89 (4.17)	0.95 (0.50)	2.68 (1.75)	6.18 (4.78)	6.61 (2.64)	0.81 (0.32)	2.70 (1.51)	−0.91 (5.45)	3.44 (2.17)	0.42 (0.26)	1.16 (0.77)	3.21 (4.51)
N	19				20				22			

further ANOVAs were calculated for the five participants with three dates, where the factors were again date (A, B, and C) and f_2 frequency (2, 3, and 4 kHz). Again, there were no significant effects involving date. Using the data from all ten people, the mean changes between dates were calculated at each frequency. For f_2=2, 3, and 4 kHz, the changes between dates were −0.04 (0.29), −0.10 (0.44), and 0.10 (0.24) dB for the average absolute effect of contralateral noise. Across frequencies, Pearson's linear correlation coefficient was 0.79 ($p<0.001$). For the maximum absolute ef-

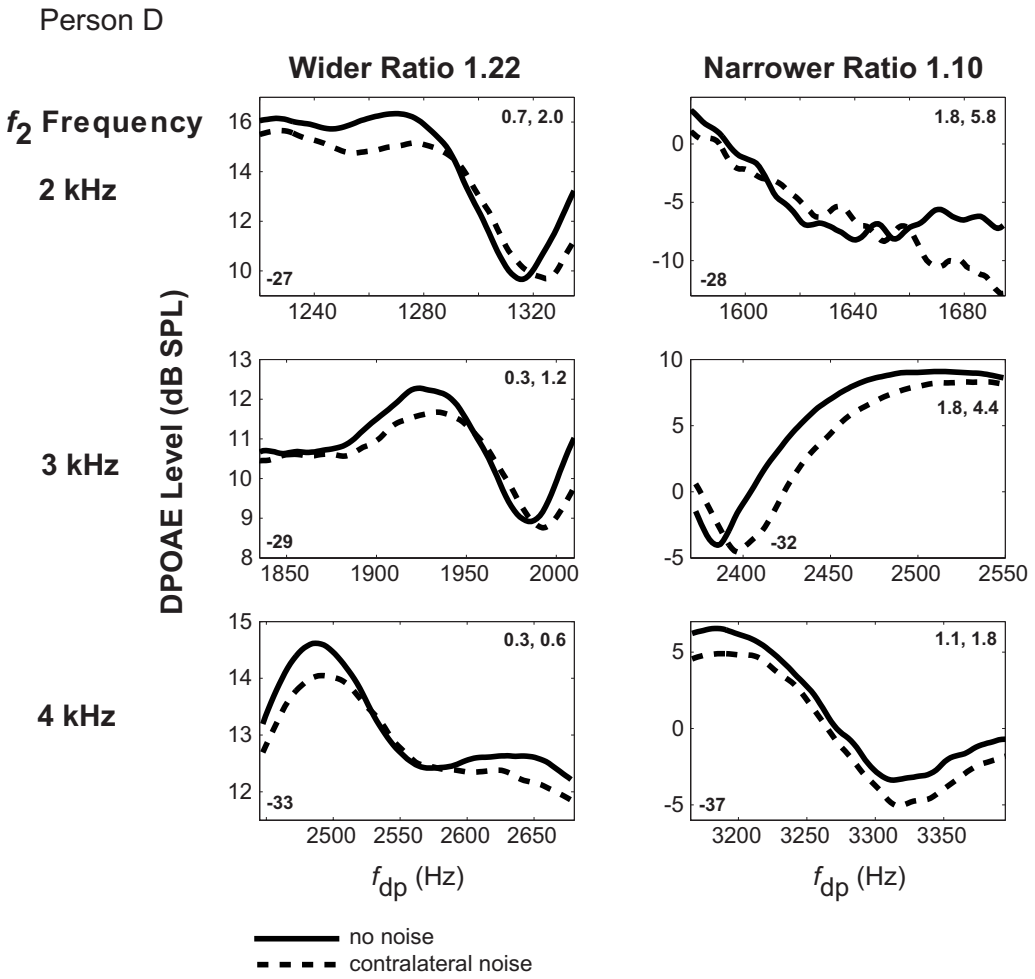


FIG. 6. Example individual DPOAE level data from one person in Experiment 3. From top to bottom, each row shows data for f_2 frequencies from 2 to 4 kHz. The columns show results for the two stimulus frequency ratios with the wider clinical ratio on the left. DPOAE level in dB SPL is given on the vertical axes, and emission frequency is on the horizontal axes. Note that the vertical ranges in the left column are smaller than for the narrower ratio in the right column. Solid lines are for the first half of the sweep with no contralateral noise, and dashed lines are for the second half of the sweep with contralateral noise. Values given in the lower left corners of each plot are the estimated noise floors in dB SPL. The two values given in the top right corner of each plot are the mean absolute difference and maximum absolute difference in dB that were obtained by subtracting the two curves in each panel.

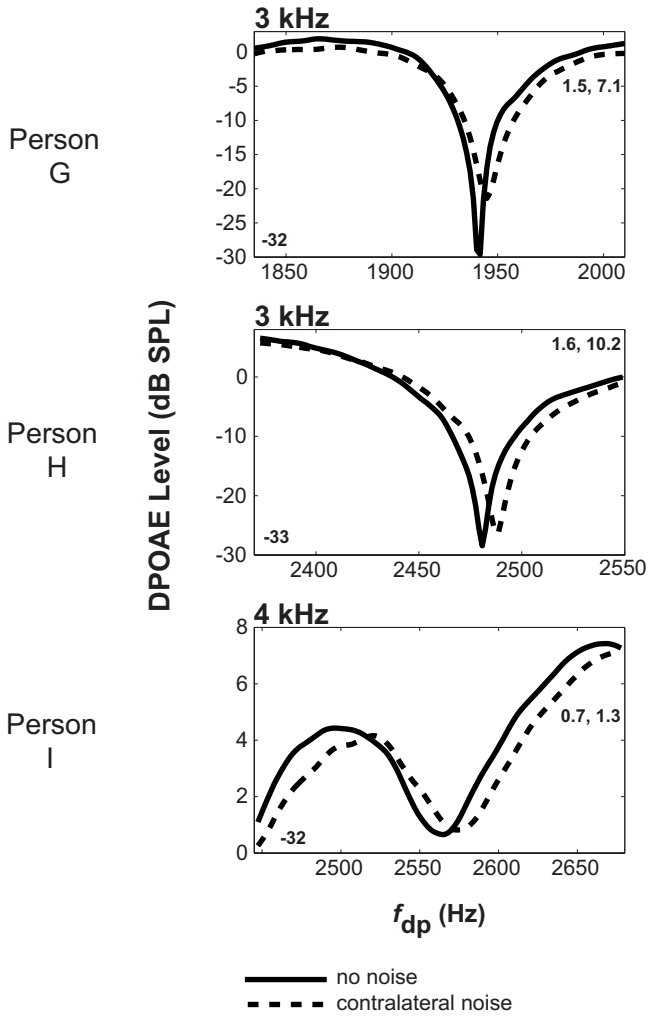


FIG. 7. Example data from an additional three individuals that were chosen to show interesting morphology and the changes that occurred with contralateral noise. Each row is a participant whose data has not been shown previously. DPOAE level in dB SPL is given on the vertical axes and emission frequency is shown on the horizontal axes in Hz. The top and bottom plots are from Experiment 3 for the wider stimulus frequency ratio when f_2=3 and 4 kHz, respectively. The middle plot is from Experiment 2 (and therefore the narrower ratio) for the condition with 60 dB SPL contralateral noise and f_2=3 kHz. As with previous figures, the estimated noise level in the average sweep is shown in the lower left corner of each plot in dB SPL. The values in the top right corners are the difference metrics, mean absolute change and maximum absolute change, in dB that are obtained by subtracting the two curves in each panel.

fect of contralateral noise, the changes between dates were −0.02 (1.18), −0.51 (2.28), and 0.00 (1.32) dB, respectively. Across frequencies, Pearson's correlation coefficient was 0.73 ($p<0.001$) for this metric.

E. Detailed emission morphology

Figure 7 shows additional data from another three participants who have not appeared in previous plots. These data were chosen to show more examples of interesting emission features that were observed across the measured DPOAE frequency ranges, and of the changes that occurred with contralateral acoustic stimulation. The first two plots show deep nulls of over 30 dB in the response. The third plot shows an example of response details that occurred over a much smaller range of emission level. In each of these cases, the presence of contralateral noise shifted the emission morphology upwards in frequency. This was a very common change pattern across the experiments when distinctly featured morphology was present.

IV. DISCUSSION

A. Validity of DPOAE change functions

When trying to measure the effects of MOC efferent activity on otoacoustic emissions, there is the potential to observe changes that are due to the activity of the middle ear muscles. To avoid clear contamination of the results, participants were only included in this study if their broadband noise contralateral reflex thresholds were 70 dB HL or greater. In the experiments described here, the maximum contralateral noise level was 60 dB SPL, so middle ear muscle reflexes should not have been elicited. Many other studies have employed a contralateral 60 dB SPL broadband noise to obtain estimates of human MOC efferent effects (e.g., James *et al.*, 2002; Lisowska *et al.*, 2002; James *et al.*, 2005; Müller *et al.*, 2005; Wagner *et al.*, 2007; Zhang *et al.*, 2007). James *et al.* (2005) note from their real-time analysis method that the time course of middle ear muscle responses is slower and the size of the changes are an order of magnitude larger. They felt the changes they observed were reflective of MOC efferent activity. Sun (2008) also concluded that DPOAE changes in response to contralateral stimuli lower in level than the contralateral ART probably do not reflect middle ear muscle activity. One of the most compelling arguments against the middle ear muscles significantly contaminating the results is that DPOAE level enhancement would not be observed if the middle ear muscles were responsible for the changes (Müller *et al.*, 2005). This is because the stimuli would be reduced in level on the inward crossing of the middle ear, which would generally lead to a reduction in DPOAE magnitude at the stimulus levels used here (i.e., before input-output saturation, Whitehead *et al.*, 1995b). DPOAE level would be further reduced on the outward crossing of the middle ear. The fact that enhancement occurred at some frequencies for most ears suggests that real MOC efferent effects were being observed.

A second concern is that background noise in the ear canal could contaminate DPOAE estimates and lead to small changes from one measurement to the next that could be confused with true MOC efferent effects. Twenty-one sweeps were included in the average to reduce background noise; of course more sweeps are desirable, but compromises must be made to limit measurement time. No DPOAE estimates were included in the analysis unless the estimated signal-to-noise ratio exceeded 6.9 dB [this is the signal-to-noise ratio required for $p<0.01$ using the F-ratio $F(2,80)=4.88$]. Since the noise estimate itself was from a frequency band lower than the DPOAE frequencies, the actual minimum signal-to-noise ratio was probably slightly higher (due to increasing ear canal noise at lower frequencies). Additional constraints were placed on maximum allowable noise estimates as outlined in the methods section. At f_2=3 and 4 kHz, the signal-

TABLE V. Summary of changes in DPOAE level observed in example studies of contralateral acoustic stimulation in humans. Stimulus levels are in dB SPL. In the final column, mean changes with standard deviations (when available) are given in dB. The f_2 stimulus frequency follows in parentheses. Since the present study had multiple measurements at the narrower ratio with 60 dB SPL contralateral noise, data were combined across studies. The data of individuals with repeated measurements were averaged across dates before inclusion in the grand means for this table. Contralateral acoustic stimulus type was broadband noise in the studies listed.

		DPOAE stimulus			Contralateral Stimulus		DPOAE Change [mean ± s.d. dB]
Study	N	f_2/f_1 ratio	L1 (dB SPL)	L2 (dB SPL)	Level (dB SPL)	Type	(f_2 frequency)
Bassim *et al.* (2003)	24	1.21	70	65	60	BBN	1.10 (1.8–9.7 kHz)
James *et al.* (2005)	10	1.22	70	65	50–60	BBN	1.14 (4.4 kHz)
Lisowska *et al.* (2002)	10	1.22	65	55	60	BBN	1.00 (1.5 kHz)
		1.22	55	40	60	BBN	1.50 (2 kHz)
Williams and Brown (1995)	6	1.225	55	40	55	BBN	1.99 ± 0.69 (2 kHz)
							1.12 ± 1.00 (3 kHz)
Zhang *et al.* (2007)	20	1.22	55	40	40	BBN	0.31 (1–6 kHz)
		1.22	55	40	60	BBN	1.26 ± 0.89 (1–1.5 kHz)
							1.59 ± 2.40 (1.5–2.5 kHz)
							0.62 ± 0.61 (3.5–4.5 kHz)
		1.22	55	40	70	BBN	1.44 (1–6 kHz)
Present	19	1.20–1.25	60	55	60	BBN	0.95 ± 0.50 (2 kHz)
	20						0.81 ± 0.32 (3 kHz)
	22						0.42 ± 0.26 (4 kHz)
	25	1.08–1.12	60	55	60	BBN	1.33 ± 0.68 (2 kHz)
	26						1.09 ± 0.47 (3 kHz)
	26						0.87 ± 0.42 (4 kHz)
	15	1.08–1.12	60	55	50	BBN	1.08 ± 0.61 (2 kHz)
	17						0.78 ± 0.32 (3 kHz)
	17						0.55 ± 0.34 (4 kHz)

to-noise ratio typically exceeded 20 dB. It generally exceeded 10 dB for f_2=2 kHz, where canal noise is higher.

Since absolute values were used to quantify the changes between the first and second halves of the average sweep, background ear canal and other measurement noise led to small positive values, as shown for the control condition in Fig. 2. DPOAE changes in the test condition with contralateral stimulation were significantly larger and showed an effect of f_2. The fact that the control condition did not show an effect of f_2 is favorable because it demonstrates that background ear canal noise was not a driving factor in the observed f_2 effects for the test condition (where the response was smaller at 4 kHz than at 2 and 3 kHz, which were not themselves statistically different).

For a clinical test, averaging time might be reduced if a control measurement was included for each client. This could provide an internal benchmark for the variability expected before contralateral noise is actually introduced. For noisy individuals, the averaging time could be extended until acceptable variability is demonstrated.

B. Changes in DPOAE level with contralateral acoustic stimulation

The changes in DPOAE level with contralateral acoustic stimulation that were observed in Experiments 1, 2, and 3 are similar to those reported in previous studies. The average areas of the difference functions that are reported in Tables II–IV will be more usable if future studies publish data using ramped techniques. Table V summarizes some example data that can be compared to values obtained here for the mean absolute change metric. With the exception of the present study, estimates were obtained near the clinically favored f_2/f_1 ratio 1.22. The changes reported in the present study are somewhat smaller than observed at f_2=2 kHz by Williams and Brown (1995) and Lisowska *et al.* (2002), which may be partially due to the Williams and Brown (1995) finding that lower level stimulus tones elicit DPOAEs that are influenced to a greater extent by contralateral acoustic stimulation. The present study chose higher stimulus tone levels to evoke larger DPOAEs and thus improve signal-to-noise ratios. Zhang *et al.* (2007) also used lower level stimulus tones and employed a search strategy, where the tabulated values are average changes at frequencies identified as emission peaks in the ranges listed. Values from standard fixed audiometric frequencies were smaller.

The studies in Table V observed contralateral suppression at one or more sets of stimulus frequencies, whereas the present study ramped f_1 while maintaining a constant f_2 to record a continuous response function that may contain emission maxima and minima. Although the mean absolute changes reported here are similar to values found in the literature (about 1 dB), it is important to note that the present values represent the average across each DPOAE frequency range. The maximum absolute change was larger; across studies the mean values were between 3 and 4 dB (the largest value found was 11.8 dB). Studies employing search strategies that find fine structure and subsequently make a large set of measurements with small increments in stimulus level have reported a related value. For the entire measurement set, the range is determined between the largest en-

hancement and the largest suppression that were found. Range values that have been reported are 9.1 dB (Wagner *et al.*, 2008), 9.2 dB (Wagner *et al.*, 2007), and 14.1 dB (Müller *et al.*, 2005). In this last report, the average maximum suppression and enhancement were 7.3 and 6.8 dB, respectively (7.3+6.8=14.1 dB). In the present study, we chose to report the maximum absolute value (the largest suppression or enhancement), which are therefore smaller than the range between maximum suppression and enhancement. Both measurements could be vulnerable to background ear canal noise events that have the potential to contaminate a small set of measurements. Great care is taken to avoid including noisy data, but the average absolute change discussed in the next paragraph is more robust against noise events.

The choice of a narrower stimulus frequency ratio was made to encourage fine structure, and DPOAE changes in response to contralateral acoustic stimulation were somewhat larger compared to the wider ratio in Experiment 3. The generation mechanism responsible for the DPOAEs measured in the ear canal here is, however, unknown. This passive, standardized approach did not elicit changes as large as those found in search strategies where $f_2/f_1=1.2$. In response to contralateral noise of 60 dB SPL and in a fine structure dip, Wagner *et al.* (2007) report mean changes across a large set of stimulus levels of 2.17 dB. A second set of measurements at a frequency region chosen without a distinct fine structure dip (for the more standard stimulus levels used in the initial search) gave a smaller mean change of 1.29 dB. This value is closer to those found in the present and other studies listed in Table V. Directed searching can elicit larger efferent effects than the approach used here; however, it may be possible to demonstrate the presence of efferent activity in a shorter amount of time using the present continuous frequency range approach. Additionally, the two methods survey efferent activity for different dimensions of the DPOAE stimulus space (i.e., frequency versus stimulus levels), and therefore are complementary in the information that they provide.

In Experiment 3 for the wider stimulus frequency ratio, the mean absolute changes across DPOAE ranges were not significantly different from those obtained at the single f_{dp} frequency in the middle of the ranges, where the ratio was exactly $f_2/f_1=1.22$. However, the maximum absolute changes were significantly (and substantially) larger. These maximums are only available because efferent activity was measured over a range of DPOAE frequencies. In a clinical situation the individual outcome matters, and there would be clients with intact MOC efferent systems whose single point measurement fell very close to zero because of unfortunate transitions from suppression to enhancement close to that single measurement frequency. For example, of the 20 participants in Experiment 3 with data for $f_2=3$ kHz, two individuals had single point absolute changes less than 0.20 dB. Across the DPOAE frequency range for the wider ratio, these two had mean absolute changes of 1.5 and 0.5 dB, and maximum absolute changes of 7.1 and 1.4 dB, respectively. While still small for the second person, the range of DPOAE frequencies surrounding $f_2/f_1=1.22$ afforded more opportunity for the MOC efferent system to reveal itself. If the DPOAE function graphs were not available for inspection, it would, however, be prudent to run a control condition without contralateral noise to determine the expected DPOAE variability within each of these individuals. Data for the first person are shown in the top panel of Fig. 7 (person G). The shift in the deep emission null to higher frequencies with contralateral noise leads to the larger change metrics. Note that the DPOAEs in the deepest part of the notch would not be included in the change metric calculations because they would approach the noise estimate and would be considered invalid. The single point measurement occurs at 1918 Hz, which is very close to where the two DPOAE functions intersect one another on the left side of the large dip. Inspection of the curves for the second person (not presented here) showed a DPOAE response resembling a single cycle of a sinusoid (e.g., a sine function starting at phase 0 rad with peak-to-peak "amplitude" of about 5 dB). With contralateral noise, that DPOAE response function shifted slightly to higher frequencies and the change was mostly suppression. The small change registered for the single point measurement occurred because 1918 Hz again happened to fall at one of the intersections of the two DPOAE response functions (without and with contralateral noise). In both these examples, it is clear that background ear canal noise is not responsible for the observed changes between measurements because the morphology of the DPOAE response function was repeated in each measurement, albeit shifted along the frequency axis. Zhang *et al.* (2007) observed and raised the issue that single point measurements at audiometric frequencies produced smaller average DPOAE changes than using their search strategy (see their Table 1 and Fig. 3).

A significant effect of frequency was found in the experiments, where changes in DPOAE level with contralateral noise were smaller at $f_2=4$ kHz. There were no statistically significant differences between f_2 of 2 and 3 kHz. In Experiment 3, changes were larger for the narrower ratio $f_2/f_1=1.10$ compared to the wider ratio 1.22. This was particularly evident for $f_2=4$ kHz, where the average results were more than double for both the mean absolute change and the maximum absolute change. If 4 kHz is determined to be an important frequency to test clinically, then the narrower ratio would be favorable. Until clinical requirements are better established, a simple demonstration of MOC efferent activity within an individual using this method is likely to be more robust at f_2 of 2 or 3 kHz. Other studies have also reported that changes can be larger for lower f_2. For example, Lisowksa *et al.* (2002) found $f_2=1.5$ and 2 kHz to be favorable (see their Fig. 3). This is consistent with other reports suggesting greater changes in the region 1 to 3 kHz (Moulin *et al.*, 1993; Chery-Croze *et al.*, 1993; Williams and Brown, 1997). Zhang *et al.* (2007) also show larger changes for lower frequencies (see their Table 1). Background ear canal noise is greater at lower frequencies, but with due care to exclude noisy data, the results of Experiment 1 suggest that this frequency effect is real. The control condition in Experiment 1 did not show an effect of f_2, whereas it was observed to be a factor in all experiments when contralateral noise was presented to induce MOC efferent activity.

In Experiment 2, larger DPOAE changes were observed with higher level contralateral noise. This is consistent with the literature where an increase in DPOAE changes are generally observed with increasing level of the contralateral stimulus (e.g., Fig. 5 in James *et al.*, 2002; Fig. 4 in James *et al.*, 2005; Fig. 5 in Zhang *et al.*, 2007). Harrison *et al.* (2008) have shown in a chinchilla model that DPOAE changes are continuously linearly modulated without a threshold effect by using an amplitude modulated contralateral stimulus.

Repeatability of the DPOAE change measurements in the same ear on different dates was very good. Across the group of individuals with repeated measurements, there were no statistically significant changes and the mean differences were distributed around small values approaching zero. This suggests that the influence of background ear canal noise was sufficiently attenuated to allow good response estimates and that, as expected, the cochlear, middle ear, and neural systems involved are stable over time. These are attractive features for potential clinical utility. Reasonably good repeatability has also been reported previously in the literature (Müller *et al.*, 2005).

C. Detailed emission morphology

Müller *et al.* (2005) identified large bipolar changes in the effect of contralateral stimulation in humans (from suppression to enhancement), as a result of small shifts in stimulus tone level. This was particularly evident at frequencies where DPOAE fine structure exhibited pronounced dips. These dips were located in the DPOAE stimulus frequency space with f_2 steps of 47 Hz (using a fixed ratio $f_2/f_1=1.2$), and subsequently changes with contralateral noise were observed by stepping the stimulus tone levels in small increments. The range of DPOAE magnitude changes elicited in this manner are far greater than the change found for a single set of fixed stimulus parameters. This is because the stimulus level combination eliciting the greatest enhancement of the distortion product and the combination producing the greatest suppression are used to calculate the overall range in each individual (as mentioned above). Wagner *et al.* (2007) were able to locate fine structure dips and consequent bipolar changes due to contralateral stimulation in all 22 of their participants. The highest incidence of fine structure dips was found in the frequency range below 4.2 kHz. Fine structure dips likely occur when the two DPOAE sources interfere with similar magnitudes and near opposite phase polarities. If the contralateral acoustic stimulation has more effect on one source than the other, or if the sources are modulated on different trajectories, then the change in emission level near fine structure could be pronounced. This might not be as dramatic in the case of a single dominant source, or sources changing on the same trajectory. These concepts are supported by studies showing that the largest efferent effects are found at frequencies where DPOAE fine structure dips are exhibited (Müller *et al.*, 2005; Wagner *et al.*, 2007; Wagner *et al.*, 2008). However, peaks in emission level have also been reported to be more favorable for displaying changes (suppression) with contralateral stimulation. Williams and Brown (1997) report for four participants that suppression was most pronounced at fine structure peaks, whereas at dips it was suppressed less, unchanged, or enhanced. They also suggested from analysis of a small data set that the longer latency component (i.e., from the f_{dp} place source) of DPOAEs in the ear canal can be most affected by contralateral stimulation. This, however, was only observed in some of their data. Zhang *et al.* (2007) report that the points identified as peaks in their data exhibited greater suppression due to contralateral stimulation than those identified as dips, but the range of expected changes including suppression and enhancement appears larger for the points identified as dips in their Fig. 3. Their Fig. 2 shows the most enhancement was observed in the dips, and they noted more between-subject variability in the dips.

In the present study, an attempt was made to encourage interference between DPOAE sources and hence fine structure by choosing a narrower stimulus frequency ratio $f_2/f_1=1.10$. As mentioned above, however, the actual generation mechanism of the recorded emissions is unknown. Using a narrower ratio, the DPOAE frequencies are higher than for $f_2/f_1=1.22$, which means the background noise was lower (on average by 3.5 dB across frequencies in Experiment 3). As expected, however, the DPOAEs were also smaller (on average by 8.2 dB across frequencies), and this is of course the motivation for using $f_2/f_1=1.22$ in clinical measurements (Harris *et al.*, 1989; Abdala, 1996). Through rejection of noisy epochs prior to averaging, averaging itself, and background noise criteria applied to the average sweep, changes in the test condition reported in Experiment 1 were significantly larger than for the control condition without contralateral noise. As discussed above, the benefit of the narrower ratio was most evident for $f_2=4$ kHz, where the influence of ear canal noise would be expected to be smallest, supporting that it is a real effect.

Extreme examples showing deep DPOAE nulls were presented in Fig. 4 (person C, top row) and Fig. 7 (persons G and H, top two plots). It was originally hypothesized that the narrower ratio would produce larger changes with contralateral noise by facilitating dramatic DPOAE response functions. Distinctly featured responses were not, however, remarkably more prevalent at one stimulus frequency ratio or another. The larger mean changes with contralateral noise that were observed at the narrower ratio are probably the aggregate of larger changes in the more subtle emission morphology.

The bandwidth of deep emission nulls discussed in the previous paragraph (and reported in other studies: e.g., Dhar *et al.*, 2005) is sufficiently narrow that search strategies could overstep them without sufficiently closely spaced measurements. Wagner *et al.* (2007) use an f_2 step size of 47 Hz, which has an f_{dp} step of 31 Hz for $f_2/f_1=1.2$. The deepest part of the nulls in Figs. 4 and 7 might be overlooked without fortuitously placed f_2. Although the nulls discussed above are probably too deep for practical measurement of DPOAEs (due to background noise), it is possible that search strategies may not always converge on the optimal dip. The challenge posed in defining peaks and valleys was identified by Zhang *et al.* (2007), where they noted that the definitions might

change with the resolution of the frequency steps used. A continuous DPOAE ramp technique like that described here could help to more fully characterize the stimulus frequency space prior to mapping fine structure along stimulus level dimensions.

A very common result was an apparent shift in the morphology of the DPOAE response function to higher frequencies in the presence of contralateral noise. Enhancement of the DPOAE often occurred where the slope of the function was steeply negative (decreasing DPOAE level with increasing frequency), such as shown in the left column of Fig. 6 and each panel of Fig. 7. This common shift in morphology suggests that the emission measured in the ear canal is affected in a relatively consistent way by MOC efferent activity elicited with contralateral noise, despite the complexity of DPOAE generation within the cochlea.

V. CONCLUDING REMARKS

The purpose of this study was to evaluate DPOAE change functions obtained over continuous frequency ranges in response to contralateral acoustic stimulation. The method appears promising for relatively rapid evaluation of the human MOC efferent system and is complementary to techniques exploring efferent effects by manipulating stimulus levels. Repeatability between measurement sessions was good and responses were shown to be significantly larger than changes observed due to background noise in the ear canal. Measurements of this type must be vigilant to avoid background noise that could mimic the effects of MOC efferents on DPOAE amplitude. The use of a control condition without contralateral stimulation may be prudent for clinical applications; however, the graphical nature of the response allows some critical judgment to be applied to DPOAE curves obtained without and with contralateral noise.

The smallest changes with contralateral noise were observed at f_2=4 kHz, and there were no statistical differences between 2 and 3 kHz. The use of a narrower stimulus frequency ratio f_2/f_1=1.10 elicited larger changes with contralateral stimulation compared to f_2/f_1=1.22, particularly at 4 kHz where the observed changes were more than doubled. DPOAE response function morphology was observed to shift to higher frequencies with contralateral noise. To permit clinical use of contralateral acoustic stimulation as a test of the human MOC efferent system, further data must be collected to determine the range of normal and expectations for different disorders.

ACKNOWLEDGMENTS

The authors would like to thank the Hearing Foundation of Canada and the Canada Foundation for Innovation for their support of this research.

Abdala, C. (**1996**). "Distortion product otoacoustic emission (2f1-f2) amplitude as a function of f2/f1 frequency ratio and primary tone level separation in human adults and neonates," J. Acoust. Soc. Am. **100**, 3726–3740.

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.

Chery-Croze, S., Moulin, A., and Collet, L. (**1993**). "Effect of contralateral sound stimulation on the distortion product 2f1-f2 in humans: Evidence of a frequency specificity," Hear. Res. **68**, 53–58.

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.

Frisina, R. D., Newman, S. R., and Zhu, X. (**2007**). "Auditory efferent activation in CBA mice exceeds that of C57s for varying levels of noise," J. Acoust. Soc. Am. **121**, EL29–EL34.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

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.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). "Acoustic distortion products in humans: Systematic changes in amplitudes as a function of f2/f1 ratio," J. Acoust. Soc. Am. **85**, 220–229.

Harrison, R. V., Sharma, A., Brown, T., Jiwani, S., and James, A. L. (**2008**). "Amplitude modulation of DPOAEs by acoustic stimulation of the contralateral ear," Acta Oto-Laryngol. **128**, 404–407.

Heitmann, J., Waldmann, B., Schnitzler, H., Plinkert, P., and Zenner, H. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near 2f1-f2 removes DP-gram fine structure-Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Jacobson, M., Kim, S., Romney, J., Zhu, X., and Frisina, R. D. (**2003**). "Contralateral suppression of distortion-product otoacoustic emissions declines with age: A comparison of findings in CBA mice with human listeners," Laryngoscope **113**, 1707–1713.

James, A. L., Harrison, R. V., Pienkowski, M., Dajani, H. R., and Mount, R. J. (**2005**). "Dynamics of real time DPOAE contralateral suppression in chinchillas and humans," Int. J. Audiol. **44**, 118–129.

James, A. L., Mount, R. J., and Harrison, R. V. (**2002**). "Contralateral suppression of DPOAE measured in real time," Clin. Otolaryngol. **27**, 106–112.

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.

Kim, S., Frisina, D. R., and Frisina, R. D. (**2002**). "Effects of age on contralateral suppression of distortion product otoacoustic emissions in human listeners with normal hearing," Audiol. Neuro-Otol. **7**, 348–357.

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

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

Kujawa, S. G., and Liberman, M. C. (**2001**). "Effects of olivocochlear feedback on distortion product otoacoustic emissions in guinea pig," J. Assoc. Res. Otolaryngol. **2**, 268–278.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). "Optimal L(1)-L(2) primary tone level separation remains independent of test frequency in humans," Hear. Res. **146**, 47–56.

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.

Moulin, A., Collet, L., and Duclaux, R. (**1993**). "Contralateral auditory stimulation alters acoustic distortion products in humans," Hear. Res. **65**, 193–210.

Müller, J., Janssen, T., Heppelmann, G., and Wagner, W. (**2005**). "Evidence

for a bipolar change in distortion product otoacoustic emissions during contralateral acoustic stimulation in humans," J. Acoust. Soc. Am. **118**, 3747–3756.

Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (**2004**). "Human temporal auditory acuity as assessed by envelope following responses," J. Acoust. Soc. Am. **116**, 3581–3593.

Purcell, D. W., Van Roon, P., John, M. S., and Picton, T. W. (**2006**). "Simultaneous latency estimations for distortion product otoacoustic emissions and envelope following responses," J. Acoust. Soc. Am. **119**, 2869–2880.

Regan, D. (**1989**). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*, Elsevier Science Publishing Co., Inc., New York.

Roup, C. M., Wiley, T. L., Safady, S. H., and Stoppenbach, D. T. (**1998**). "Tympanometric screening norms for adults," Am. J. Audiology **7**, 55–60.

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

Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.

Silva, A., and Ysunza, A. (**1998**). "Effect of contralateral masking on the latency of otoacoustic emissions elicited by acoustic distortion products," Int. J. Pediatr. Otorhinolaryngol. **44**, 125–132.

Sun, X. M. (**2008**). "Contralateral suppression of distortion product otoacoustic emissions and the middle-ear muscle reflex in human ears," Hear. Res. **237**, 66–75.

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.

Varghese, G. I., Zhu, X., and Frisina, R. D. (**2005**). "Age-related declines in distortion product otoacoustic emissions utilizing pure tone contralateral stimulation in CBA/CaJ mice," Hear. Res. **209**, 60–67.

Wagner, W., Frey, K., Heppelmann, G., Plontke, S. K., and Zenner, H. P. (**2008**). "Speech-in-noise intelligibility does not correlate with efferent olivocochlear reflex in humans with normal hearing," Acta Oto-Laryngol. **128**, 53–60.

Wagner, W., Heppelmann, G., Müller, J., Janssen, T., and Zenner, H. P. (**2007**). "Olivocochlear reflex effect on human distortion product otoacoustic emissions is largest at frequencies with distinct fine structure dips," Hear. Res. **223**, 83–92.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). "Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **98**, 3200–3214.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L1, L2 space," J. Acoust. Soc. Am. **97**, 2359–2377.

Williams, D. M., and Brown, A. M. (**1995**). "Contralateral and ipsilateral suppression of the 2f1-f2 distortion product in human subjects," J. Acoust. Soc. Am. **97**, 1130–1140.

Williams, D. M., and Brown, A. M. (**1997**). "The effect of contralateral broad-band noise on acoustic distortion products from the human ear," Hear. Res. **104**, 127–146.

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.

Cochlear nonlinearity in normal-hearing subjects as inferred psychophysically and from distortion-product otoacoustic emissions

Peter T. Johannesen and Enrique A. Lopez-Poveda[a)]
Unidad de Audición Computacional y Psicoacústica, Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, 37007 Salamanca, Spain

(Received 28 April 2008; revised 14 July 2008; accepted 14 July 2008)

The aim was to investigate the correlation between compression exponent, compression threshold, and cochlear gain for normal-hearing subjects as inferred from temporal masking curves (TMCs) and distortion-product otoacoustic emission (DPOAEs) input–output (I/O) curves. Care was given to reduce the influence of DPOAE fine structure on the DPOAE I/O curves. A high correlation between compression exponent estimates obtained with the two methods was found at 4 kHz but not at 0.5 and 1 kHz. One reason is that the DPOAE I/O curves show plateaus or notches that result in unexpectedly high compression estimates. Moderately high correlation was found between compression threshold estimates obtained with the two methods, although DPOAE-based values were around 7 dB lower than those based on TMCs. Both methods show that compression exponent and threshold are approximately constant across the frequency range from 0.5 to 4 kHz. Cochlear gain as estimated from TMCs was found to be ~16 dB greater at 4 than at 0.5 kHz. In conclusion, DPOAEs and TMCs may be used interchangeably to infer precise individual nonlinear cochlear characteristics at 4 kHz, but it remains unclear that the same applies to lower frequencies.
 [DOI: 10.1121/1.2968692]

PACS number(s): 43.64.Jb, 43.66.Dc, 43.64.Ri, 43.64.Bt [BLM] Pages: 2149–2163

I. INTRODUCTION

It is almost certain that our ability to perceive sounds over a 120 dB level range is accomplished via a form of compression that takes place in the basilar membrane (BM) (Bacon, 2004). Listeners with sensorineural hearing loss show reduced auditory dynamic ranges and this is typically interpreted as an indication of reduced cochlear compression. It is controversial, however, that this is actually the case (e.g., Heinz and Young, 2004; Plack *et al.*, 2004; Lopez-Poveda *et al.*, 2005) and yet knowing the degree of residual peripheral compression might help in diagnosing the type of hearing loss as well as improving its treatment with hearing aids. The long-term goal of this work is to design a fast and reliable technique to estimate the degree of *individual* residual compression in listeners with sensorineural hearing loss. This paper describes a first effort to determine whether distortion-product (DP) otoacoustic emission (DPOAE) input/output (I/O) functions could be used for that purpose in normal-hearing listeners. The motivation of the present work is thus similar to that of Müller and Janssen (2004).

There exist several psychoacoustical methods to estimate the amount of peripheral compression in humans [reviewed in Bacon (2004)]. Of these, the temporal masking curve (TMC) method of Nelson *et al.* (2001) (see also Lopez-Poveda *et al.*, 2003) is perhaps the most accurate because it minimizes off-frequency listening effects that might occur when the probe level is varied. This method consists of measuring the level of a tonal forward masker required to just mask a *fixed* tonal probe as a function of the time interval between the masker and the probe. A TMC is a graphical representation of the resulting masker levels against the corresponding masker–probe intervals. Because the probe level is fixed, the masker level increases with increasing masker–probe time interval and hence TMCs have positive slopes. Nelson *et al.* (2001) argued that the slope of any given TMC depends simultaneously on the amount of BM compression affecting the masker at a cochlear place whose characteristic frequency (CF) equals approximately the probe frequency and on the rate of decay of the internal (postcochlear) masker effect. By assuming that the decay rate is the same across masker frequencies, BM I/O functions may be estimated by plotting the masker levels of a linear reference TMC (i.e., the TMC for a masker that is processed linearly by the BM) against the levels for any other masker frequency, paired according to masker–probe delays [Nelson *et al.* (2001) provide a full justification of these assumptions; see also Lopez-Poveda and Alves-Pinto (2008)].

The TMC method has been used to infer BM I/O curves in normal-hearing and hearing-impaired listeners in a number of studies (e.g., Lopez-Poveda *et al.*, 2003, 2005; Lopez-Poveda and Alves-Pinto, 2008; Nelson *et al.*, 2001; Nelson and Schroder, 2004; Plack *et al.*, 2004; Rosengard *et al.*, 2005). It is arguably a reliable method but time consuming and requires the active participation of the listener. This would make it inconvenient for clinical purposes, particularly for testing newborns and the elderly.

On the other hand, DPOAE I/O curves share many characteristics with BM I/O functions (Cooper and Rhode, 1997; Dorn *et al.*, 2001; Neely *et al.*, 2003). Specifically, both of

a)Author to whom correspondence should be addressed. Electronic mail: ealopezpoveda@usal.es

them are generally linear at low levels but become compressive above a certain compression threshold (Dorn *et al.*, 2001; Kummer *et al.*, 1998) and both of them are similarly labile to outer hair cell damage (Rhode, 2007). Furthermore, measuring DPOAEs does not require the active participation of the listener. This suggests that DPOAE I/O functions could be used as a faster and universal way to infer individual BM I/O functions in clinical conditions (Müller and Janssen, 2004). The aim of the present study is to test this hypothesis by measuring the degree of correlation between BM I/O functions as inferred from TMCs and from DPOAEs in the same subject. Clearly, only if the results of the two methods correlate well will it be possible to add support to the use of DPOAEs to infer *individual* cochlear response characteristics. If they disagree, however, it will be difficult to resolve which of the two methods is more appropriate to reveal the true nonlinear characteristics of the underlying BM responses. Although the long-term goal is to extend the study to hearing-impaired listeners, the focus here is on normal-hearing listeners.

Several earlier studies have addressed this or related questions. Müller and Janssen (2004) investigated the similarity of loudness and DPOAE I/O curves in the same subject sample [Neely *et al.* (2003) had done it previously using different subject samples and slightly different methods]. They found a high resemblance between the characteristics (gain and compression) of the two sets of *average* I/O curves in normal-hearing and hearing-impaired listeners. Müller and Janssen (2004) acknowledged, however, that loudness may be affected by retrocochlear mechanisms (see also Heinz and Young, 2004) and it is also thought that loudness is affected by off-frequency effects (e.g., different spreads of excitation at different levels), which make it difficult to establish a one-to-one relationship between loudness and underlying BM I/O curves (Moore, 2003). This undermines the conclusions of Müller and Janssen (2004). Furthermore, their conclusions applied to *average* I/O curves and frequencies of 2–4 kHz, and thus may not be valid individually or for other frequencies, particularly 0.5 and 1 kHz.

Gorga *et al.* (2007) measured the degree of cochlear compression in a very large sample ($N=103$) of normal-hearing listeners as estimated from DPOAE I/O functions at 0.5 and 4 kHz. As a consequence, the I/O functions they reported likely provide a good description of *average* normal responses. Their results supported the conclusion of earlier psychophysical studies that the degree of compression is similar for apical and basal cochlear sites (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003). However, their study did not include within-subject psychophysical/physiological comparisons.

Williams and Bacon (2005) inferred cochlear I/O curves from TMCs and DPOAEs in four listeners and for frequencies of 1, 2, and 4 kHz. The results revealed that both methods yield similar *average* compression estimates. Like the above-mentioned studies, this study was not intended to investigate within-subject correlations between the results of both methods. Further, their DPOAE I/O curves could have been influenced by the DP fine structure. Indeed, Gaskill and Brown (1990) showed rapid variations (known as "fine structure") of the magnitude of the $2f_1-f_2$ DPOAE with changing the frequencies of the primaries (f_1 and f_2, with $f_2/f_1=1.21$) only slightly. This fine structure is thought to be the result of constructive and destructive interference between DPs generated at two spatially distant sites (Kummer *et al.*, 1995; Stover *et al.*, 1996; Gaskill and Brown, 1996; Heitmann *et al.*, 1998; Talmadge *et al.*, 1998; 1999; Mauermann *et al.*, 1999; Mauermann and Kollmeier, 1999; Shera and Guinan, 1999). The principal generation site is the BM region of maximum overlap between the excitation caused by the two primaries, that is the BM site with CF=f_2 (e.g., Kummer *et al.*, 1995). This component propagates back toward the oval window but also to the cochlear site with CF=$2f_1-f_2$ where it excites a second generation source. The DP generated at this second source propagates back toward the oval window and is summed with the response of the first source. The fine structure is thought to originate from vector summation of these two components, whose varying phases give rise to constructive or destructive interference and thereby to peaks and valleys in the DP-gram. The f_2 generator site is the dominant source at high stimulus levels [Fig. 3 of Mauermann and Kollmeier (2004)], which explains why the fine structure is more pronounced at low levels.

The fine structure has a large influence on DPOAE I/O curves, especially at low levels. He and Schmiedt (1993) mentioned that the DPOAE magnitude can change by as much as 20 dB for a change in f_2 of 1/32 octave. Mauermann and Kollmeier (2004) reported that the response varied by 10–15 dB when f_2 varied over the interval from 2250 to 2610 Hz. The influence of the fine structure is greater for individual than for average (across subjects) I/O curves, but can also affect average curves when the sample size is small. The sample size was small ($N=4$) in the study of Williams and Bacon (2005). Thus, fine-structure effects may have complicated the interpretation of their results or even led to wrong conclusions.

The present report extends these earlier studies in several respects. First, the focus here is on within-subject as opposed to average psychophysical/physiological correlations. Second, psychophysical BM I/O curves were inferred using what is arguably the most accurate method available to date for this purpose [see Nelson *et al.* (2001) for a full justification; but see also Sec. IV of the present paper]. Third, special care was exercised to reduce the influence of the fine structure on individual DPOAE I/O curves by averaging the magnitude of the $2f_1-f_2$ DPOAE for five f_2 frequencies near the frequency of interest. Fourth, the frequency range considered (0.5–4 kHz) included low and high frequencies. Fifth, the physiological/psychophysical comparisons extended to parameters pertaining to cochlear nonlinearity other than compression magnitude; specifically, compression threshold and the level at which maximum compression occurs.

It will be shown that reasonable correlation exists between the characteristics of individual TMC-based and DPOAE I/O curves at 4 kHz but not at 0.5 and 1 kHz. Reasons for the observed discrepancies at low frequencies will be discussed.

TABLE I. Thresholds (in dB SPL) measured with Etymotic ER2 insert earphones for all subjects and for tone durations of 300 ms (absolute threshold), 110 ms (masker threshold), and 10 ms (probe threshold), respectively. n.a. stands for not available.

	Frequency (kHz)			
Tone duration (ms)	0.5 300/110/10	1 300/110/10	2 300/110/10	4 300/110/10
S1	13/15/39	6/8/30	12/16/34	13/13/33
S2	14/21/43	10/15/30	20/22/n.a.	21/21/40
S3	13/16/39	7/8/32	5/8/27	4/10/25
S4	17/20/n.a.	10/16/34	16/21/42	21/24/44
S5	23/24/n.a.	13/17/41	7/12/n.a.	0/3/23
S6	9/10/36	10/11/31	9/13/n.a.	4/4/24
S7	6/10/n.a.	10/11/n.a.	16/20/n.a.	12/15/31
S8	11/18/34	5/7/38	25/28/n.a.	11/10/30
S9	12/n.a./n.a.	9/n.a./n.a.	10/n.a./n.a.	10/15/31
S10	5/n.a./n.a.	−2/n.a./n.a.	10/n.a./n.a.	3/1/23

II. METHOD

A. Subjects

Ten normal-hearing subjects participated in the study. Their age ranged from 20 to 39 years. Their absolute thresholds were measured using a two-down, one-up adaptive procedure. Signal duration was 300 ms, including 5 ms cosine-squared onset and offset ramps. All subjects had thresholds within 20 dB HL at the frequencies considered in this study (0.5, 1, 2, and 4 kHz, see Table I).

B. TMC stimuli

TMCs were measured for probe frequencies (f_P) of 0.5, 1, 2, and 4 kHz and for masker frequencies equal (on-frequency) to the f_P. Additional TMCs were measured for a probe frequency of 4 kHz and a masker frequency of 1.6 kHz (f_M=0.4f_p). The latter were selected as the linear references (Lopez-Poveda and Alves-Pinto, 2008) and used to infer BM I/O curves for all probe frequencies (Lopez-Poveda *et al.*, 2003). The masker–probe time intervals ranged from 5 to 100 ms in 5 ms steps with an additional interval of 2 ms. The duration of the masker was 110 ms, including 5 ms cosine-squared onset and offset ramps. The probe duration was 10 ms, including 5 ms cosine-squared onset/offset ramps and no steady state portion. The level of the probe was fixed at 9 dB sensation level (SL) (i.e., 9 dB above the individual absolute threshold for the probe), except for subject S5 for whom it was 15 dB SL. Stimuli were generated with a Tucker Davies Technologies Psychoacoustics Workstation (System 3) operating at a sampling rate of 48.8 kHz and with analog to digital conversion resolution resolution of 24 bits. If needed, signals were attenuated with a programmable attenuator (PA-5) before being output through the headphone buffer (HB-7). Stimuli were presented to the listeners through Etymotic ER-2 insert earphones. TMC sound pressure levels (SPL) were calibrated by coupling the earphones to a sound level meter through a Zwislocki DB-100 coupler. Calibration was performed at 1 kHz only and the obtained sensitivity was used at all other frequencies because the earphone manufacturer guarantees an approximately flat (±2 dB) frequency response between 200 Hz and 10 kHz.

C. TMC procedure

The procedure was identical to that of Lopez-Poveda and Alves-Pinto (2008). Masker levels at threshold were measured using a two-interval, two-alternative forced-choice paradigm. Two sound intervals were presented to the listener in each trial. One of them contained the masker only and the other contained the masker followed by the probe. The interval containing the probe was selected randomly. The subject was asked to indicate the interval containing the probe. The initial masker level was set sufficiently low that the subject always could hear both the masker and the probe. The masker level was then changed according to a two-up, one-down adaptive procedure to estimate the 71% point on the psychometric function (Levitt, 1971). An initial step size of 6 dB was applied, which was decreased to 2 dB after three reversals. A total of 15 reversals were measured. Threshold was calculated as the mean of the masker levels at the last 12 reversals. A measurement was discarded if the standard deviation (s.d.) of the last 12 reversals exceeded 6 dB. Three threshold estimates were obtained in this way and their mean was taken as the threshold. If the s.d. of these three measurements exceeded 6 dB, a fourth threshold estimate was obtained and included in the mean.

The maximum SPL was set to 104 dB to prevent subject discomfort and/or temporary threshold shifts. A measurement run was stopped and discarded when the subject reached this limit on more than two consecutive trials over the last 12 reversals. Masker levels at threshold were measured for masker–probe time intervals in increasing order. This was done to minimize the possibility that the measurements would be affected by potential temporary thresholds shifts that might have occurred if intervals had been presented in random order and a long interval (high masker level) immediately preceded a short interval. An attempt was made to measure masker levels for all masker–probe time intervals. Missing data indicate that the maximum output level (104 dB SPL) was reached for the time interval in question or that it was impossible within six to ten attempts to obtain three threshold estimates with s.d. ⩽6 dB.

The listeners' absolute threshold for the maskers and probes were measured using the same equipment and conditions used to measure the TMCs. At least three measurements were obtained and averaged. Results are shown in Table I. Listeners were trained in the forward-masking task for several hours, at first with a higher probe level of 15 dB SL and later with a probe level of 9 dB SL, until performance became stable. Listeners sat in a double-wall sound attenuating chamber during all measurements.

D. Inferring BM I/O functions from TMCs

BM I/O functions were inferred from TMCs by plotting the levels for the linear reference TMC against the levels for any other masker frequency paired according to masker–probe time interval (Nelson *et al.*, 2001). The off-frequency

TMC for f_P=4 kHz was used as the linear reference to infer I/O curves for all other frequencies, as suggested by Lopez-Poveda *et al.* (2003). It was sometimes necessary to extrapolate the linear references to longer masker–probe time intervals to infer BM I/O functions over the wider possible range of levels. In similar situations, some authors have fitted the linear reference TMC with a straight line (e.g., Lopez-Poveda *et al.*, 2003; 2005; Nelson *et al.*, 2001; Plack *et al.*, 2004). There is strong evidence, however, that the decay of forward masking is better described with two time constants (e.g., Lopez-Poveda and Alves-Pinto, 2008; Meddis and O'Mard, 2005; Oxenham and Moore, 1994; Plack and Oxenham, 1998). Based on this, the individual linear-reference TMCs were fitted here (using a least-squares procedure) with a double exponential function of the form:

$$L_m(t) = L_0 - 20\log_{10}[\alpha e^{(-t/\tau_a)} + (1-\alpha)e^{(-t/\tau_b)}], \quad (1)$$

where $L_m(t)$ is the masker level required to mask the probe at masker–probe time interval t; L_0 is masker level for a masker–probe interval of zero; τ_a and τ_b are time constants; and α determines the level at which the second exponential takes over from the first one. α, τ_a, τ_b, and L_0 were fitting parameters and were allowed to vary freely within certain boundaries: α was restricted to the interval [0, 1]; L_0 was restricted to vary within [50,120] dB; and τ_a and τ_b to the interval [1,200] ms. The lowest correlation between actual and predicted masker levels was r=0.94, which shows that the goodness of fit was excellent.

E. DPOAE stimuli

DPOAE I/O curves were obtained by plotting the magnitude (in dB SPL) of the $2f_1-f_2$ DP emission as a function of the level, L_2, of the primary tone f_2. DPOAEs were measured only for f_2 frequencies equal to the probe frequencies for which TMCs had been previously measured (0.5, 1, 2, and/or 4 kHz). The f_2/f_1 ratio was fixed at 1.2. L_2 ranged from 20 to 75 dB SPL in 5 dB steps, except for f_2 =0.5 kHz for which it ranged from 45 to 75 dB SPL. L_1 and L_2 were related according to $L_1=0.4L_2+39$ dB, the rule proposed by Kummer *et al.* (1998) to obtain maximum-level DPOAEs for $L_2 \leqslant 65$ dB SPL. In the current study, this rule was extrapolated to $L_2 > 65$ dB SPL.

In an attempt to reduce the potential variability of the I/O curves caused by the DP fine structure, DPOAE I/O curves were measured for five f_2 frequencies near the frequency of interest, and the resulting I/O curves were averaged. This procedure is supported by Kalluri and Shera (2001) and Mauermann and Kollmeier (2004), who showed that a "cleaned" DP-gram (i.e., a DP-gram where the fine structure has been accounted for) resembles very closely a moving average of the original DP-gram with fine structure. The five adjacent frequencies were selected to differ by as much as 2% of the frequency of interest based on a suggestion of Mauermann and Kollmeier (2004) that this frequency spacing is appropriate to reveal (or to account for) the influence of the fine structure. That meant measuring DPOAE I/O curves for frequencies of $0.98f_2$, $0.99f_2$, f_2, $1.01f_2$, and $1.02f_2$. For instance, the final DPOAE I/O curve at 4 kHz was the mean of five I/O functions for f_2 ={3920,3960,4000,4040,4080 Hz}. To further assess the potential influence of the fine structure around the frequency of interest, DPOAEs were measured for four additional adjacent f_2 frequencies on each side of the five main frequencies. In the case of 4 kHz these were: f_2 ={3760,3800,3840,3880,…,4120,4160,4200,4240 Hz}. DPOAEs for these latter frequencies were measured only for L_2={70,60,50,40 dB} and the resulting I/O curves were not included in the final mean I/O curve.

The influence of standing waves must be taken into consideration when measuring DPOAE at high frequencies. Siegel (1994) and Whitehead *et al.* (1995) found that restricting DPOAE measurements to $f_2 < 6$ kHz avoids the majority of the idiosyncratic variations of the sound pressure at the eardrum due to standing waves. For this reason, it was decided to restrict DPOAE measurements to $f_2 \leqslant 4$ kHz. Measuring DPOAEs for $f_2 < 1$ kHz is also problematic due to increased physiological subject noise. The moderate-to-high level part of the DPOAE I/O curve for f_2=0.5 kHz could still be measured for most listeners, however, by considering L_2 above 45 dB SPL only.

F. DPOAE stimulus calibration

DPOAE stimuli were calibrated with a Zwislocki DB-100 coupler for each f_1,f_2 frequency. In some studies, calibrated levels are further adjusted with the probe *in situ* to account for the acoustic effects associated with ear-canal resonances. This *in situ* adjustment, however, was not applied here for two reasons. First, Siegel (1994) has shown that it does not always work for frequencies above 2–3 kHz because of the errors in predicting the level at the eardrum from measurements made at the plane of the probe (where the standing waves interact). Second, as this is a comparison study, it was deemed important that the two methods applied the same stimulus level control. Given that the psychophysical equipment did not allow easy *in situ* level adjustment, this option was disabled in the OAE instrument.

G. DPOAE system artifacts

When measuring DPOAE I/O curves at high primary levels it is necessary to control for cubic distortion produced by the measurement instrument. This system artifact was assessed by measuring the magnitude of the $2f_1-f_2$ DP in two different couplers: a DB-100 Zwislocki coupler and a plastic syringe having a volume of approximately 1.5 cc. The test was performed for L_2 from 50 to 80 dB SPL with the same equipment and in the same conditions used to measure DPOAEs. The measurement time was prolonged to maximize the chances of discovering any artifacts. The magnitude of the cubic DP would be $-\infty$ dB SPL for an ideal OAE system. The system artifact limit was set to the higher of the responses for the two couplers that was also 2 s.d. above the mean level of ten adjacent frequency bins in its corresponding spectrum. This procedure was repeated for each of the f_2s considered when measuring DPOAEs.

It is common to accept DPOAE measurements when they are above the system artifact limit. This may be toler-

able in a clinical context where it comes to making a "pass/refer" decision. The magnitude of the measured DPOAE is the *vector* sum of DP contributions from any nonlinearity along the signal path, be it from the instrument or from the subject. If the clinical rule was applied, then the true physiological response would be any value within the range $(-\infty, +6]$ dB around the measured DPOAE in the worst possible case (i.e., when the measured DPOAE magnitude just exceeds the artifact limit and the physiological DP has opposite or equal phase to the system's DP, respectively). This uncertainty range, however, seems too broad for the present study where the slope of the I/O curve is of interest. Therefore, a more restrictive rule was applied. A DPOAE measurement was accepted as valid only when it exceeded the artifact limit by 6 dB or more. This guaranteed that the physiological DP contribution was within the range $[-6, +3.5]$ dB in the worst possible case (i.e., when the DPOAE measurement just met the present criterion). Therefore measurements were rejected if they were less than 6 dB above the system artifact limit. This was the case for subject S4 at 4 kHz, for whom most data points were discarded based on this criterion, and also for subject S3 at 1 kHz at high stimulus levels (75 dB SPL).

H. DPOAE procedure

DPOAE measurements were obtained with an IHS Smart system (with SmartOAE software version 4.52) equipped with an Etymotic ER-10D probe. During the measurements, subjects sat comfortably in a double-wall sound attenuating chamber and were asked to remain as steady as possible.

The probe fit was checked before and after each measurement session. The probe remained in the subject's ear throughout the whole measurement session to avoid measurement variance from probe fit. DPOAEs were measured for a preset measurement time. For f_2=4 kHz, the measurement time ranged from 60 s at L_2=20 dB SPL to 8 s at high L_2; for f_2=0.5 kHz, it ranged between 60 and 30 s for L_2 between 45 and 75 dB SPL, respectively. A DPOAE measurement was considered valid when it was 2 s.d. above the measurement noise floor (defined as the mean level over ten adjacent frequency bins in the spectrum). When a response did not meet this criterion, the measurement was repeated and the measurement time was increased if necessary. The probe remained in the same position during these remeasurements. If the required criterion was not met after successive attempts, the measurement point was discarded.

Each recording session consisted of measuring DPOAE I/O curves for one frequency of interest (consisting of five adjacent frequencies) (see Sec. II E) and was allowed to take up to 1 h. Three DPOAE measurements were obtained per condition (i.e., per f_2 and L_2) and averaged, except for subject S2 for whom only one measurement was obtained. Therefore, each point in the final I/O curves was the mean of 15 (3 measurements $\times$ 5 adjacent frequencies) measurements for each L_2 level. Occasionally, it was not possible to obtain all 15 points, particularly for the lower L_2 levels. In those cases, the direct mean of the available points would have been biased toward the DPOAE values for the frequencies giving the stronger responses. To minimize this potential bias, a mean was calculated only when two (out of the three possible) measurements were available per frequency and when eight of the new ten possible measurements were available. Otherwise, the corresponding point was neglected for further analysis.

III. RESULTS

A. TMCs

Figure 1 shows the TMCs for the ten subjects for probe frequencies of 0.5, 1, 2, and 4 kHz. Note that TMCs were not measured for all frequencies and for all subjects. Subject S5 was unable to perform the TMC task using a probe level of 9 dB SL, thus a probe level of 15 dB SL was used in that case. Even so, it was not possible to measure high masker levels at 4 kHz for subject S5 and the resulting TMC did not allow estimating the true degree of compression (open circles in panel S5 of Fig. 1). Therefore these data were discarded from further analysis.

The shapes of the present TMCs are generally consistent with those of previous studies (e.g., Nelson *et al.*, 2001; Lopez-Poveda *et al.*, 2003; Lopez-Poveda and Alves-Pinto, 2008; Plack *et al.*, 2004; Rosengard *et al.*, 2005). The linear-reference TMCs (open squares) can be described either by a straight line (e.g., S4, S5) or by a shallow and gradually saturating function (e.g., S3, S8). The latter justifies the decision to fit the linear-reference TMCs with a double-exponential function. The continuous, thick lines illustrate these fits. Several on-frequency TMCs show a shallow segment for short masker–probe time intervals (or gaps) followed by a steeper segment for moderate intervals. Others, however, are better described by a segment steeper than the linear reference followed sometimes by a shallower section at high masker levels (e.g., S1, S2, S6, and S8). In any case, all of the on-frequency TMCs have segments that are much steeper than the linear reference TMC. Assuming that the off-frequency masker condition used to generate the linear reference is processed linearly by the BM and that the rate of decay of the internal masker effect is identical across frequencies (Lopez-Poveda *et al.*, 2003; Lopez-Poveda and Alves-Pinto, 2008), the steeper segments may be interpreted to indicate BM compression (Nelson *et al.*, 2001). The validity of these assumptions is discussed in Sec. IV F.

B. DPOAEs

Figure 2 shows a typical example of the amount of data measured to estimate one DPOAE I/O curve (2 kHz in this particular case). Each data point is the average of three measurements. Figure 2 also serves to illustrate the influence of the DP fine structure on the resulting I/O function. The inset indicates L_2 (in dB SPL). Note that DPOAEs were measured for a wider range of stimulus levels (from 20 to 75 dB SPL) for the five central adjacent frequencies only. A narrower range of levels (40, 50, 60, and 70 dB SPL) was considered for frequencies outside this frequency range. This was suffi-

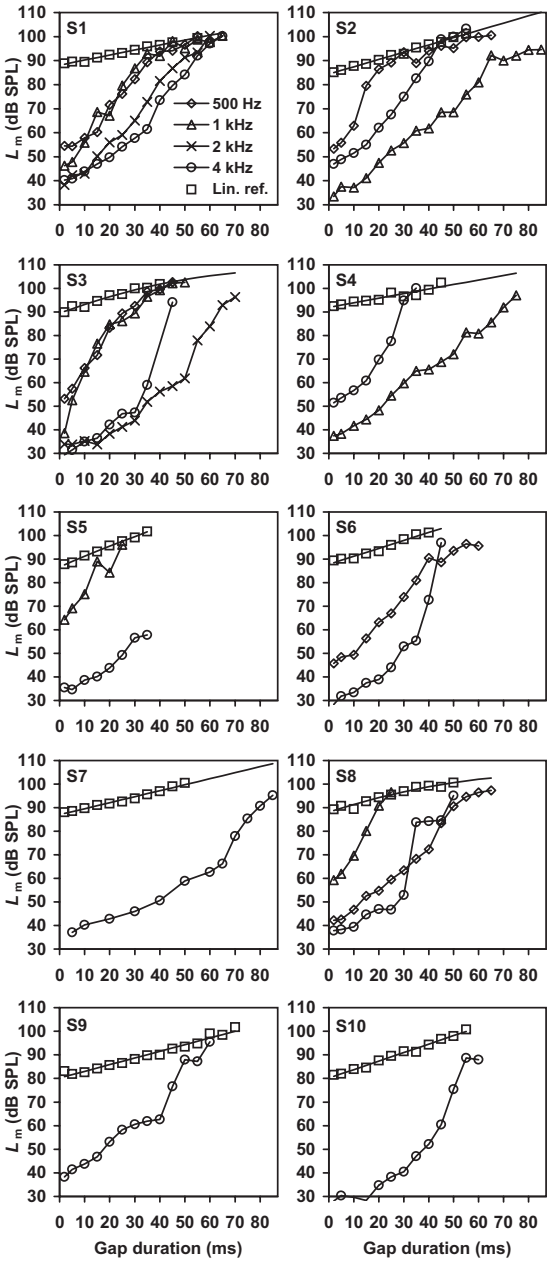


FIG. 1. TMCs for all subjects and probe and masker frequencies. Each panel illustrates the data for one subject. Open squares illustrate the linear reference TMC; i.e., the TMC for a probe frequency of 4 kHz and a masker frequency ($0.4f_P$) of 1.6 kHz. The smoother continuous lines illustrate fits to the linear reference TMC with a double-exponential function. Other symbols illustrate on-frequency TMCs for different probe frequencies (as indicated by the inset in the top-left panel). The probe level was 9 dB SL except for subject S5 for whom it was 15 dB SL.

cient, however, to get an idea of the surrounding DPOAE fine structure and its potential influence on the DPOAE I/O at the frequency of interest.

Obviously the DPOAE I/O curve would change considerably by changing f_2 within a narrow frequency range of 100 Hz, which emphasizes the need to take into account the effect of fine structure when estimating the actual I/O curve. The final I/O curve, representing the BM I/O function at 2 kHz, would be the mean of the five I/O curves for the five central frequencies. Figure 3 illustrates these five I/O curves for subjects S1 and S2 and for all frequencies considered in

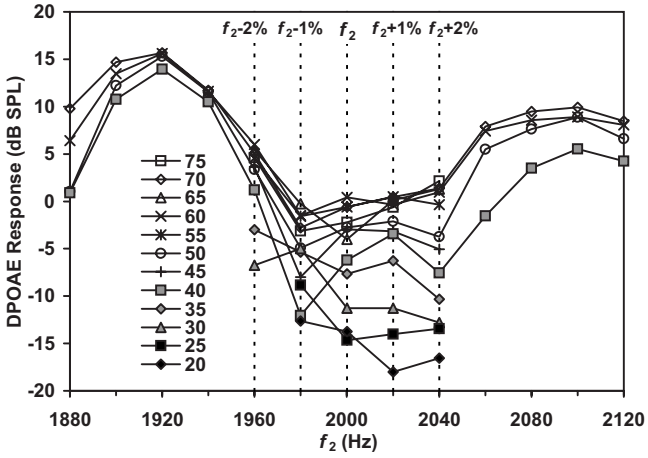


FIG. 2. An example of the influence of the DPOAE fine structure at 2 kHz. An example data set recorded to estimate every I/O curve is also illustrated. The DPOAE magnitude is shown for 12 f_2 frequencies around the frequency of interest (2 kHz) and for different L_2 levels (in dB SPL) as indicated in the inset.

the present study. As can be seen, there is variability across subjects and across frequencies. Nevertheless, variability appears to be greater for the lower stimulus levels. This seems reasonable given that the fine structure is more obvious at low stimulus levels (Mauermann and Kollmeier, 2004).

C. Comparison of BM I/O curves inferred from TMCs and DPOAEs

Figures 4–7 allow within-subject comparisons of DPOAE I/O curves (open squares) with BM I/O curves inferred from TMCs (closed circles) for corresponding cochlear sites with CFs of 0.5, 1, 2, and 4 kHz, respectively. Each panel illustrates the results for one subject. The associated solid lines are third-order polynomials fitted to the data. Error bars denote 1 standard error (s.e.) of the mean. For DPOAE I/O curves, the s.e. was based on up to 15 measurement points for each stimulus level, which explains why error bars are so short. The average error across all subjects and levels was 0.7 dB. The vertical and horizontal error bars of TMC-based I/O curves illustrate the s.e. of the linear reference or the TMC in question, respectively, based on at least three measurements. The average errors across all subjects and time gaps for the linear reference and the on-frequency TMCs were 0.9 and 2.3 dB, respectively.

The open triangles in the top-right panel of Figs. 4–7 and their associated error bars illustrate the mean DPOAE noise floor plus 2 s.d. (based on three measurements for the five adjacent frequencies considered) for one example subject. The noise levels were similar for the other subjects. In general, the noise has the effect of increasing the DPOAE magnitude, particularly at low levels. However, the strict criteria used here should have avoided this influence.

In general, both DPOAE and TMC-based I/O curves are similar in that they are linear at low levels and become gradually compressive with increasing level. There is a tendency in both sets for I/O curves to become linear again at the highest levels tested (e.g., S6 at 0.5 kHz in Fig. 4; or S2 and S5 at 4 kHz in Fig. 7). The degree of similarity between

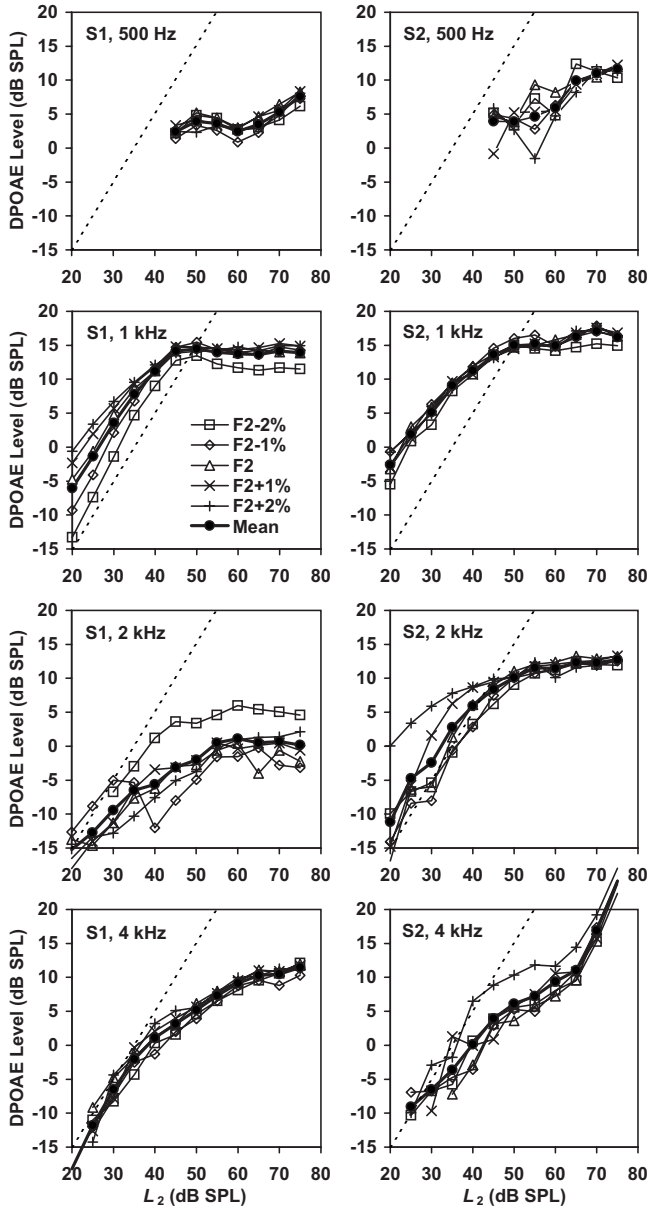


FIG. 3. Example DPOAE I/O curves for subjects S1 (left panels) and S2 (right panels), for frequencies from 0.5 (top) to 4 kHz (bottom) in octave steps. Each final I/O curve (closed circles, thick line) was obtained as the average of five I/O curves for frequencies of $0.98f_2$, $0.99f_2$, f_2, $1.01f_2$, and $1.02f_2$, illustrated with different symbols according to the inset. The dashed line illustrates a linear response for comparison.

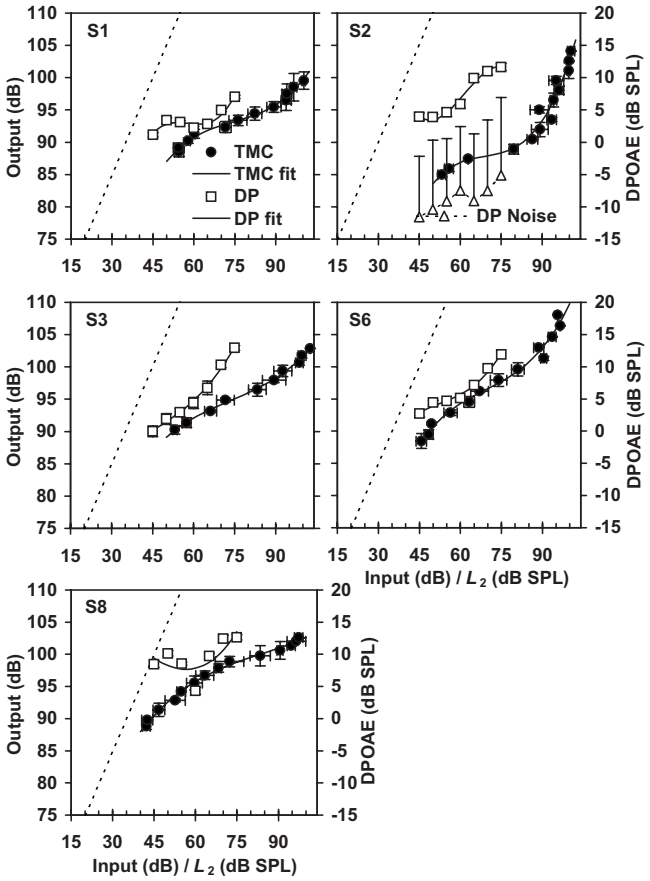


FIG. 4. Experimental DPOAE (open squares) and TMC-based (closed circles) I/O curves at 0.5 kHz ($=fp=f_2$). Continuous lines illustrate third-order polynomial fits to the experimental I/O curves. Error bars denote 1 s.e. of the mean. Horizontal bars (only for TMC-based curves) represent the standard error of the input level (i.e., the standard error for the on-frequency masker level). Each panel illustrates the result for a different subject. The panel for subject S2 also illustrates the mean DP noise floor and its corresponding 2 s.d. Thin dashed lines illustrate a linear response for comparison.

the two sets of I/O curves is greatest at 4 kHz (Fig. 7). At this frequency, both sets of I/O curves indicate equally mild compression for subjects S1, S2, S5, and S7, and equally strong compression for S3 and S8. The strongest disagreement between the two sets of I/O curves at 4 kHz occurs for subject S9 (Fig. 7). S4 deserves a special mention because her DPOAEs could not be measured for levels outside the 40–50 dB SPL level range despite her having normal hearing at 4 kHz (her DPOAE readings for $L_2<40$ dB SPL were below the physiological noise level and for $L_2>50$ dB SPL they did not meet the instrument artifact criterion, see Sec. II G).

The degree of similarity between the shapes of the two sets of I/O curves is, however, much lower for 0.5 and 1 kHz. It is noteworthy that, for these frequencies, some DPOAE I/O curves show plateaus and notches that do not have a clear correlate with TMC-based I/O curves (e.g., S1, S2, and S8 at 0.5 kHz in Fig. 4; or S3, S4, and S5 at 1 kHz in Fig. 5). Possible explanations for the notches and plateaus are discussed in Sec. IV B.

It was not possible to draw conclusive results regarding the degree of correlation between the two sets of I/O curves at 2 kHz because data at this frequency were collected for two subjects only (Fig. 6). Good correspondence (akin to what was observed at 4 kHz) was found for one of the two subjects (S3). The DPOAE I/O curve for the other subject (S1), however, exhibited a negative gradient at moderate-to-high levels that was not present in the TMC-based I/O curves. This negative gradient was typical of the I/O curves at frequencies of 0.5 and 1 kHz.

D. Comparison of derived cochlear nonlinearity parameters

Third-order polynomials were fitted by least squares to the TMC-based and DPOAE I/O curves (continuous lines in Figs. 4–7) and used to derive the following parameters pertaining to cochlear nonlinearity: minimum compression ex-

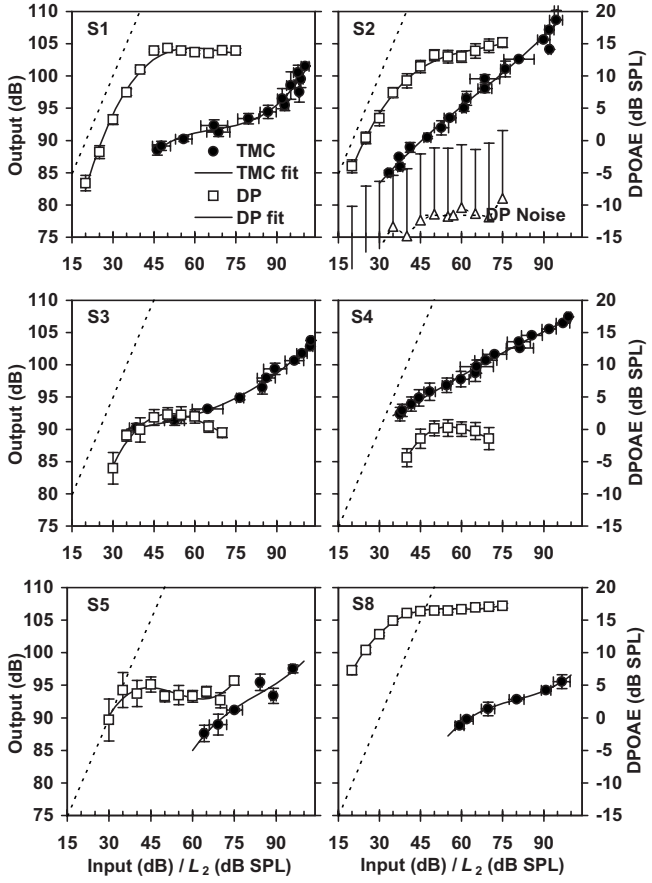


FIG. 5. The same as Fig. 4, but for a frequency of 1 kHz.

ponent, the level at which it occurred, compression threshold, cochlear gain, and the level at which the I/O curves returned to linearity at high levels.

The minimum compression exponent was estimated as the minimum slope of the fitted I/O curves over the measured range of input levels. Figure 8(A) compares the minimum compression exponent as inferred from the I/O curves obtained with the two methods. Closed symbols show the results for 4 kHz separately. Open symbols illustrate the results for the other frequencies (0.5, 1, and 2 kHz), as indicated by the inset. The solid and dashed lines illustrate linear regression fits to the 4 kHz data only or to all data points (including also the 4 kHz points), respectively. Both regression lines were forced to pass through the origin of the graph. Pearson's correlation coefficients were (high) 0.92 for the 4 kHz subsample and (low) 0.19 when all data points (all

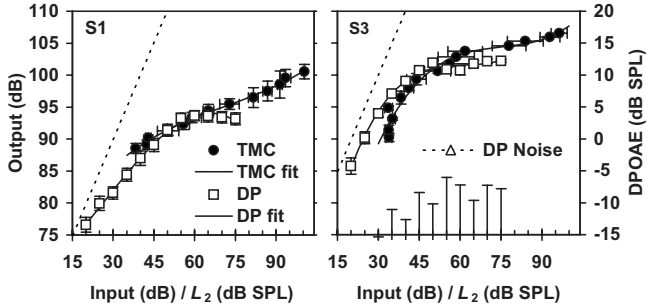


FIG. 6. The same as Fig. 4, but for a frequency of 2 kHz.

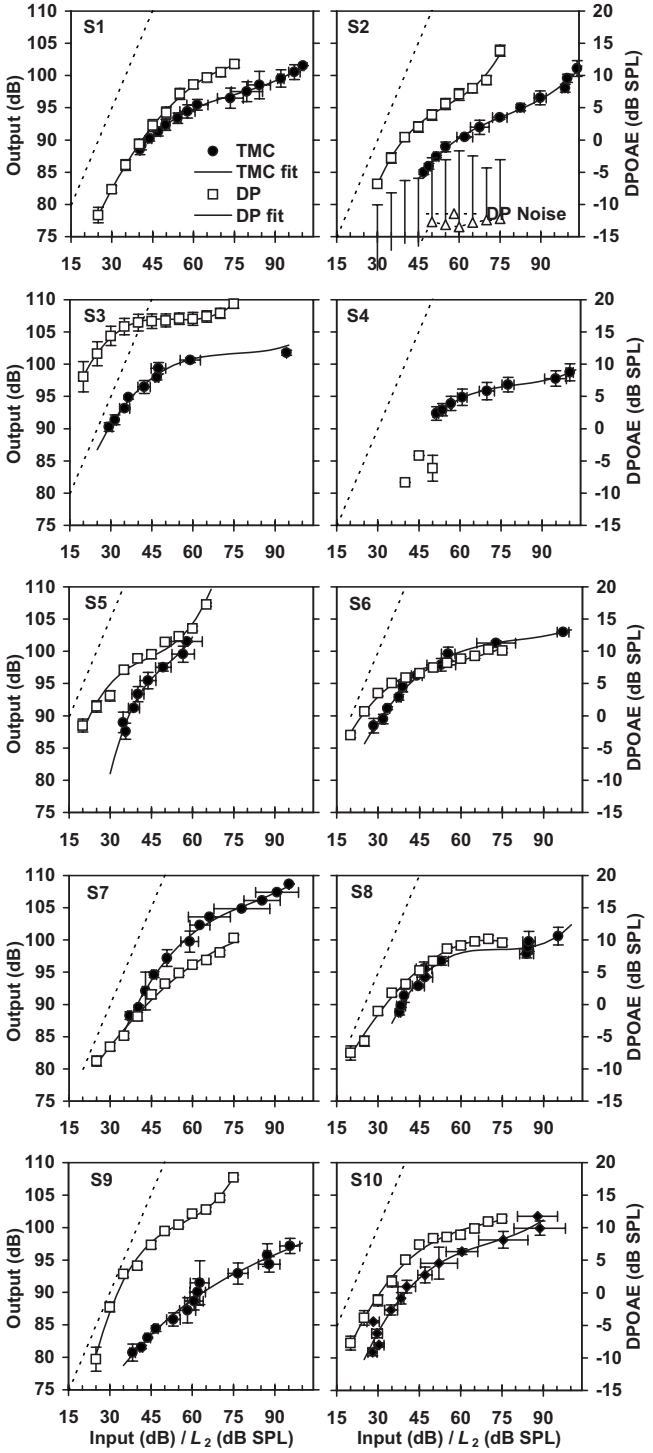


FIG. 7. The same as Fig. 4, but for a frequency of 4 kHz.

frequencies) were considered. The mean compression exponents across frequencies were 0.13 and 0.04 for the TMC-based and the DPOAE data sets, respectively. They, however, were 0.10 and 0.11, respectively, considering only the 4 kHz data points. The mean difference was significant when all data points were considered ($p<0.005$ for a paired two-tailed Student's t-test) but not for the 4 kHz subsample ($p=0.81$). This is clearly seen from Fig. 8(A) in that DPOAE-based compression estimates are lower overall than TMC-based ones for 0.5 and 1 kHz.

The reason for the disagreement between the results at

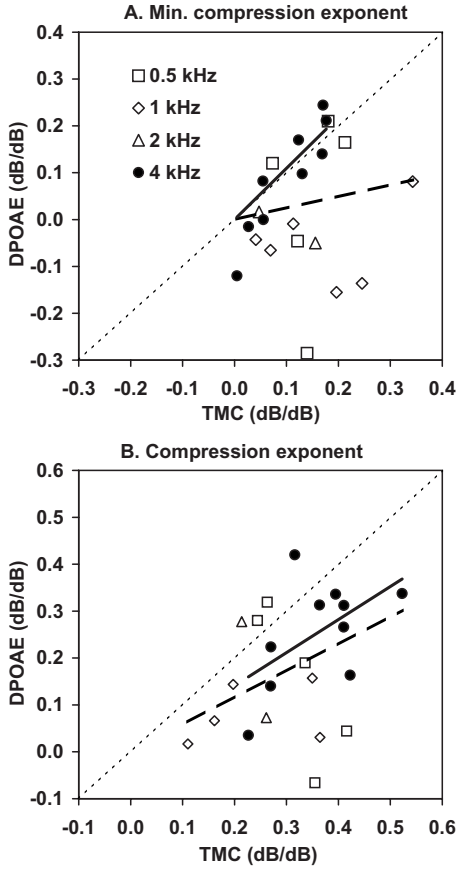


FIG. 8. Correlation between compression-exponent estimates obtained from DPOAE- and TMC-based I/O curves (i.e., each point represents data for one subject). (A) Minimum compression exponent based on third-order polynomial fits to the I/O curves. (B). Compression exponent based on linear-regression fits to the I/O curves over the input-level range 40–65 dB SPL. Closed and open symbols illustrate the 4 kHz data and the data at other frequencies (0.5, 1, and 2 kHz), respectively. Solid lines show linear regression fits to the 4 kHz data points constrained to cross through the origin. Thick dashed lines are linear regression fits considering all data points (including 4 kHz) constrained to cross through the origin. Diagonal thin-dashed lines illustrate perfect correlation.

0.5 and 1 kHz is, almost certainly, that many DPOAE I/O curves for the lower frequencies showed plateaus or notches, which resulted in unexpectedly low compression exponents (slopes <0 dB/dB). Table II summarizes the number of cases across frequencies when the fitted I/O curve exhibited plateaus or negative compression exponents.

These results contrast with those of Williams and Bacon (2005), who reported a high correlation between compression estimates inferred from TMC-based and DPOAE I/O curves also at low frequencies. The compression estimates of Williams and Bacon (2005), however, were calculated differently. They were calculated as the slope of linear segments fitted to the DPOAE and TMC-based I/O curves over the range of input levels from 40 to 65 dB approximately (they determined the actual level range by visual inspection). This fitting method was also employed here as it could prove less sensitive to notches. Figure 8(B) shows the correlation between compression exponent estimates for the two methods derived by linear regression fits to the I/O curves over the level range 40–65 dB. The 4 kHz data are shown by the closed symbols; open symbols illustrate the results for frequencies of 0.5, 1, and 2 kHz. The solid and dashed lines are linear regression fits to the 4 kHz data and to all data points (including also the 4 kHz data), respectively. Pearson's correlation coefficient considering all data points was 0.32 and Fig. 8(B) illustrates that DPOAE-based compression exponents were generally lower (most points are below the diagonal) than those inferred from TMCs. Group mean compression exponents were 0.31 and 0.18 for TMCs and DPOAEs, respectively. This difference was statistically significant ($p<0.0005$). A moderately higher correlation was found when analyzing the 4 kHz data separately ($r=0.53$), but the associated mean difference (0.36 and 0.25 for TMCs and DPOAEs, respectively) was still statistically significant ($p<0.01$).

TABLE II. Incidence of plateaus and notches in DPOAE I/O curves as a function of f_2.

	Frequency (kHz)			
	0.5	1	2	4
Number of I/O measured curves	5	6	2	10
Number of I/O curves with plateaus or notches	3	5	1	2

Another characteristic of cochlear nonlinearity is the level at which maximum compression (or, equivalently, minimum compression exponent) occurs. Correlations were sought between estimates of this parameter for the I/O curves obtained with the two methods. Figure 9(A) reveals that high correlation occurs in a few cases only (points close to or on the diagonal). Most points are below the diagonal, which means that the level at which minimum compression occurred was lower for DPOAE I/O curves than for I/O curves inferred from TMCs. The average input level for minimum compression exponent was 61 and 76 dB for DPOAEs and TMC-based I/O curves, respectively. The difference was statistically significant ($p<0.001$). Results were similar when the data for each frequency were analyzed separately.

Compression threshold was defined as the input level at which the slope of the fitted polynomial decreased from a value close to one at low levels to 0.4 dB/dB at a higher level. This is an arbitrary definition, but seems reasonable for our purpose. When the slope of the I/O curve at the lowest level for which a data point existed was reasonably close to 0.4 dB/dB, the fitted polynomial was extrapolated up to 5 dB to identify the compression threshold and the value was noted for further analysis. An extrapolated compression estimate was thus included in the analysis only if it was less than 5 dB from an existing data point. Figure 9(B) illustrates the correlation between the compression threshold estimates inferred from DPOAE and TMC-based I/O curves. Three compression estimates were considered outliers [symbols surrounded by an open circle in Fig. 9(B)] and thus excluded from the statistical analysis. The possible reasons for these outliers will be discussed in Sec. IV D. The regression line was constrained to cross through the graph origin. The line indicated a high degree of correlation between the estimates of the two methods and, indeed, Pearson's correlation coefficient was reasonably high ($r=0.8$). The average compression threshold estimates were lower for the DPOAE than for

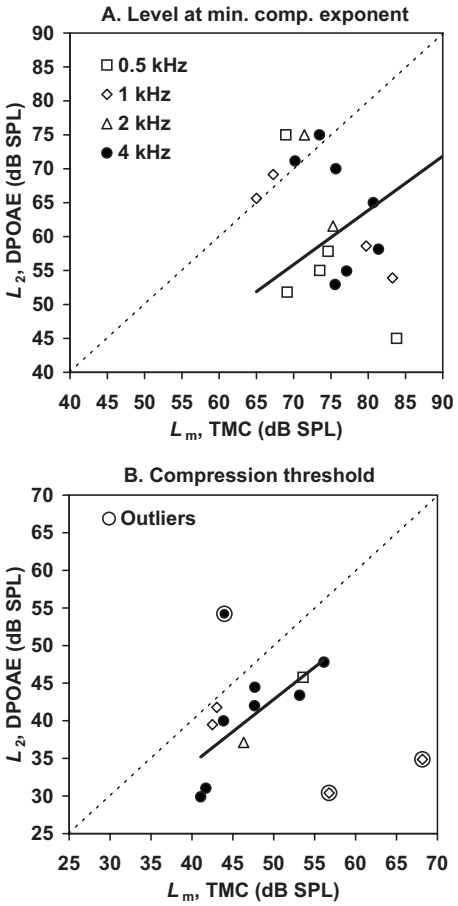


FIG. 9. (A) Correlation between the level at which maximum compression occurs based on the I/O curves inferred from DPOAE and TMCs (i.e., each point represents data for one subject). The solid line shows a linear regression fit to the data constrained to cross through the origin. (B) Correlation between compression-threshold estimates obtained with the two methods. Threshold was defined as the input level at which the I/O curve slope decreased from linearity to 0.4 dB/dB. The solid line shows the linear regression constrained to cross through the origin. Symbols surrounded by open circles illustrate three data points regarded as outliers. Diagonal thin-dashed lines illustrate perfect correlation.

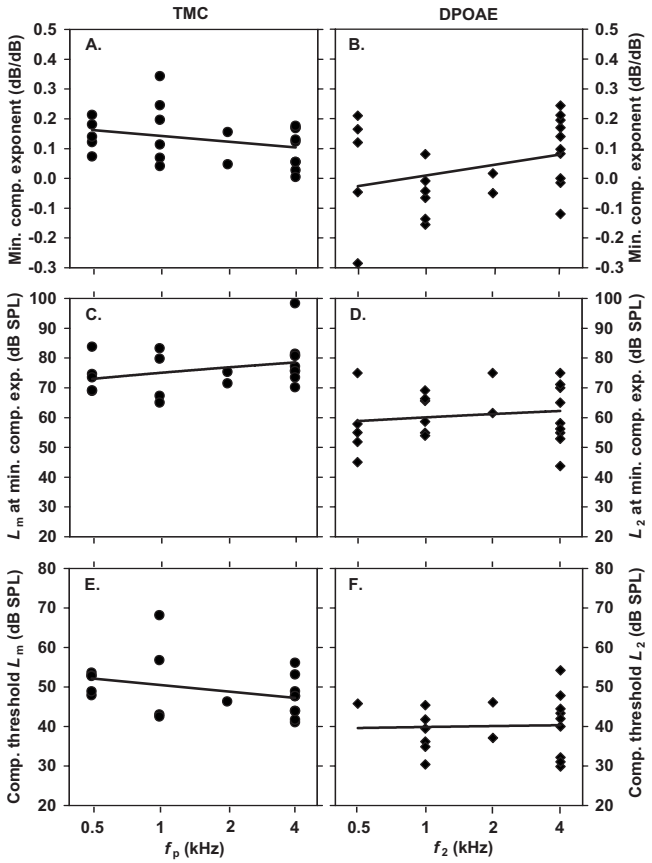


FIG. 10. Frequency dependency of cochlear nonlinearity parameters as estimated from TMCs (left panels) and DPOAEs (right panels). (A), (B) Minimum compression exponent. (C), (D) The level at which maximum compression occurs. (E), (F) Compression threshold.

the TMC-based I/O curves (40 and 47 dB SPL, respectively), which is also clearly seen in Fig. 9(B). This difference was statistically significant ($p<0.0001$).

E. Cochlear nonlinearity dependency on characteristic frequency

This section addresses whether the parameters considered in the preceding section vary with frequency similarly when they are inferred from TMCs or DPOAEs. Two additional parameters were considered based on TMCs, namely gain and the threshold of return to linearity at high levels. I/O curves may be generally described as having a linear segment (slope of 1 dB/dB) at low levels, followed by a compressive segment (slope <1 dB/dB) at moderate levels, followed by linear segment at high levels (e.g., Lopez-Poveda *et al.*, 2003). Perhaps, the clearest examples of this pattern were the TMCs for S6 at 0.5 kHz and S2 at 4 kHz. Few TMC-based I/O curves showed a linear segment at high levels (e.g., S2 at 0.5 kHz or S1 at 1 kHz). The slope of the I/O curves, however, always increased with increasing level beyond the inflection point of the curve (defined as the level at which maximum compression occurs or, equivalently, the level at which the second derivative of the I/O curve equals zero). This suggests that the I/O curves might approach linearity at very high levels. Since the minimum slopes were always <0.4 dB/dB [Fig. 8(A)], the return-to-linearity threshold was arbitrarily defined here as the level at which the slope of the fitted I/O equated to 0.4 dB/dB for levels above the inflection point. Gain was defined as the difference between the return-to-linearity and the compression thresholds (in decibels).

Figure 10 shows the frequency dependency of the minimum compression exponent (top panels), the level at which it occurs (middle panels), and the compression threshold (lower panels), respectively. Left and right panels show the value of these parameters as inferred from TMCs and DPOAE I/O curves, respectively. Figure 10 illustrates that all these parameters remain approximately constant across frequencies. Indeed, no significant differences were found between the mean values of every parameter across frequencies. The same applied to parameters estimated with both methods.

Figure 11 shows the gain and the return-to-linearity threshold as inferred from TMCs only. Both parameters tend to increase with increasing frequency. Since the compression threshold remained approximately constant across frequencies [Fig. 10(E)], the frequency dependency of the gain is

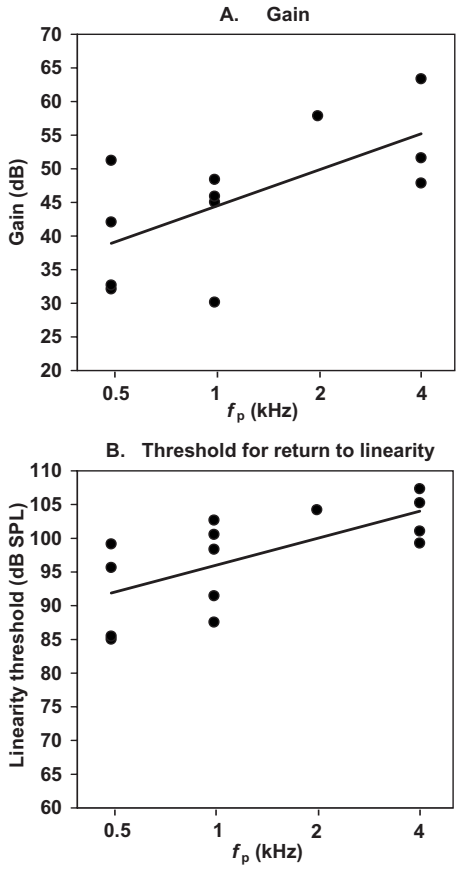


FIG. 11. Frequency dependency of gain (A) and the threshold of return to linearity (B) derived from TMC I/O curves. The latter was defined as the input level at which the I/O curve reached a slope of 0.4 dB/dB from a lower value for increasing input level.

fully attributed to the increase of the return-to-linearity threshold with increasing frequency. In any case, a one-to-one correspondence between these two parameters should not be expected because they are based on data for different subjects. Indeed, it was not possible to estimate the return-to-linearity threshold in several cases (one different subject for each frequency of 0.5, 1, and 2 kHz, and six subjects at 4 kHz). The I/O curves for these subjects may show a return to linearity at input levels higher than those considered in the present study. That the majority of these cases occurred at 4 kHz supports the idea that the return to linearity at higher frequencies occurs at higher input levels than those considered here.

IV. DISCUSSION

The goal of this study was threefold. The first objective was to compare cochlear nonlinearity parameters inferred from TMCs and DPOAEs, and, if coinciding, add support to the notion that they are two equivalent manifestations of cochlear nonlinearity. A second aim was to evaluate the feasibility of using DPOAE I/O curves as a fast tool for estimating individual parameters of cochlear nonlinearity. A third objective was to investigate the frequency dependency of parameters describing cochlear nonlinearity, as inferred from DPOAEs and TMCs.

A. Equivalence between DPOAE and TMC-based I/O curves

The degree of correlation between compression estimates inferred with the two methods was high (0.92) at 4 kHz but much lower at 0.5 and 1 kHz. The assumption has been made that DPOAE I/O curves reflect the characteristics of the BM response to single tones at the f_2 place. Measuring DPOAEs, however, requires presenting the two primaries (f_1 and f_2) *simultaneously*; hence the f_1 primary may have suppressed the BM response to the f_2 primary at the f_2 site. Indeed, Rhode (2007) measured BM excitation and DPOAEs simultaneously in the same preparation and showed that the f_1 primary suppresses the BM response to the f_2 (with a fixed level of 60 dB SPL) for L_1 above 60 dB SPL (see his Fig. 1). As a result, DPOAE I/O curves may not correspond directly to single-tone BM I/O curves, as is commonly assumed.

Unlike DPOAE I/O curves, the I/O curves inferred from TMCs would not be affected by suppression because the masker and the probe tones were not presented simultaneously (in fact, this is one of the reasons that they are so widely used to infer BM I/O curves). Therefore, one might think that mutual suppression between the primary tones may have influenced DPOAE but not TMC-based I/O curves and that the effect would be more pronounced at low CFs because the nonlinear effects extend to a wider bandwidth (Rhode and Cooper, 1996; Lopez-Poveda *et al.*, 2003). This explanation, however, is unlikely to account for the low correlation between compression estimates obtained with the two methods at low frequencies. There is physiological and psychophysical evidence that suppression leads to I/O curves steeper than single-tone I/O curves (Nuttall and Dolan, 1993; Rhode, 2007; Yasin and Plack, 2007). This is true particularly for suppressor/suppressee combinations similar to the primary-tone combinations used here. If suppression had affected the DPOAE I/O curves, they should indicate less compression than TMC-based I/O curves and this has been found *not* to be the case (Fig. 8). Therefore, the most likely explanation for the low correlation at low CFs between the compression estimates obtained with the two methods is the presence of notches and plateaus in the DPOAE I/O curves, which occur more frequently at low CFs (Table II).

The group mean compression-exponent estimates at 4 kHz obtained in the present study (0.10 and 0.11 dB/dB for TMC and DPOAE, respectively) are in agreement with those from Gorga *et al.* (2007), who reported a minimum slope value of ~0.12 dB/dB at 4 kHz based on DPOAEs (estimated from their Fig. 4 at L_2=60 dB SL). As for compression estimated using linear regression over the midlevel range [Fig. 8(B)], the group mean exponent values obtained in the present study (0.36 and 0.25 for TMC and DPOAE, respectively) were moderately higher than those reported by Williams and Bacon (2005) (0.26 and 0.15 for TMC and DPOAE at 4 kHz, respectively). In any case, the present study shows that the estimated degree of compression differs considerably depending on the method used to infer it (e.g., polynomial versus linear regression fits), which emphasizes the need to specify clearly the method used in every study.

The minimum compression exponent found here at 4 kHz (based on polynomial fits) is lower than previously

reported values obtained with the same (or different) methods [e.g., 0.14, third-order polynomial, Nelson and Schroder (2004), 0.20, third-order polynomial, Plack and Drga (2003), 0.13, sum of linear and sigmoidal function, and 0.23, straight line, Rosengard *et al.* (2005), 0.20, straight line, Plack *et al.* (2004), 0.25, straight line, Lopez-Poveda *et al.* (2003)]. The difference in TMC-based estimates may relate to differences in the linear reference used by different studies. Here, the linear reference was the TMC for a masker frequency of $0.4f_P$, whereas the above-mentioned studies used the TMC for a masker frequency between $0.5f_P$ and $0.6f_P$. Lopez-Poveda and Alves-Pinto (2008) have suggested that the latter may still undergo as much as 2:1 compression; hence they could lead to an underestimate of the degree of on-frequency compression. The agreement between the present compression estimates obtained from TMCs and DPOAEs at 4 kHz provides circumstantial support to the conclusion of Lopez-Poveda and Alves-Pinto (2008).

B. DPOAE notches and plateaus

Notches and plateaus are common in the present DPOAE I/O curves (Figs. 4–7, especially at the lower f_2 (Table II). This contrasts with the conclusion of Kummer *et al.* (1998), who reported that notches and plateaus were less common when DPOAE I/O curves were measured with their proposed primary-level rule than with other rules. The Kummer *et al.* level rule is based on a group average and thus in some cases it may deviate considerably from the individual optimal (the optimal rule would be the one that evokes the strongest possible DPOAEs at all levels). Therefore, one possible explanation for the present observations is simply that the rule of the Kummer *et al.* (1998) was not optimal for the subjects used in the present study. This explanation is supported by Neely *et al.* (2005), who reported that L_1 should be systematically higher than prescribed by Kummer *et al.* (1998) and that the L_1–L_2 relationship should vary with f_2. Johnson *et al.* (2006) confirmed the latter and further suggested that the f_2/f_1 ratio should vary slightly with f_2. Interestingly, it is for low frequencies and moderate–high stimulus levels where the rule of Kummer *et al.* deviates most from the rules of Neely *et al.* (2005) and Johnson *et al.* (2006). These are also the conditions where plateaus and dips are most common in the present data, which suggests that the rule of Kummer *et al.* is not optimal for the subjects considered in this study at low frequencies. On the other hand, Kummer *et al.* (2000) verified their original paradigm [derived from data of Gaskill and Brown (1990)] with a larger sample and still found it to be independent of frequency.

A second explanation for the plateaus and notches is that some of the DPOAE I/O curves still could have been influenced by the fine structure despite the precautions taken to minimize its effects. Plateaus occur for levels around 45–50 dB SPL (e.g., S1 and S8 at 1 kHz in Fig. 5; S3 at 4 kHz in Fig. 7), where the fine structure certainly can have influence. This is, however, an unlikely explanation for the notches because they always occurred at moderate-to-high levels (60–70 dB SPL) and the fine structure has a higher influence at low levels (Mauermann and Kollmeier, 2004).

A third explanation might be that another DP generation mechanism starts playing a role at high stimulus levels and notches reflect destructive interference between the DPs generated by this "new" high-level source and the normal source (see Mills, 1997). The measurement system is unlikely to be the source in question because a rather strict exclusion criterion was applied in the present study to eliminate system-generated DPs. Liberman *et al.* (2004) showed that genetically modified mice without the necessary prestin protein to drive outer-hair-cell electromotility still generated attenuated DP responses and thus supported the existence of a second possible DP-generation mechanism at high levels. Some of their DPOAE I/O curves showed notches at similar stimulus levels to the notches found in the current study. On the other hand, Avan *et al.* (2003) attributed low and high stimulus level DPOAEs to the same nonlinear mechanism. Also, Lukashkin *et al.* (2002) and Lukashkin and Russell (2002) have shown that a single saturating nonlinearity is sufficient to explain a notch in a DP I/O function.

C. The level at which the minimum compression exponent occurred

Low correlation was found between estimates of this parameter obtained with the two methods (DPOAEs and TMCs). The reason for this is uncertain. Maybe DPOAEs grow faster with increasing stimulus level at high stimulus levels because of the contribution from the second high level DP generation source discussed in the preceding sections.

D. Compression threshold

A moderately high correlation was found between compression-threshold estimates inferred with the two methods [Fig. 9(B)]. DPOAE-based estimates were, however, on average 7 dB lower than TMC-based estimates. A total of 22 I/O curves were measured in the ten subjects, but the compression threshold could be estimated in only 14 of these 22 cases. This could be interpreted as an argument against the apparent equivalence of the two methods with respect to estimating this parameter. Further analysis reveals, however, that there are good reasons why a compression threshold could not be estimated in the remaining eight cases. Four of them corresponded to DPOAE I/O curves at 0.5 kHz (Fig. 4) that extended over a range of input levels above 45 dB SPL that was too narrow to reveal a compression threshold. Another case was S4 at 4 kHz (Fig. 7), who did not have sufficiently strong DPOAE despite her hearing threshold being normal at that frequency. For the three remaining cases, the TMCs did not show a clear compression threshold or did not reach the criterion slope of 0.4 dB/dB (see Sec. III D). These cases were S1 at 2 kHz in Fig. 6; S3 and S4 at 1 kHz in Fig. 5. Three of the 14 cases where both methods demonstrated a compression threshold were considered outliers and excluded from the comparison of compression threshold: S1 at 4 kHz, and S5 and S8 at 1 kHz [depicted as circles in Fig. 9(B)]. The data for S5 and S8 were excluded because these listeners had great difficulties performing the TMC task (they needed six to eight attempts at high masker levels to obtain three measurements each having a standard deviation below

6 dB; Sec. II C). There was no obvious reason to exclude the data point of S1. This data point corresponded to the first condition on which the subject was tested and he might not have been sufficiently trained. In any case, it is noteworthy that DPOAE I/O curves allow estimating a compression threshold much more easily than do TMC-based I/O curves.

The reason why DPOAE-based compression threshold estimates were lower than corresponding TMC-based estimates is uncertain. Maybe the DPOAE response was influenced by mutual suppression of the primaries. This could have linearized the I/O curve (e.g., by decreasing cochlear gain) and thus *increased* the compression threshold suggested by DPOAE I/O curves. This explanation does not fit the data as the actual compression thresholds estimated from DPOAEs were *lower* than those estimated from TMCs. Perhaps the difference in compression threshold estimate for the two methods is caused by the use of suboptimal DPOAE parameters (see Sec. IV B).

The moderately high correlation between the compression-threshold estimates of the two methods also at low frequencies [Figs. 9(B), 10(E), and 10(F)] may seem surprising given the low correlation between the estimates of minimum compression exponent. One possible explanation could be that the DPOAE parameters were adequate for lower L_2, where the compression threshold occurs, but not for higher L_2 levels, at which, coincidentally, plateaus and notches occur.

The average compression threshold estimates at 4 kHz found in the present study were 47 and 40 dB SPL for TMC and DPOAE, respectively. These values are comparable to those (average 37 dB SPL at 3–4 kHz) found psychophysically by Yasin and Plack (2003) based on a three-line segment fitting procedure (Plack *et al.*, 2004). A value of ~35 dB SPL is obtained when applying the present definition of compression threshold to the DPOAE data of Neely *et al.* (2003). A compression threshold was also estimated from the DPOAEs reported by Gorga *et al.* (2007). After discarding their data for lowest stimulus levels because they were most likely contaminated by noise and applying the definition used in the present study (level at which I/O curve slope equals 0.4 dB/dB) to their data in their Fig. 3, the resulting compression thresholds were 30 and 45 dB SPL at 0.5 and 4 kHz, respectively. The present values are reasonably in accordance with their results at 4 kHz but not at 0.5 kHz. Our results indicate that the compression threshold is approximately constant across frequencies [Figs. 10(E) and 10(F)].

E. Gain and the level of return to linearity

It is still controversial that BM I/O functions become linear at very high input levels in healthy cochleae (Robles and Ruggero, 2001). Assuming, however, that this is a truly physiological characteristic, DPOAEs are not considered a reliable predictor of the threshold level of return to linearity at high levels. First, there may be another mechanism involved in generation of DPOAE at the higher stimulus levels, as explained above (Sec. IV B). Second, the best frequency of any BM site shifts with increasing level, hence the BM site where the two primaries cause maximum excitation is likely to shift accordingly with level. This may change the DPOAE response as the region of overlap of the two primaries changes with level. Because of this, the gain and return-to-linearity thresholds were not estimated based on DPOAE I/O curves and not compared with those inferred from TMCs.

The TMC-based I/O curves suggested that the cochlear gain increased with increasing CF because of a parallel increase in the threshold of return to linearity at high levels (Fig. 11). This is in agreement with physiological data (Robles and Ruggero, 2001). Gain estimates based on tip-to-tail level differences of DPOAE suppression tuning curves (Gorga *et al.*, 2008) showed the same tendency for the gain to decrease with decreasing frequency.

F. On the merits of the DPOAEs and TMCs for estimating cochlear I/O curves

The presence of plateaus and notches in the DPOAE I/O curves at 0.5 and 1 kHz results in zero or negative compression-exponent estimates which do not occur in corresponding TMC-based curves. While deep notches in apical BM I/O functions have been reported [e.g., Fig. 7(a) of Rhode and Cooper (1996)], they typically occur for stimulation frequencies higher than the CF (Rhode and Cooper, 1996). Therefore, the notches reported here are unlikely to reflect notches in the underlying BM responses (see also Sec. IV B).

It would be wrong to conclude, however, that it is inappropriate to use DPOAEs to infer cochlear I/O functions at low frequencies. As discussed earlier (Sec. IV B), the present notches are possibly due to using suboptimal primary levels at low frequencies and it might be possible to find DPOAE stimulus parameters that would lead to higher correlations between TMC-based and DPOAE I/O curves.

It would also be wrong to conclude that I/O curves inferred from TMCs are more correct (i.e., reflect more closely the underlying BM responses) than DPOAE I/O functions at low frequencies. The TMC method is an indirect, psychophysical method, thus its results may be influenced by retrocochlear mechanisms unknown to date. Indeed, there exist within-subject differences between I/O functions inferred with different psychophysical methods (e.g., Rosengard *et al.*, 2005). Furthermore, the TMC method rests on several assumptions, the main of which is that the rate of decay of the internal masker effect is identical across frequencies and levels (Nelson *et al.*, 2001; Lopez-Poveda *et al.*, 2003). The validity of these assumptions is still controversial. Stainsby and Moore (2006) have argued that the decay rate is faster for low probe frequencies (or equivalently, low CFs), at least for hearing-impaired listeners. By contrast, Lopez-Poveda and Alves-Pinto (2008) have provided indirect evidence for frequency-independent decay rates, at least for normal-hearing listeners. Additionally, there is evidence that for any given frequency, the decay rate is slower at high levels (Lopez-Poveda and Alves-Pinto, 2008; Wojtczak and Oxenham, 2007). These issues complicate the selection of the linear reference TMC and thus cast doubts on the correspond-

ing I/O curves, particularly at low frequencies (Stainsby and Moore, 2006; Lopez-Poveda and Alves-Pinto, 2008).

In summary, the lack of correlation between the results of the two methods at low frequencies is uninformative at present of their relative accuracy for inferring cochlear I/O curves. Future studies should investigate the reason for the low correlation at low frequencies and whether higher correlations would be obtained using different DPOAE parameters and/or different psychophysical methods or assumptions.

The present study did not evaluate the merit of DPOAE I/O curves as a (*clinical*) tool for assessing residual cochlear compression characteristics in hearing-impaired listeners. The present results, however, suggest that they might be useful to assess residual compression in listeners with presbyacusis, who are mostly affected by high-frequency loss.

V. CONCLUSIONS

(1) The correlation between individual compression exponent estimates inferred from TMCs and DPOAEs is reasonably high at 4 kHz, but low at 0.5 and 1 kHz. Both methods suggest that maximum compression is approximately 10:1 and constant across the characteristic frequency range from 0.5 to 4 kHz.
(2) The low correlation at low frequencies cast doubts on the postulates and interpretation of I/O curves inferred with either (or both) of the two methods. The most likely reason for the lack of correlation at low frequencies (0.5–1 kHz) is the presence of notches and plateaus in the DPOAE I/O curves. This suggests that the DPOAE stimulus paradigm of Kummer *et al.* (1998) may not be optimal (i.e., does not produce maximum DP magnitude) at low frequencies.
(3) A high correlation was found between estimates of compression threshold inferred from DPOAEs and TMC-based I/O curves between 1 and 4 kHz. The DPOAE and the TMC methods indicate that the compression threshold equals 40 and 47 dB SPL, respectively, and is approximately constant across the range of frequencies from 0.5 to 4 kHz for TMCs and from 1 to 4 kHz for DPOAEs.
(4) Cochlear gain and return-to-linearity thresholds were inferred from the TMCs only. Both parameters increased by ~16 dB with increasing characteristic frequency from 0.5 to 4 kHz.
(5) It seems reasonable to use TMCs and DPOAE I/O curves interchangeably to infer cochlear I/O curves at 4 kHz but doubts exist that the same applies to lower frequencies of 0.5 and 1 kHz.

ACKNOWLEDGMENTS

The authors are grateful for the comments and suggestions provided by two anonymous reviewers to improve an earlier version of this paper. This work was supported by IMSERSO 131/06, PROFIT CIT-390000-2005-4, MEC BFU-2006-07536, and The Oticon Foundation.

Avan, P., Bonfils, P., Gilain, L., and Mom, T. (**2003**). "Physiopathological significance of distortion-product otoacoustic emissions at 2f1-f2 produced by high-versus low-level stimuli," J. Acoust. Soc. Am. **113**, 430–441.

Bacon, S. P. (**2004**). "Overview of auditory compression," in *Compression. From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), Chap. 1, pp. 1–17.

Cooper, N. P., and Rhode, W. S. (**1997**). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea," J. Neurophysiol. **78**, 261–270.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

Gaskill, S. A., and Brown, A. M. (**1996**). "Suppression of human acoustic distortion product: Dual origin of 2f1-f2," J. Acoust. Soc. Am. **100**, 3268–3274.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2007**). "Low-frequency and high-frequency cochlear nonlinearity in humans," J. Acoust. Soc. Am. **122**, 1671–1680.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2008**). "Low-frequency and high-frequency distortion product otoacoustic emission suppression in humans," J. Acoust. Soc. Am. **123**, 2172–2190.

He, N. J., and Schmiedt, R. A. (**1993**). "Fine structure of the 2f1-f2 acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

Heinz, M. G., and Young, E. D. (**2004**). "Response growth with sound level in auditory-nerve fibers after noise induced hearing loss," J. Neurophysiol. **91**, 784–795.

Heitmann, J., Waldmann, B., Schnitzler, H., Plinkert, P. K., and Zenner, H. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near 2f1-f2 removes DP-gram fine structure—Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (**2006**). "Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 418–428.

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.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). "Suppression tuning characteristics of the 2f1-f2 distortion product otoacoustic emission in humans," J. Acoust. Soc. Am. **98**, 197–210.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the 2F1-F2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). "Optimal L1-L2 primary tone level separation remains independent of test frequency in humans," Hear. Res. **146**, 47–56.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 466–477.

Liberman, M. C., Zuo, J., and Guinan, J. J., Jr. (**2004**). "Otoacoustic emissions without somatic motility: Can stereocilia mechanics drive the mammalian cochlea?," J. Acoust. Soc. Am. **116**, 1649–1655.

Lopez-Poveda, E. A., and Alves-Pinto, A. (**2008**). "A variant temporal-masking-curve method for inferring peripheral auditory compression," J. Acoust. Soc. Am. **123**, 1544–1554.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

Lopez-Poveda, E. A., Plack, C. J., Meddis, R., and Blanco, J. L. (**2005**). "Cochlear compression in listeners with moderate sensorineural hearing loss," Hear. Res. **205**, 172–183.

Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). "One source for distortion product otoacoustic emissions generated by low- and high-level primaries," J. Acoust. Soc. Am. **111**, 2740–2748.

Lukashkin, A. N., and Russell, I. J. (**2002**). "Modifications of a single saturating non-linearity account for post-onset changes in 2f1-f2 distortion product otoacoustic emission," J. Acoust. Soc. Am. **112**, 1561–1568.

Mauermann, M., and Kollmeier, B. (**1999**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPAOE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of frequency ratio f2/f1," J. Acoust. Soc. Am. **106**, 3473–3483.

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.

Mills, D. (**1997**). "Interpretation of distortion product otoacoustic emission measurements. I. Two stimulus tones," J. Acoust. Soc. Am. **102**, 413–429.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).

Müller, J., and Janssen, T. (**2004**). "Similarity in loudness and distortion product otoacoustic emission input/output functions: Implications for an objective hearing aid adjustment," J. Acoust. Soc. Am. **115**, 3081–3091.

Neely, S. T., Gorga, M. P., and Dorn, P. A. (**2003**). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **114**, 1499–1507.

Neely, S. T., Johnson, T. A., and Gorga, M. P. (**2005**). "Distortion-product otoacoustic emission measured with continuously varying stimulus level," J. Acoust. Soc. Am. **117**, 1248–1259.

Nelson, D. A., and Schroder, A. C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2221–2233.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

Nuttall, A. L., and Dolan, D. F. (**1993**). "Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig," J. Acoust. Soc. Am. **93**, 390–400.

Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.

Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.

Rhode, W. S. (**2007**). "Distortion product otoacoustic emissions and basilar membrane vibration in the 6–9 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **122**, 2725–2737.

Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*," Aud. Neurosci. **3**, 101–121.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.

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

Siegel, J. H. (**1994**). "Ear-canal standing waves and high frequency calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.

Stainsby, T. H., and Moore, B. C. J. (**2006**). "Temporal masking curves for hearing-impaired listeners," Hear. Res. **218**, 98–111.

Stover, J. S., Neely, S. T., and Gorga, M. P. (**1996**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.

Talmadge, C., Tubis, A., Long, G. R., and Pisorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structure in humans," J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Whitehead, M. L., Stagner, B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **98**, 3200–3214.

Williams, E. J., and Bacon, S. P. (**2005**). "Compression estimates using behavioral and otoacoustic emission measures," Hear. Res. **201**, 44–54.

Wojtczak, M., and Oxenham, A. J. (**2007**). "Verification of the assumption of frequency-independent recovery from forward masking," J. Acoust. Soc. Am. **121**, 3133.

Yasin, I., and Plack, J. P. (**2003**). "The effect of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions," J. Acoust. Soc. Am. **114**, 322–332.

Yasin, I., and Plack, J. P. (**2007**). "The effect of low- and high-frequency suppressors on psychophysical estimates of basilar-membrane compression and gain," J. Acoust. Soc. Am. **121**, 2832–2841.

Hearing preservation surgery: Psychophysical estimates of cochlear damage in recipients of a short electrode array

René H. Gifford
Department of Otolaryngology, Mayo Clinic, 200 First Street Southwest, Rochester, Minnesota 55906

Michael F. Dorman, Anthony J. Spahr, and Sid P. Bacon
Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

Henryk Skarzynski and Artur Lorens
International Center of Hearing and Speech, 17 Mokra Street, Kajetany, 05-830 Nadarzyn, Poland

(Received 4 February 2008; revised 23 May 2008; accepted 3 July 2008)

In the newest implementation of cochlear implant surgery, electrode arrays of 10 or 20 mm are inserted into the cochlea with the aim of preserving hearing in the region apical to the tip of the electrode array. In the current study two measures were used to assess hearing preservation: changes in audiometric threshold and changes in psychophysical estimates of nonlinear cochlear processing. Nonlinear cochlear processing was evaluated at signal frequencies of 250 and 500 Hz using Schroeder phase maskers with various indices of masker phase curvature. A total of 15 normal-hearing listeners and 13 cochlear implant patients (7 with a 10 mm insertion and 6 with a 20 mm insertion) were tested. Following surgery the mean low-frequency threshold elevation was 12.7 dB (125–750 Hz). Nine patients had postimplant thresholds within 5–10 dB of preimplant thresholds. Only one patient, however, demonstrated a completely normal nonlinear cochlear function following surgery—although most retained some degree of residual nonlinear processing. This result indicates (i) that Schroeder phase masking functions are a more sensitive index of surgical trauma than audiometric threshold and (ii) that preservation of a normal cochlear function in the apex of the cochlea is relatively uncommon but possible.
 [DOI: 10.1121/1.2967842]

PACS number(s): 43.64.Me, 43.66.Nm, 43.64.Kc, 43.66.Dc [JCM] Pages: 2164–2173

I. INTRODUCTION

In one of the newest applications of cochlear implants, electrode arrays ranging in length from 10 to 20 mm are inserted into the scala tympani of individuals with bilateral low-frequency hearing with the aim of preserving residual hearing apical to the tip of the array. A successful surgical outcome allows for an electric stimulation of basal neural tissue without damaging the apical cochlear structures that transmit low-frequency acoustic information (e.g., von Ilbert *et al.*, 1999; Skarzynski *et al.*, 2004, 2006, 2007; Gantz and Turner, 2003, 2004; Gstoettner *et al.*, 2004; Gantz *et al.*, 2005, 2006; Kiefer *et al.*, 2005; Leutje *et al.*, 2007). The mean hearing loss following this procedure ranges from 10 to 20 dB depending on the electrode array and the nature of the surgical technique (Gantz and Turner, 2004; Gantz *et al.*, 2005; Skarzynski *et al.*, 2003, 2006, 2007; Gstoettner *et al.*, 2004, 2005; Kiefer *et al.*, 2005). The combination of binaural low-frequency acoustic hearing and monaural high-frequency electric hearing—termed combined electric and acoustic stimulation (EAS)—has been shown to improve speech understanding in quiet and in noise beyond that achieved by aided acoustic or electric hearing alone (Wilson *et al.*, 2002; Brill *et al.*, 2002; Gantz *et al.*, 2005; Kiefer *et al.*, 2005; Gstoettner *et al.*, 2004). Gantz *et al.* (2005) reported a mean consonant nucleus consonant (CNC) (Peterson & Lehiste, 1962) score in the combined EAS condition (implant plus binaural hearing aids) of 67% correct for 11 hybrid recipients (10 mm electrode), and Gstoettner *et al.* (2006) reported a mean CNC score (also for the combined EAS condition) of 75% correct for 20 mm, EAS patients. In both studies, the postoperative combined EAS scores represented a significant improvement in speech perception performance relative to the best aided scores obtained preoperatively. These EAS scores represent an above average performance relative to the mean 55%–60% monosyllabic word recognition typically reported for conventional implant recipients of newer generation technology (e.g., Baumgartner *et al.*, 2007; Balkany *et al.*, 2007; Gifford *et al.*, 2008).

The research reviewed above makes it clear that hearing can be preserved to within 10–20 dB of preimplant levels following the insertion of a short electrode array. What is not clear is whether the presence of an electrode array in the scala tympani affects cochlear mechanics and function beyond that which is revealed by the audiogram. To date, the only pre- and postimplant measures of auditory function with EAS patients have been clinical assessments of the audiometric threshold. To provide a broader view, the current study assesses the effect of electrode insertion on a fundamental aspect of normal cochlear function—nonlinear cochlear processing. Nonlinear cochlear processing was assessed because in the normal cochlea it is responsible for

high sensitivity, broad dynamic range, sharp frequency tuning, and enhanced spectral contrasts via suppression (e.g., Oxenham and Bacon, 2004).

Psychophysical estimates of nonlinear cochlear processing can be obtained by measuring masked thresholds for Schroeder phase harmonic complexes. Positive-Schroeder ($m+$) and negative-Schroeder ($m-$) phase complexes have identical amplitude spectra but different phase spectra (Schroeder, 1971). Although $m+$ and $m-$ complexes have identical flat envelopes, $m+$ complexes tend to produce less masking. This difference in masking is often referred to as the phase effect. Several researchers have hypothesized that the phase effect results from the $m+$ complexes producing a more peaked response along the basilar membrane (BM) coupled with fast-acting compression (e.g., Kohlrausch and Sander, 1995; Carlyon and Datta, 1997; Summers and Leek, 1998; Summers, 2000). Consistent with the possibility that the phase effect is influenced by the cochlear nonlinearity are results showing that it is reduced in subjects with cochlear hearing loss (Summers and Leek, 1998; Summers, 2000; Oxenham and Dau, 2004). Given that hearing impairment likely involves some degree of outer hair cell destruction and/or dysfunction, a reduction in psychophysical estimates of compression is a plausible outcome.

Physiological evidence supporting the relationship between the Schroeder phase effect and cochlear compression has been provided by Recio and Rhode (2000), who measured BM responses to $m+$ and $m-$ complexes in chinchilla cochleas. Their findings corresponded well with psychophysical data in that the BM responses to the $m+$ complexes were much peakier than those to the $m-$ complexes. Additionally, the BM response difference between $m+$ and $m-$ complexes was reduced or absent in peripherally damaged cochleaes for which BM compression was substantially reduced or absent.

Summers (2001) examined the effects of phase for Schroeder phase harmonic complexes by measuring the overshoot for listeners with normal hearing and listeners with sensorineural hearing loss. Overshoot refers to the increase in threshold as a brief signal is moved from the temporal center to the onset of a broadband masker. Overshoot is known to be reduced in individuals with both permanent sensorineural hearing loss (Bacon and Takahashi, 1992) and temporary aspirin-induced hearing loss affecting the outer hair cell function (McFadden and Chamlin, 1990). Summers (2001) found that the overshoot with the Schroeder phase maskers was greatest for listeners with normal hearing and for positive phase ($m+$) Schroeder maskers. Summers (2001) concluded that the findings were consistent with previous reports that the masking effectiveness of positive phase Schroeder harmonic complexes is more influenced by nonlinear cochlear processing.

Further evidence linking the Schroeder phase effect with nonlinear cochlear processing was provided by Oxenham and Dau (2001). They used a temporal window model that incorporated a static nonlinearity, similar to that seen with BM compression. With a normal amount of compression, the output of the model was similar to behavioral results from normal-hearing subjects (Oxenham and Dau, 2001). Without compression, the phase effect was completely eliminated.

In the current study, estimates of audiometric threshold and nonlinear cochlear function were obtained before and after surgery for patients implanted with 10 and 20 mm electrode arrays. The following questions were asked: (i) What is the average change in audiometric threshold following surgery? (ii) Is the psychophysical measure of nonlinear cochlear function more sensitive to cochlear damage than audiometric threshold? (iii) Is it possible to preserve cochlear function within the range of normal following electrode insertion? If so, how common is this outcome?

II. THE EFFECTS OF SHORT ELECTRODE IMPLANTATION ON MASKER PHASE CURVATURE

A. Subjects

Two subject groups were evaluated for the present study. The first group included 15 listeners with normal hearing to be used as the control group representing a normal cochlear function. All 15 listeners with normal hearing were female with a mean age of 22.6 years (range of 19–31). The second group of subjects included 13 individuals with various degrees of hearing loss who had been identified as candidates for EAS. Of the 13 individuals, 6 were recipients of a 20 mm electrode array (MED EL, EAS) and 7 were recipients of a 10 mm electrode array (Cochlear, Nucleus Hybrid). Of the 13 EAS recipients, 10 individuals were available for both pre- and postimplant assessments of nonlinear cochlear processing. The remaining three subjects were only available for a postimplant evaluation. Postimplant evaluation of an audiometric threshold and Schroeder masking functions was obtained 2–13 months postoperatively (with a mean duration of 8 months). The mean age of the 13 EAS subjects was 46.0 years (range of 28–75 years) with 11 female and 2 male participants.

For the six subjects receiving the 20 mm MED EL EAS electrode array, preoperative audiometric thresholds gave ranges of from 10–55 dB Hearing Level (HL) at 500 Hz, 20–50 dB HL at 750 Hz, 55 to >120 dB HL at 1000–1500 Hz, and 80 to >120 dB HL at 2000 Hz and above. All six subjects were implanted using a round window surgical technique described by Skarzynski *et al.* (2007).

For the seven subjects receiving the 10 mm Nucleus Hybrid electrode array, preoperative audiometric thresholds—obtained using standard audiometric techniques—gave ranges of 0–30 dB HL at 500 Hz, 20–50 dB HL at 750 Hz, 25–115 dB HL at 1000–1500 Hz, and 95 dB HL and higher at 2000 Hz and above. All seven subjects were implanted via standard implantation techniques with the use of a cochleostomy described by Gantz *et al.* (2005). Figure 1 displays individual and mean audiometric thresholds for all 13 subjects for the preimplant (filled circles) and postimplant (shaded triangles) conditions. The error bars represent ±2 standard error. For those frequencies where no behavioral responses were obtained, a symbol was placed at 125 dB

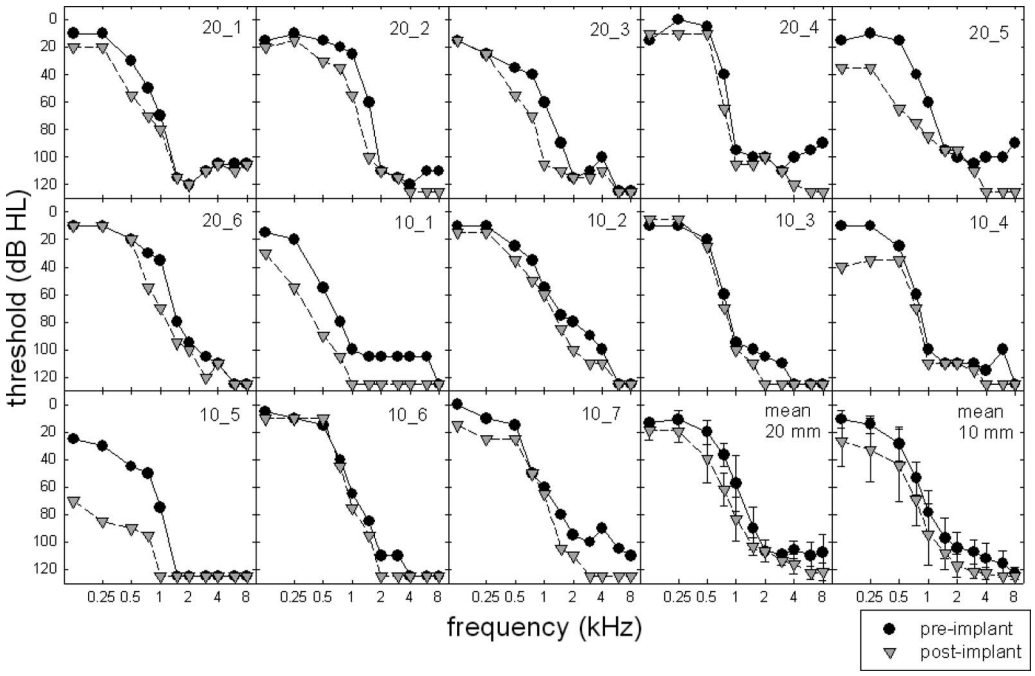


FIG. 1. Individual and mean audiometric thresholds (in dB HL) obtained preoperatively (filled circles) and postoperatively (shaded inverted triangles).

HL—which is beyond the limits of the audiometer for all frequencies tested. Further details will be provided in Sec. II C.

B. Methods

Psychophysical estimates of nonlinear cochlear processing were measured by obtaining thresholds for pure-tone signals, 250 and 500 Hz, in the presence of positively and negatively scaled Schroeder phase harmonic complexes. The phases of the masker components were selected according to an equation originally proposed by Schroeder (1971) and more recently modified by Lentz and Leek (2001),

$$\theta_n = C\pi n(n-1)/N, \quad -1 \leqslant C \leqslant 1,$$

where C is a scalar factor and N is the number of components in the harmonic complex. A true positive $(m+)$ or negative $(m-)$ Schroeder phase complex is achieved when the scalar factor C is equal to 1 or −1, respectively. A sine-phase complex is produced when C is equal to 0. By varying the scalar value C from −1 to 1, a range of frequency sweep rates, or masker phase curvatures, can be produced (Lentz and Leek, 2001). When the phase curvature of the masker is equal, but opposite in sign, to that of the auditory filter centered at f_s, masker effectiveness will be at a minimum. Thus, for the first ten subjects in the present study, masked thresholds were obtained for scalar factors, C, ranging from −1 to 1 in increments of 0.25. For the last three subjects enrolled, masked thresholds for scalar factors of −1, +1, −0.25, and +0.25 were obtained (more details are provided in Sec. II C). Based on the findings of Oxenham and Dau (2001), the spectral range of the masker encompassed the frequency range between $0.4f_s$ and $1.6f_s$. The fundamental frequency of the masker complex was 25 Hz. The overall level of the masker was fixed at 75 dB sound pressure level (SPL), and the signal level was varied adaptively. The durations of the masker and the signal were 400 and 200 ms, respectively (including 10 ms $\cos^2$ rise/fall times). The signal was placed in the temporal center of the masker.

Relatively low signal frequencies were chosen so that quiet thresholds across the range of the masker spectrum were likely either normal or near normal. Also, Oxenham and Dau (2001) examined masker phase effectiveness at 250 Hz and found a large phase effect (>20 dB) in normal-hearing listeners. Furthermore, a 250 Hz signal was chosen so that it was possible to obtain EAS subjects with normal or near-normal hearing for a (limited) range of frequencies above f_s. The reason that normal or near-normal hearing above the test frequencies is important is that the cochlear location of the nonlinearity generally ranges from 1/3 to 1/2 of an octave above that of the cochlear place corresponding to the characteristic frequency of the test stimulus (e.g., Davis, 1983; Chatterjee and Zwislocki, 1997; Rhode and Recio, 2000). Thus, it seems reasonable that to obtain a true estimate of the nonlinear cochlear function in the low-frequency cochlear region, one must ensure that cochlear functioning above f_s is also relatively normal. Gifford *et al.* (2007) examined the effects of masker phase curvature in subjects meeting audiologic candidacy for EAS and reported that nonlinear cochlear processing was present in the majority of subjects tested. Given that subjects were required to have hearing thresholds less than 55 dB HL at 500 Hz (see Sec. II A), the nonlinear properties of the cochlea would be expected to be present (though perhaps slightly reduced) at the test frequency (Neely and Kim, 1986). However, it is the comparison between the pre- and postimplant phase effects that was of interest, not simply the magnitude of the phase effect.

Thresholds for Schroeder phase masking were measured in an adaptive three-interval forced-choice paradigm with a 3-down, 1-up stepping rule to track 79.4% correct (Levitt,

1971). A run consisted of eight reversals. The first two reversals were discarded, and the threshold was determined using the average signal level at the remaining six reversal points. The initial step size of 5 dB was decreased to 2 dB after the second reversal. On the rare occasion that an estimate had a standard deviation greater than 5 dB, that run was discarded. All reported thresholds represent the mean of at least two estimates. If the difference between the two thresholds was greater than 3 dB, one additional run was completed and averaged. A third run was required for just nine of the individual thresholds obtained in the current study. The maximum level for the dynamically varying stimulus was fixed at 92 dB SPL. During a run, it was permissible for the threshold track to reach the ceiling value; however, if the tracking procedure called for a higher level, that run was discarded. If two runs for a particular condition were discarded on this basis, it was concluded that a threshold for that condition could not be achieved. Subjects were provided with a minimum of 1 h training on simultaneous masking with the Schroeder phase maskers. This amount of training was found to be sufficient for the majority of subjects in order to achieve stable threshold estimates. One subject, however, required nearly 2 h in order to become fully acquainted with the task, as evidenced by threshold stabilization.

All stimuli were generated and produced digitally at a 20 kHz sampling rate. Stimuli were routed monaurally to one channel of Sennheiser HD250 Linear II stereo headphones via an Echo Indigo input/output laptop soundcard. Subjects were tested in a double-walled sound-attenuating booth. Observation intervals were signaled by the highlighting of visual stimuli on a laptop computer monitor. The signal was presented randomly in one of three intervals. Subjects responded by either pressing a button on the number keypad or using a mouse. Visual feedback was provided.

C. Results

1. Audiometry

Figure 1 displays individual pre- and postoperative audiometric thresholds for frequencies from 125 to 8000 Hz in the implanted ear. An absent behavioral response to sound was denoted by a threshold of 125 dB HL, which was beyond the limits of the audiometer for all frequencies tested. A two-way, repeated-measures analysis of variance (ANOVA) was completed with frequency and time of measurement (pre-versus postimplant) as the variables. There was a significant effect of time ($F_{(1,12)}=57.73$, $p<0.001$). That is, at the group level, postimplant thresholds were found to be significantly poorer than preimplant thresholds. Mean thresholds that collapsed across frequency were 73.6 and 86.3 dB HL for the pre- and postimplant conditions, respectively. There was also a significant effect of frequency, which was expected given the sloping nature of the hearing losses ($F_{(12,10)}=193.547$, $p<0.001$). A statistical analysis also revealed that there was no significant interaction between the frequency and the time of the threshold assessment ($F_{(1,10)}=1.80$, $p=0.067$). In other words, the degree of threshold change between the pre- and postimplant assessments was not found to be significantly different across the range of signal frequencies tested.

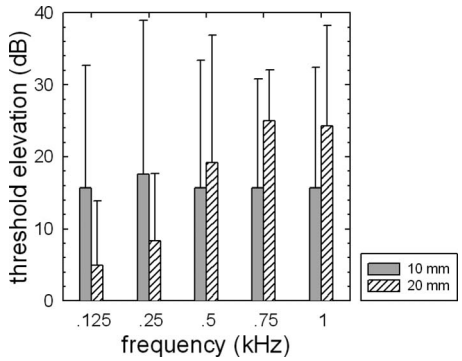


FIG. 2. Mean degree of audiometric threshold elevation for the subjects receiving 10 mm (shaded bars) and 20 mm (hatched bars) electrode arrays. The error bars represent a ±2 standard error.

The statistical analysis revealed significantly elevated thresholds following the surgical insertion of an electrode array; however, if we were to use a ±10 dB test-retest variability as the clinical estimate of a nonsignificant difference in behavioral thresholds for clinical audiometric procedures (Stuart *et al.*, 1991), there were five subjects for whom no difference in threshold was observed at 500 Hz, postoperatively. Of these five subjects, two were recipients of the 20 mm array (20_4 and 20_6) and three were recipients of the 10 mm array (10_2, 10_3, and 10_6).

Mean postoperative threshold elevation (in decibels) as a function of frequency is plotted in Fig. 2. The shaded and hatched bars represent the 10 and 20 mm recipients, respectively. The error bars represent a +2 standard error. An ANOVA was completed, analyzing the effect of the electrode array length on the degree of threshold elevation. Only results from thresholds at 125–750 Hz were entered into the analysis because hearing losses at higher frequencies tended to be near the ceiling. The results of the statistical analysis revealed no main effect of the electrode array length on the degree of threshold elevation ($F_{(1,11)}=0.071$, $p=0.795$). The mean threshold elevation for the 10 mm group was 16.4 dB. The mean threshold elevation for the 20 mm group was 14.4 dB. There was an effect of frequency ($F_{(3,11)}=4.276$, $p=0.012$) and a significant interaction between the array length and frequency ($F_{(1,3)}=5.549$, $p=0.003$). An inspection of Fig. 2 suggests that the threshold elevation for the 10 mm group tended to be constant across frequencies, whereas the threshold elevation for the 20 mm group increased with frequency. The degree of threshold elevation, however, was not significantly different across the 10 and 20 mm groups.

2. Schroeder phase masking

Figure 3 displays mean masked thresholds for the Schroeder phase maskers as a function of masker phase curvature, or scalar factor (C), for the normal-hearing listeners. The filled and unfilled symbols represent thresholds for the 250 and 500 Hz signals, respectively. The error bars represent a ±2 standard error. The Schroeder masking functions for the normal-hearing listeners exhibit the typical curved

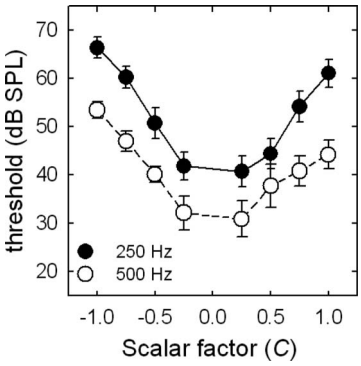


FIG. 3. Mean masked thresholds (in dB SPL) for the normal-hearing listeners as a function of masker phase curvature, or scalar factor (C). The filled and unfilled circles represent masked thresholds for the 250 and 500 Hz signals, respectively. The error bars represent a +2 standard error.

pattern that has been reported elsewhere (Oxenham and Dau, 2001, 2004). That is, the masked thresholds vary as a function of both masker phase curvature and signal frequency. The mean Schroeder phase effect, which was calculated as the threshold difference between the peak ($C=-1$) and valley ($C=0.25$) of the function, was 26.8 dB for 250 Hz and 24.1 dB for 500 Hz. Oxenham and Dau (2001, 2004) presented similar data with Schroeder phase effects of 22 dB at 250 Hz and 21 dB at 500 Hz. The phase curvature associated with the peak ($C=-1$) and valley ($C=0.25$) of the function was also similar to that reported by Oxenham and Dau (2001); however, Oxenham and Dau (2004) used a 50 Hz fundamental for a masker complex centered at 500 Hz in contrast to the 25 Hz fundamental used in the current study.

3. Signal frequency: 250 Hz

Figures 4 and 5 display the masking functions for the EAS subjects obtained both pre- and postoperatively for signal frequencies of 250 and 500 Hz. Preoperative masked thresholds are represented by the filled circles, and postoperative thresholds are represented by the shaded inverted triangles. The shaded area at the bottom of each figure represents mean masked thresholds for the normal-hearing listeners and a ±2 standard error about the mean. After data were collected from the first nine subjects, it was realized that the most important masker scalar factors for determining the shape of the masking function were +1, −1, +0.25, and −0.25. Thus, in an effort to save time during the time-intensive sessions, the last three subjects were run using only these four scalar factors for each signal frequency. If a masked threshold could not be obtained for a given condition (i.e., the signal level would have to be increased beyond the 92 dB SPL maximum), the "threshold" was plotted at 92 dB SPL and an asterisk was placed next to the symbol. This occurred for just one subject at 250 Hz and for two subjects at 500 Hz.

Examining the preimplant data for the 250 Hz signal (filled circles), there were nine subjects for whom the minima of the masking function were located in the shaded region—or near the mean thresholds for normal-hearing listeners. This included all three of the 20 mm subjects for whom we obtained preimplant data and six out of the seven 10 mm subjects (excluding 10_5). This outcome should not be surprising given that the poorest audiometric threshold at 250 Hz obtained preoperatively for these subjects was 25 dB HL. That is, research has shown that the effects of cochlear damage on psychophysical tasks thought to reflect nonlinear cochlear processing tend to become evident once thresholds reach 30 dB HL (e.g., Laroche *et al.*, 1992). Thus this group of nine subjects exhibited minima thresholds near the mean normal-hearing listeners' minima thresholds. The one 10 mm subject who did not exhibit normal cochlear nonlinearity

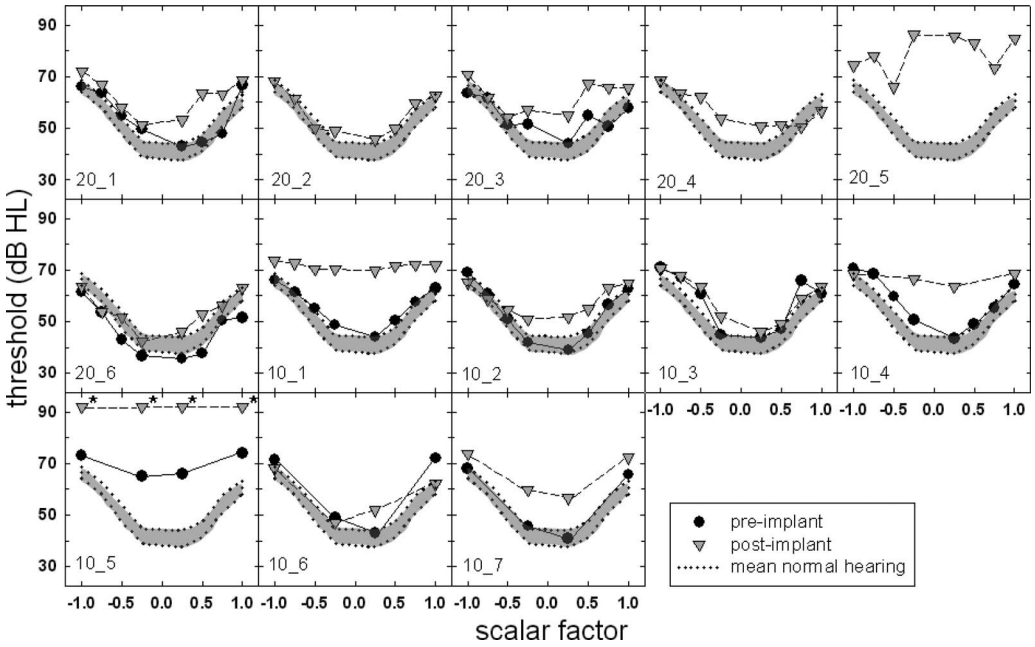


FIG. 4. Individual masked thresholds (in dB SPL) at 250 Hz for the implanted subjects. The filled symbols represent the preoperative condition, and the shaded inverted triangles represent the postoperative condition. The shaded area outlined by a dotted line encompasses the area of the mean masked thresholds and a +2 standard error for the normal-hearing listeners.

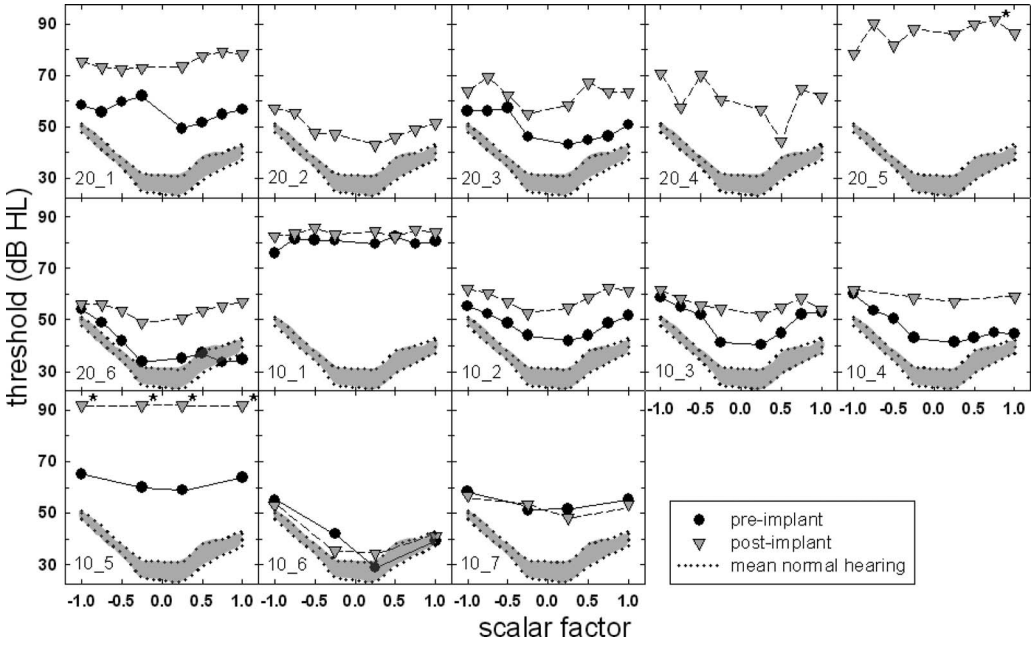


FIG. 5. Individual masked thresholds (in dB SPL) at 500 Hz for the implanted subjects. The filled symbols represent the preoperative condition, and the shaded inverted triangles represent the postoperative condition. The shaded area outlined by a dotted line encompasses the area of the mean masked thresholds and a +2 standard error for the normal-hearing listeners.

prior to surgery (10_5) had preoperative thresholds in the range of 25–50 dB HL through 750 Hz. Thus, it is not surprising that she demonstrated little nonlinear cochlear processing even before surgery.

Postoperative results (inverted triangles) demonstrate a different pattern of results. For the 250 Hz signal shown in Fig. 4, none of the 10 nor 20 mm subjects' minima thresholds were within the shaded region. This suggests that the surgical insertion of the electrode array resulted in a significant elevation of all subjects' masked thresholds for the 0.25 scalar factor. The minima of masking function for subjects 20_2, 20_6, and 10_3, however, were just marginally outside of the shaded region.

Recall that five subjects demonstrated no significant postoperative elevation in the audiometric threshold at 250 Hz (20_4, 20_6, 10_2, 10_3, and 10_6). None of these five subjects, however, exhibited masked thresholds in the shaded region of normal hearing for *both* the pre- and postimplant conditions. Further examination of these data revealed that subject 10_2 exhibited an abnormal *postoperative* masking function despite the fact that her postimplant audiometric thresholds increased by just 5 dB at 250 Hz and by 10 dB at 500 Hz. Subject 10_6 also demonstrated abnormal masking functions postoperatively even though her pre- and postimplant audiometric thresholds were identical at 250 Hz and were within 5 dB of one another at 500 Hz. On the other hand, subjects 20_5, 10_1, 10_4, and 10_5 demonstrated considerably reduced or absent nonlinear cochlear processing evidenced by a flat or relatively flat postoperative masking function. Given that these subjects exhibited considerable postimplant elevations in audiometric thresholds, this outcome was not surprising. Subject 10_7, however, demonstrated much reduced nonlinear function following surgery even though her pre- and postimplant audiometric thresholds varied by 15 dB at 250 Hz and by 10 dB at 500 Hz. These results suggest that while pre- and postimplant audiometric thresholds may differ by a value that is just outside the acceptable range for audiometric test-retest variability, the Schroeder masking pattern may be significantly altered postoperatively. Thus, Schroeder phase masking appears to be a more sensitive index of surgically related damage to the cochlea.

4. Signal frequency: 500 Hz

Figure 5 displays masking data for the 500 Hz signal. Examining the preimplant data at 500 Hz (circles), there was just one subject (10_6) for whom the minima of the masking function was located within the shaded region representing the normal-hearing data. There were other subjects, however, whose preoperative masking functions may have been elevated but still exhibited clear maxima and minima (20_3, 20_6, 10_2, 10_3, and 10_4).

Examining postoperative results at 500 Hz, there were no subjects for whom pre and postimplant masking functions were both within the shaded region depicting the mean data (+2 standard error) for the normal-hearing listeners. Examining those subjects whose preoperative masking function was either normal or elevated but still exhibited curvature with a definitive maximum and minimum (20_3, 20_6, 10_2, 10_3, 10_4, and 10_6), five of these individuals (excluding 10_6) had nearly lost all evidence of nonlinear cochlear processing at 500 Hz following surgery—based on the masked thresholds displayed in Fig. 5. This was a surprising outcome—particularly for subjects 10_2 and 10_3—as their postoperative audiometric thresholds increased by just 5–15 dB at 500 and 750 Hz. Thus one might have hypothesized that there would have been a postimplant reduction in the Schroeder phase effect but not a near elimination of the effect.

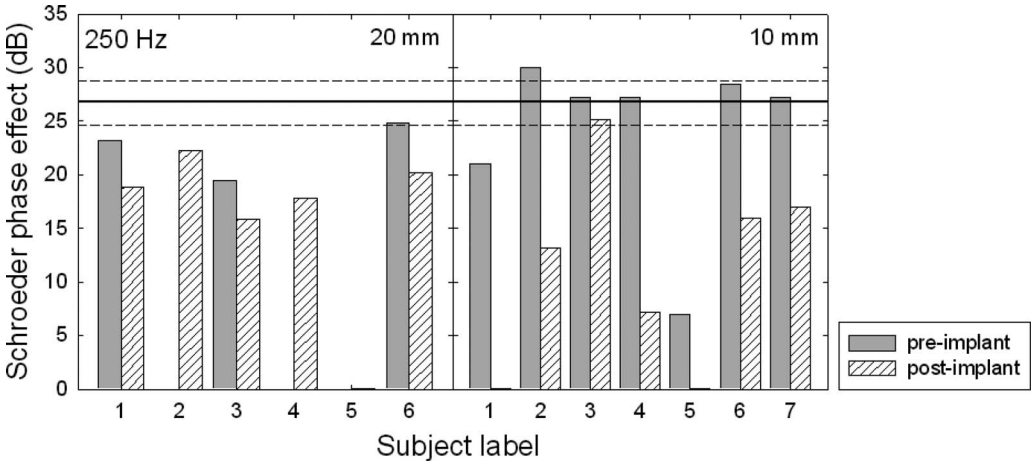


FIG. 6. Schroeder phase effect (in dB) at 250 Hz for the preoperative (shaded bars) and postoperative (hatched bars) conditions. The 20 and 10 mm subjects' data are shown in the left and right panels, respectively. The solid line at 26.8 dB represents the mean phase effect for the 15 normal-hearing subjects, and the dashed lines represent a ±2 standard error.

The threshold functions shown in Figs. 4 and 5 do not necessarily convey whether the phase effects may have been within the range expected for normal-hearing listeners. Furthermore, the masked thresholds also do not necessarily demonstrate whether a postoperative reduction in the phase effect—or a psychophysical estimate of nonlinearity—was observed. That is, an individual may demonstrate higher than normal masked thresholds but a normal or near-normal phase effect. Thus the effects of masker phase curvature, or the Schroeder phase effect, were calculated for each EAS subject by determining the difference in masked threshold between the maximum and the minimum of each function for each signal frequency. These data are shown in Figs. 6 and 7.

For those instances where there may have been a negative-Schroeder phase effect (the masked threshold for a scalar factor of 0.25 was higher than that for −1.0), a value of zero was entered, indicating the lack of nonlinear cochlear function. Figures 6 and 7 display the Schroeder phase effect (in decibels) at 250 and 500 Hz for each of the individual subjects both preoperatively (shaded bars) and postoperatively (hatched bars). The six 20 mm subjects' data are shown in the left panel, and the seven 10 mm subjects' data are shown in the right panel. The solid horizontal line represents the mean phase effect for the 15 normal-hearing listeners, and the dashed lines above and below represent a ±2 standard error.

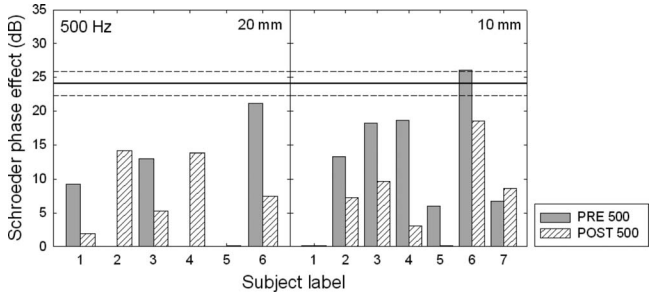


FIG. 7. Schroeder phase effect (in dB) at 500 Hz for the preoperative (shaded bars) and postoperative (hatched bars) conditions. The 20 and 10 mm subjects' data are shown in the left and right panels, respectively. The solid line at 24.1 dB represents the mean phase effect for the 15 normal-hearing subjects and the dashed lines represent a ±2 standard error.

Examining the 250 Hz data in Fig. 6, one of the 20 mm subjects (20_6) and five of the 10 mm subjects (20_2, 20_3, 20_4, 20_6, and 20_7) had preoperative phase effects in the range expected for normal-hearing listeners. Postoperatively, however, only one subject, 10_3, retained a phase effect that was within a ±2 standard error of the mean for normal-hearing listeners. This suggests that all but a single subject demonstrated an abnormal degree of nonlinear cochlear processing at 250 Hz following surgery. There were several subjects, however, who did exhibit a considerable phase effect following surgery—albeit not within the expected range for a normal cochlear function. Thus, EAS surgery resulted in the preservation of some residual cochlear nonlinearity, but not necessarily within the range of *normal*.

Examining the 500 Hz data in Fig. 7, just one subject (10_6) exhibited a normal phase effect preoperatively. Postoperatively, none of the 13 subjects retained a *normal* degree of nonlinear cochlear processing. There were several subjects, however, who exhibited a postoperative phase effect, which suggests that these implant recipients retained some degree of nonlinear cochlear processing. Nevertheless, the effects of surgical trauma appeared to be more apparent at 500 Hz.

The statistical analysis using a repeated-measures ANOVA was computed comparing phase effects for the *preoperative* 10 and 20 mm recipients to the normal-hearing listeners at 250 and 500 Hz. The 10 and 20 mm data were combined into one group termed EAS. A significant effect of subject group was found such that the Schroeder phase effect was found to be larger for the normal-hearing listeners relative to the preimplant subjects ($F_{(1,23)}=14.36$, $p<0.001$). This trend is more apparent in Fig. 7 for the 500 Hz signal. The mean Schroeder phase effects that collapsed across frequencies were 18.3 dB for the preoperative EAS subjects and 25.5 dB for the normal-hearing subjects. A significant effect of signal frequency was also found ($F_{(1,1)}=33.26$, p

<0.001); the mean Schroeder phase effects were 25.2 dB for 250 Hz and 18.6 dB for 500 Hz. There was also a significant interaction between the frequency and the subject group ($F_{(1,1)}=11.42$, $p=0.003$). An all pairwise multiple comparison using the Tukey test revealed that there was a significant difference between the preimplant EAS and the normal-hearing data at 500 Hz ($q=7.06$) but not at 250 Hz ($q=2.10$). This is another finding that is easily observed in Figs. 6 and 7. Clearly more subjects achieved normal phase effects preoperatively at 250 Hz than at 500 Hz. The *post hoc* analysis also revealed that the Schroeder phase effect was not significantly different across signal frequencies for the normal-hearing listeners ($q=2.67$) but was significantly different across frequencies for the EAS subjects ($q=8.35$). This was not unexpected given the sloping nature of the audiometric thresholds in the preimplant subjects, resulting in a reduced effect of masker phase curvature at 500 Hz.

Another statistical analysis was completed examining the effect of time of testing (pre—versus postimplant) for the EAS Schroeder phase effect. There was a significant main effect of time (pre—versus postimplant) on the Schroeder phase effect ($F_{(1,9)}=43.6$, $p<0.001$). That is, the preimplant Schroeder phase effect was found to be significantly greater than the postimplant effect. The phase effects that collapsed across frequencies were 18.34 and 9.9 dB for the pre- and postimplant conditions, respectively. The mean preoperative phase effects were 23.6 and 13.1 dB at 250 and 500 Hz. The mean postoperative phase effects were 13.5 and 6.4 dB at 250 and 500 Hz. This confirms the fact that the surgical insertion of the electrode array significantly reduced nonlinear cochlear processing in the lower-frequency region.

III. DISCUSSION

It is reasonable to suppose that the insertion of a relatively stiff foreign body (an electrode array) into the scala tympani could alter cochlear mechanics or cochlear function. In the Introduction it was noted that the audiogram provides only a very narrow view of changes in cochlear function. To broaden that view the current study assessed changes in both audiometric thresholds and low-frequency nonlinear cochlear processing following the insertion of electrode arrays 10 and 20 mm into the cochlea.

All of the 13 patients tested had some degree of hearing preservation. The electrode insertion depth was related in a frequency-specific manner to the preservation of audiometric thresholds. Hearing loss with a 10 mm insertion was relatively constant across a frequency range of 125–750 Hz. For a 20 mm insertion hearing loss increased over a range of 125–750 Hz. Thus although no significant differences were noted in the degree of hearing preservation across the 10 and 20 mm groups, it is reasonable to expect that a deeper electrode insertion will be associated with greater hearing loss at higher frequencies within the 125–750 Hz range.

Nine subjects had postoperative audiometric thresholds that were within 10 dB of the preoperative estimates at the signal frequencies tested, which is considered within the limits of acceptable test-retest variability (Stuart *et al.*, 1991). Out of these nine subjects, only one demonstrated a normal nonlinear cochlear function, as evidenced by the Schroeder phase effect following surgery at 250 Hz (Fig. 6), and none of the subjects demonstrated a normal cochlear function postoperatively at 500 Hz (Fig. 7). However, there were subjects for whom the postoperative Schroeder phase effect was clearly abnormal but who still exhibited some phase effect and thus retained some degree of nonlinear cochlear processing. Based on these findings, we conclude that the measure used in the current study is a more sensitive index of post-surgical alterations to cochlear function than the standard audiometric threshold.

In the light of this more sensitive index, is it possible to preserve a normal cochlear function at the apical end of the cochlea following the insertion of electrodes arrays 10 and 20 mm into the cochlea? The answer is "yes" for the 10 mm insertion—for a single subject (10_3) at 250 Hz. This is likely due to the higher degree of threshold elevation just above 500 Hz than above 250 Hz. Although only one subject demonstrated a normal nonlinear cochlear function postoperatively, most subjects still exhibited a postimplant phase effect (more so at 250 Hz than at 500 Hz). Thus, although retention of normal nonlinear cochlear processing is not the most common outcome of hearing preservation surgery, preservation of some cochlear nonlinearity (with postoperative phase effects >5 dB) was achieved for ten subjects at 250 Hz and for seven subjects at 500 Hz.

If the preservation of low-frequency cochlear function is possible following surgery, does it make a difference for speech recognition? Speech recognition and psychophysical data were available for only 7 of the 13 patients—the native English speaking patients (see Table I). The European clinical trial of EAS did not require a evaluation of postoperative acoustic only nor of ipsilateral EAS speech perception for the implanted ear. Thus, these data were not available for the six Polish speaking 20 mm EAS subjects. Table I displays CNC monosyllabic word recognition data for the seven 10 mm English speaking subjects for the pre- and postimplant acoustic only (A) conditions (for the implanted ear), postimplant electric only (E), and postimplant ipsilateral EAS (A+E). For the most part, the 10 mm subjects demonstrated considerable preservation of acoustic only speech perception in the implanted ear following surgery. Only one subject (10_5) demonstrated a significant decrement in acoustic only word recognition postoperatively using a binomial distribution statistic for 50-item word lists (Thornton and Raffin, 1978). Subject 10_5, however, was the only subject with a near complete loss of hearing.

For the seven English speaking 10 mm subjects, there was a significant correlation between the magnitude of the threshold elevation and the postoperative acoustic-only speech perception. The greater the elevation, the poorer the understanding ($r=-0.82$, $p=0.02$). This is not surprising. There was no evidence that the Schroeder data provided predictive information about speech understanding that was not supplied by audiometric thresholds. However, a much larger data set is needed to assess this rigorously. The ease of the

TABLE I. Individual and mean CNC word recognition scores, in percent correct, for the acoustic only (A) both pre- and postimplant, electric only (E), and ipsilateral electric (E) plus acoustic (A) conditions.

Subjects	A only implanted ear preimplant	A only implanted ear postimplant	E only	Ipsilateral E+A
10_1	16	10	70	66
10_2	38	44	76	78
10_3	20	20	32	38
10_4	16	16	6	30
10_5	8	0	48	58
10_6	44	44	54	76
10_7	40	38	0	52
Mean (Standard deviation)	26.0 (14.3)	24.6 (17.5)	40.9 (29.6)	56.9 (18.3)

Schroeder masking task and the need for just two to four thresholds per frequency provides motivation to use this measure on a much larger sample of hearing preservation patients.

IV. SUMMARY

The average changes in threshold up to 750 Hz following surgery were 16.4 and 14.4 dB for the 10 and 20 mm insertions, respectively. The psychophysical estimate of nonlinear cochlear function, i.e., Schroeder phase effect, was more sensitive to the effects of surgery than audiometric thresholds. The most common outcome of surgery was elevation of audiometric thresholds and decreases or elimination of nonlinear cochlear function. It is possible, but not common, to preserve a normal auditory function at 250 Hz following electrode insertion. Preservation of *some* residual nonlinear cochlear functions was common for both the 10 and 20 mm subjects, more so at 250 Hz. If claims of "better" surgical techniques for hearing preservation are made in the future, then Schroeder phase masking patterns could be used to evaluate those claims.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC006538 to R.H.G. and by NIDCD Grant No. RO1 DC00654-15 to M.F.D. A portion of the results were presented at the 2006 International Conference on Cochlear Implants and Other Implantable Auditory Technologies, in Vienna, Austria, the 2005 Hearing Preservation Workshop in Warsaw, Poland, and the 2006 Hearing Preservation Workshop in Raleigh, NC. The authors would like to thank Dr. Chris Brown at Arizona State University for his assistance with programing as well as Sharon McKarns, Andrea Kosko, and Arkadiusz Wasowski for their help with data collection. This study was conducted in strict accordance with approved Arizona State University and International Center of Hearing and Speech IRB protocols.

Bacon, S. P., and Takahashi, G. (**1992**). "Overshoot in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 2865–2871.

Balkany, T., Hodges, A., Menapace, C., Hazard, L., Driscoll, C., Gantz, B., Kelsall, D., Luxford, W., McMenomy, S., Neely, G., Peters, B., Pillsbury, H., Roberson, J., Schramm, D., Telian, S., Waltzman, S., Westerberg, B., and Payne, S. (**2007**). "Nucleus freedom north american clinical trial," Otolaryngol.-Head Neck Surg. **136**, 757–762.

Baumgartner, W. D., Jappel, A., Morera, C., Gstoettner, W., Muller, J., Kiefer, J., Van De Heyning, P., Anderson, I., and Nielson, S. B. (**2007**). "Outcomes in adults implanted with the FLEXsoft electrode," Acta Oto-Laryngol. **127**, 579–586.

Brill, S., Lawson, D. T., Wolford, R. D., and Schatzer, R. (**2002**). "Speech processors for auditory prostheses." 11th quarterly progress report on NIH Project N01—DC-8–2105.

Carlyon, R. P., and Datta, J. (**1997**). "Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system," J. Acoust. Soc. Am. **101**, 3636–3647.

Chatterjee, M., and Zwislocki, J. J. (**1997**). "Cochlear mechanisms of frequency and intensity coding. I. The place code for pitch," Hear. Res. **111**, 65–75.

Davis, H. (**1983**). "An active process in a cochlear mechanics," Hear. Res. **9**, 79-90.

Gantz, B. J., and Turner, C. W. (**2003**). "Combining acoustic and electrical hearing," Laryngoscope **113**, 1726–1730.

Gantz, B. J., and Turner, C. W. (**2004**). "Combining acoustic and electrical speech processing: Iowa/Nucleus Hybrid implant," Acta Oto-Laryngol. **124**, 334–347.

Gantz, B. J., Turner, C. W., and Gfeller, K. E. (**2006**). "Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/Nucleus Hybrid implant," Audiol. Neuro-Otol. **11**, 63–68.

Gantz, B. J., Turner, C. W., Gfeller, K. E., and Lowder, M. (**2005**). "Preservation of hearing in cochlear implant surgery: advantages of combined electrical and acoustical speech processing," Laryngoscope **115**, 796–802.

Gifford, R. H., Dorman, M. F., Spahr, A. J., and Bacon, S. P. (**2007**). "Auditory function and speech understanding in listeners who qualify for EAS surgery," Ear Hear. **28**, 114S–118S.

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (**2008**). "Speech recognition materials and ceiling effects: Considerations for cochlear implant programs," Audiol. Neuro-Otol. **13**, 193–205.

Gstoettner, W. K., Helbig, S., Maier, N., Kiefer, J., Radeloff, A., and Adunka, O. (**2006**). "Ipsilateral electric acoustic stimulation of the auditory system: Results of long-term hearing preservation," Audiol. Neuro-Otol. **11**, 49–56.

Gstoettner, W., Kiefer, J., Baumgartner, W. D., Pok, S., Peters, S., and Adunka, O. (**2004**). "Hearing preservation in cochlear implantation for electric acoustic stimulation," Acta Oto-Laryngol. **124**, 348–352.

Gstoettner, W., Pok, S. M., Peters, S., Kiefer, J., and Adunka, O. (**2005**). "Cochlear implantation with preservation of residual deep frequency hearing," HNO **53**, 784–791.

Kiefer, J., Pok, M., Adunka, O., Stuerzebecher, E., Baumgartner, W. D., and Schmidt, M. (**2005**). "Combined electric and acoustic stimulation of the auditory system: Results of a clinical study," Audiol. Neuro-Otol. **10**, 134–144.

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.

Laroche, C., Quoc, H. T., Josserand, B., and Glasberg, B. (**1992**). "Frequency selectivity in workers with noise-induced hearing loss," Hear. Res. **64**, 61–72.

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.
Luetje, C. M., Thedinger, B. S., Buckler, L. R., Dawson, K. L., and Lisbona, K. L. (**2007**). "Hybrid cochlear implantation: Clinical results and critical review of 13 cases," Otol. Neurotol. **28**, 473–478.
McFadden, D., and Chamlin, C. A. (**1990**). "Reductions in overshoot during aspirin use," J. Acoust. Soc. Am. **87**, 2634–2642.
Neely, S. T., and Kim, D. O. (**1986**). "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1780.
Oxenham, A. O., and Bacon, S. P. (**2004**). "Psychophysical manifestations of compression: Normal-hearing listeners," In *Compression from Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer-Verlag, New York).
Oxenham, A. O., and Dau, T. (**2001**). "Reconciling frequency selectivity and phase effects in masking," J. Acoust. Soc. Am. **110**, 1525–1538.
Oxenham, A. O., 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.
Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear Disord. **27**, 62–70.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.
Schroeder, M. R. (**1971**). "Synthesis of low peak-factor signal and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.
Skarzynski, H., Lorens, A., and Piotrowska, A. (**2003**). "A new method of partial deafness treatment," Med. Sci. Monit., **9**, CS20–24.
Skarzynski, H., Lorens, A., and Piotrowska, A. (**2004**). "Preservation of low frequency hearing in partial deafness cochlear implantation (PDCI) using the round window surgical approach," Acta Oto-Laryngol. **127**, 41–48.
Skarzynski, H., Lorens, A., Piotrowska, A., and Anderson, I. (**2006**). "Partial deafness cochlear implantation provides benefit to a new population of individuals with hearing loss," Acta Oto-Laryngol. **126**, 934–940.
Skarzynski, H., Lorens, A., Piotrowska, A., and Anderson, I. (**2007**). "Preservation of low frequency hearing in partial deafness cochlear implantation (PDCI) using the round window surgical approach," Acta Oto-Laryngol. **127**, 41–48.
Stuart, A., Stenstrom, R., Tompkins, C., and Vandenhoff, S. (**1991**). "Test-retest variability in audiometric threshold with supraaural and insert earphones among children and adults," Audiology **30**, 82–90.
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. (**2001**). "Overshoot effects using Schroeder-phase harmonic maskers in listeners with normal hearing and with hearing impairment," Hear. Res. **162**, 1–9.
Summers, V., and Leek, M. R. (**1998**). "Masking of tones by Schroeder-phase harmonic complexes in normal hearing and hearing-impaired listeners," Hear. Res. **118**, 139–150.
von Ilbert, C., Kiefer, J., Tillein, J., Pfennigdorff, T., Hartmann, R., Stuerzebecher, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system," ORL **61**, 334–340.
Wilson, B., Wolford, R., Lawson, D., and Schatzer, R. (**2002**). "Speech processors for auditory prostheses," Third quarter progress report on NIH Project No. N01-DC-2—1002.

The influence of noise exposure on the parameters of a convolution model of the compound action potential

M. E. Chertoff[a] and J. T. Lichtenhan[b]
Department of Hearing and Speech, University of Kansas Medical Center, Kansas City, Kansas 66160

B. M. Tourtillott and K. S. Esau
Department of Hearing and Speech, University of Kansas Medical Center, Kansas City, Kansas 66160

(Received 11 March 2008; revised 3 July 2008; accepted 8 July 2008)

The influence of noise exposure on the parameters of a convolution model of the compound action potential (CAP) was examined. CAPs were recorded in normal-hearing gerbils and in gerbils exposed to a 117 dB SPL 8 kHz band of noise for various durations. The CAPs were fitted with an analytic CAP to obtain the parameters representing the number of nerve fibers (N), the probability density function [$P(t)$] from a population of nerve fibers, and the single-unit waveform [$U(t)$]. The results showed that the analytic CAP fitted the physiologic CAPs well with correlations of approximately 0.90. A subsequent analysis using hierarchical linear modeling quantified the change in the parameters as a function of both signal level and hearing threshold. The results showed that noise exposure caused some of the parameter-level functions to simply shift along the signal level axis in proportion to the amount of hearing loss, whereas others shifted along the signal level axis and steepened. Significant changes occurred in the $U(t)$ parameters, but they were not related to hearing threshold. These results suggest that noise exposure alters the physiology underlying the CAP, some of which can be explained by a simple lack of gain, whereas others may not.
 [DOI: 10.1121/1.2967890]

PACS number(s): 43.64.Nf, 43.64.Pg, 43.64.Ri, 43.64.Wn [BLM] Pages: 2174–2185

I. INTRODUCTION

Sensorineural hearing loss results from anatomical changes in the cochlea and auditory nerve. One problem with this type of hearing loss is that damage to the cochlea and auditory nerve varies greatly among patients (Schuknecht, 1988), and clinical techniques that distinguish and quantify the anatomical changes are lacking. Although audiometric threshold measurements are clinically valuable, they fail to provide sufficient information regarding underlying cochlear pathologies and neural integrity (e.g., see Salvi *et al.*, 1983). New techniques that distinguish sensory from neural pathologies and that also estimate the extent of anatomical damage would be useful in the clinical setting to provide a site of lesion as hair cell regeneration and genetic therapy are developed and may perhaps drive advancements in signal processing of hearing aids and cochlear implants.

By studying physiological components of the compound action potential (CAP), it might be possible to develop an assay of neural integrity in the presence of sensory and/or neural hearing loss (e.g., see Hall, 1990). Because its properties reflect the time-delayed sum of neural discharges, the CAP can be described as the convolution of a probability density function and a single-unit waveform. As proposed by Goldstein and Kiang (1958),

$$\mathrm{CAP}(t) = N \int_{-\infty}^{t} P(\tau) U(t-\tau) d\tau, \tag{1}$$

where $P(t)$ represents a probability density function of single-unit discharges from a population of neurons, $U(t)$ represents a single-unit waveform recorded extracellulary, and N is the number of neurons contributing to the CAP.

Equation (1) has been used successfully to model the CAP waveform. For example, Wang (1979) recorded $U(t)$ from an electrode on the round window using a spike-triggered averaging technique. $P(t)$ was obtained from post-stimulus time histograms from single auditory nerve fibers with a wide range of characteristic frequencies. Convolving $P(t)$ and $U(t)$ produced a simulated CAP similar to the CAP recorded by an electrode on the round window. The results of Wang (1979) are similar to those of other investigators who used the spike-triggered averaging technique (Versnel *et al.*, 1992a) as well as modeling approaches of the CAP (de Boer, 1975). Minor variations between model CAPs and physiologic CAPs, however, do exist but are limited to the later peaks in the CAP waveform. Together, the results from the previous studies indicate that the convolution model [Eq. (1)] is a reasonable description of the relation between single-unit events and the whole-nerve CAP.

Equation (1) has also been used to obtain $P(t)$ and $U(t)$, given a recorded CAP. Using a recorded CAP and assumed $U(t)$, Elberling (1976) derived $P(t)$ by deconvolving the CAP by $U(t)$. Chertoff (2004) obtained $P(t)$ and $U(t)$ in Eq. (1) by convolving functional forms for $P(t)$ and $U(t)$ and fitting the results to recorded CAPs. $P(t)$ was defined as

[a]Author to whom correspondence should be addressed.
[b]Present address: Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114.

$$P(t) = \left(\frac{t-\alpha}{\beta}\right)^{\gamma-1} e^{(t-\alpha/\beta)}, \quad (2)$$

for $t \geqslant \alpha$, 0 otherwise. Parameter t is time and α represents the time delay from the stimulus onset to the onset of the CAP. When $\gamma=2$, $P(t)$ reaches its maximum at $t-\alpha=\beta$. Beta also influences the width of $P(t)$. The unit waveform was given as

$$U(t) = e^{-kt}\sin(\omega t), \quad (3)$$

where K was the decay constant and $\omega=2\pi F$. The analytic solution was fitted to recorded CAPs, and the parameters describing $P(t)$ and $U(t)$ were examined as a function of stimulus frequency and signal level. The results showed that α decreased with increasing signal level, suggesting a spread of excitation to fibers with high characteristic frequencies, and N and β (at 2, 4, 8, and 16 kHz) increased with signal level, illustrating a change in the number of fibers and synchrony among auditory nerve fibers contributing to the CAP.

One could envision that cochlear and/or neural damage could lead to changes in N, $P(t)$, or $U(t)$. Some damage may result in changes to the CAP that can be quantified with conventional measures such as amplitude and latency, whereas others may not. For instance, as inner hair cells die and the pillar cells collapse, auditory nerve fibers degenerate (Morest *et al.*, 1998). A reduction in the number of auditory nerve fibers should reduce the amplitude of the CAP and also decrease the magnitude of parameter N of Eq. (1). In contrast, if the timing of discharges among auditory fibers is altered, e.g., becoming less synchronous, it may be reflected in the summed probability density function as a widening of width of $P(t)$. Similarly, alterations in ionic conductance or changes in myelin due to genetic abnormalities or disease processes may affect $U(t)$. Importantly, some of these alterations may affect the auditory threshold and may manifest themselves only at certain signal levels or, perhaps, as signal level changes. The long-term goal of this investigation is to determine if the parameters of the convolution model of the CAP, and their change as a function of signal level, can be used as an assay of the extent of neural damage. The purpose of this study was (1) to determine if the CAP convolution model could adequately fit CAPs recorded from animals with peripheral auditory damage and (2) to determine if alterations in the growth functions of the model parameters could be explained by elevations in the hearing threshold.

II. METHODS

A. Subjects, animal preparation, and data acquisition

The CAP data from 40 animals used in this study were from Chertoff *et al.* (2003). Thus the stimuli, surgical methods, and data acquisition are only briefly explained. CAPs were collected from Mongolian gerbils (*Meriones unguiculatus*) weighing between 40 and 60 g. Once hearing was confirmed to be within normal limits using auditory brainstem responses (ABRs), the animals were assigned to either a control ($n=6$) or experimental group ($n=34$) to receive noise exposure for durations of either 0, 1, 1.5, 2, 4, 8, 16, 64, or 128 h. The noise exposure signal was a 3 kHz bandwidth noise centered at 8 kHz and presented at 117 dB SPL. After exposure, ABRs were recorded weekly for up to 7 weeks or until a permanent hearing loss was confirmed. For CAP recordings, animals were sedated with Nembutal (64 mg/kg), the bulla was opened, and an electrode was placed on the round window. A secondary electrode was inserted into the neck musculature and served as ground. Body temperature was maintained at 37 °C using a heating pad with a rectal thermometer (Harvard). CAPs were recorded to 2 ms tone bursts (1, 2, 4, 8, and 16 kHz) windowed within a $\cos^2$ function, yielding a 1 ms rise/fall time. The signal level started at 100 dB SPL and descended in 5 dB steps until 15 dB SPL. All acoustic signals were created in an array processor (Tucker-Davis Technologies, TDT). The electrophysiologic signals were amplified 5000 times (Stanford SR560 Steward VBF10), filtered (0.03–25 kHz), digitized (TDT, AD2), averaged, and stored on disk for offline analysis. A CAP consisted of at most 500 signal presentations. All animal protocols were approved by the University of Kansas Medical Center Institutional Animal Care and Use Committee.

B. Curve fitting

The analytic CAP (Chertoff, 2004) was fitted to all recorded CAPs (in millivolts) using a constrained nonlinear least squares fitting routine (TOMLAB). Lower and upper bound limits were imposed on the convolution model parameters to improve the convergence of the fitting algorithm and provide a unique solution. $P(t)$ required boundary conditions for two parameters: β and α; gamma was set to 2 and was not a free parameter. For β, the lower boundary was set to 0.02 ms and was based on the minimum jitter of a single neuron [Miller *et al.* (1999), Fig. 5(b)]. The maximum value was set to 2.4 ms and includes the maximum jitter [~1.0 ms, Miller *et al.* (1999)] and the cochlear delay of 1.4 ms for the 12 mm section of the gerbil cochlea. The minimum value for α was set to 1 ms and represents the delay of the acoustic tube connecting the headphone to the ear canal. The maximum value was chosen to be 5 ms.

In addition to varying the $P(t)$ parameters, the parameters of $U(t)$ were also allowed to vary in this study. Our motivation was to allow for the variation in the amplitude and time between peaks that was reported by Wang (1979). The boundaries for the frequency parameter, F, were determined from the data Tables I and II of Versnel *et al.* (1992b) who considered the width of the initial negative peak (σ_n) and the subsequent positive peak (σ_p) of $U(t)$ to be the standard deviation of a Gaussian function fitted separately to each peak. From their Tables I and II, we chose the values of $\sigma_n=0.15\pm0.10$ ms and $\sigma_p=0.23\pm0.12$ ms for the width of the two peaks because of the large number of nerve fibers used to estimate these values. The lower and upper boundaries of F were defined as $F_l=[3(\sigma_n+0.10+\sigma_p+0.12)]^{-1}$ and $F_u=[3(\sigma_n-0.10+\sigma_p-0.12)]^{-1}$, respectively, where F_l and F_u are the lower and upper frequencies in kilohertz. The 3 in these equations represent three standard deviations of the Gaussian distribution. The calculated values for the boundaries were $F_l=0.555$ kHz and $F_u=2.083$ kHz and were rounded to $F_l=0.5$ kHz and $F_u=2.0$ kHz. The boundaries for

TABLE I. Coefficient values for *N*-level growth functions, goodness of fit, and variability among animals.

Frequency (kHz)	A_0	B_0	A_1	B_1[a]	A_2[a]	A_3[a]	R^2	σ
16	0.796	−0.087	0	0.778	0	0	0.859	0.094
8	1.240	−0.150	0.048	1.300	−0.776	0.003	0.886	0.109
4	−4.152	−0.017	0.195	0	−2.733	0.013	0.869	0.112
2	−4.394	0	0.142	0	−1.478	0.005	0.934	0.060
1	−4.475	0	0.137	0	−1.413	0.005	0.947	0.038

[a]Values $\times 10^{-3}$.

the decay constant of $U(t)$, i.e., K, were set to 0.5 ms for a lower boundary and to 2.0 ms for an upper boundary. The 0.5 ms value was chosen because it improved the fit of the analytic CAP to the physiologic CAP recorded at the highest signal level. The 2 ms upper boundary value was based on model $U(t)$ using an exponentially decaying sine function by McMahon and Patuzzi (2002). The boundaries for N were set to a minimum of .001 because our previous experience (Chertoff, 2004) showed that values below this were not statistically different from zero. The maximum N was chosen to be infinite. Because we did not scale $U(t)$ and we fit the CAP in units of millivolts, N is a dimensionless scalar that, according to the model, should change when the number of nerve fibers contributing to the CAP changes.

Initial starting values for the parameters and boundary values were provided to the constrained nonlinear optimization routine (TOMLAB), and CAP parameters from Eqs. (1)–(3) were estimated. Using the Jacobian of the parameters and the residuals between the fit and the physiologic CAP, standard errors and confidence intervals for each parameter were obtained. If any of the CAP parameters were not statistically significant (i.e., if the confidence interval included zero), the initial parameters were varied and the curve fitting repeated. This procedure was repeated until statistically significant parameters were obtained. Using the constrained procedure, we were able to obtain statistically significant parameters at signal levels that were lower than our previous study (Chertoff, 2004). The parameters analyzed as a function of signal level and auditory threshold in this study were all statistically significant at $p<0.05$.

The covariation of parameters and accuracy of the curve fitting was examined by creating simulated CAPs using all possible combinations of five values for each of the five parameters resulting in 3125 CAPs. Each simulated CAP was fitted with the analytic CAP, and the resulting parameter values were compared to the actual parameters used to create the simulated CAPs. In no instance did the change in one parameter influence or correlate with the change in another parameter, suggesting independence of the parameters. Error in the curve fitting procedure was estimated by error=((act −pred)/act)100. The maximum error was 0.5% and was associated with the parameter K.

C. Data analysis

The goal was to obtain a mathematical description for the influence of signal level and auditory threshold on each CAP parameter's growth function (i.e., parameter-level function). The parameter-level functions were fitted with a polynomial equation having coefficients that varied as a function of hearing threshold. That is,

$$\log_{10} Z(x,y) = \sum_{n=0}^{3} C_n x^n, \tag{4}$$

$$C_n = A_n + B_n y,$$

where Z is a given CAP parameter, C_n is a polynomial coefficient, x is the signal level in dBSPL and y is the hearing threshold in dBSPL. Hearing threshold was defined as the average CAP threshold at 4, 8, and 16 kHz and will be referred to as pure tone average (PTA). These frequencies were chosen because this was the frequency range where the majority of hearing loss was present. A threshold was defined as the lowest signal level at which a CAP could be visually detected.

As indicated in Eq. (4), C_n depends on A_n and B_n, which are to be determined from the data. This was accomplished using hierarchical linear modeling (HLM) (Raudenbush and Bryk, 2002) and the statistical software HLM 6 hierachical linear and nonlinear modeling (Raudenbush *et al.*, 2000). HLM was chosen instead of ordinary least squares regression because HLM uses an empirical Bayes estimation procedure (also known as shrinkage estimators) or a maximum likelihood approach to estimate the coefficients for each animal. These estimates are more accurate than ordinary least squares when sample sizes differ, which is an inherent condition in our study because animals differed in their thresholds after noise exposure and hence differed in the number of data points used in the estimate of the coefficients. The average value of the coefficients is reported in Tables I and II.

For an initial analysis, Eq. (4) with an intercept (C_0) and linear coefficient (C_1) were used to fit the logarithm of a CAP parameter to the signal level. Additional higher-order terms were added sequentially if they significantly improved the fit ($p<0.05$). Only terms that were statistically significant were included in the complete model, and the best model was defined as the one with the lowest deviance score (Raudenbush and Bryk, 2002).

To determine if Eq. (4) provided an adequate description of the data for each animal, a coefficient of determination (R^2) was computed. This was accomplished by obtaining a "predicted" growth function for each animal using an animal's own coefficients (i.e., slope, intercept, etc.). The predicted growth function was correlated with the animal's ac-

TABLE II. Coefficient values for ∝-level growth functions, goodness of fit, and variability among animals.

Frequency (kHz)	A_0	B_0	A_1	B_1[a]	A_2[a]	R^2	σ
16	0.206	0.011	−0.010	0	0	0.909	0.146
8	0.223	0.008	−0.007	0	0	0.888	0.206
4	0.313	0.007	−0.008	0	0	0.919	0.141
2	0.466	0	−0.005	4.800	−3.800	0.969	0.024
1	0.490	0	−0.005	2.900	−4.300	0.972	0.020

[a]Values $\times 10^{-5}$.

tual data. The R^2 value from each animal was averaged across all animals for each of the stimulus frequencies. The mean and variance of R^2 are reported in the tables.

III. RESULTS

A. Hearing threshold

A relevant issue for determining the relation between the hearing threshold and the change in the parameter-level functions is the amount of variation in hearing thresholds. To adequately derive a relation, it is important to have a wide range of thresholds. This was accomplished in our previous study (Chertoff *et al.*, 2003) by exposing animals to different durations of noise. The PTAs of the animals used in the analysis for the present study are illustrated in Fig. 1 and show the range of hearing thresholds from normal hearing to hearing thresholds as high as approximately 95 dBSPL.

B. Fit of the CAP convolution model

Previously, the convolution model of the CAP was shown to adequately fit physiologically recorded CAPs from normal-hearing animals (Chertoff, 2004). In the present study, we were interested in determining if the fit varied with signal level and the amount of hearing loss. The dependent variable in the analysis was the correlation between the analytic CAP and the physiologic CAP, and the independent variables were signal level and PTA. Equation (4) was used for the analysis with a slight modification; the correlation value was considered on a linear scale (i.e., not logarithm). The coefficients were estimated by HLM. The mean (±1SD) correlation coefficients are illustrated in Fig. 2. For the majority of signal levels and frequencies, the correlation values were above 0.9. The HLM analysis indicated that a small cubic polynomial trend occurred across the signal levels for each frequency, but the coefficients did not vary with PTA, indicating that the fit of the analytic CAP to the physiologic CAP was not related to the amount of hearing loss.

C. Representative animals

Figure 3 illustrates recorded CAPs (gray lines) and analytic CAPs (black lines) to a 16 kHz tone burst for a gerbil from the control group (left panel) and the noise-exposed group (right panel) at 100 dBSPL. In both animals, the analytic CAPs fitted the recorded CAPs quite well, with coefficients of determination (i.e., R^2) of 0.86 and 0.87, respectively. There were, however, differences between the normal and noise-exposed animal. The CAP from the noise-exposed animal was smaller (notice the change in the scale of the figure) than that from the normal animal, and the first negative peak was slightly delayed compared to the delay in the normal animal.

These changes were reflected in the parameters of Eq. (1)–(3) (Fig. 4). N was smaller in the noise-exposed animal and increased at a faster rate with signal level than in the normal animal. The delay parameter α showed a similar lin-

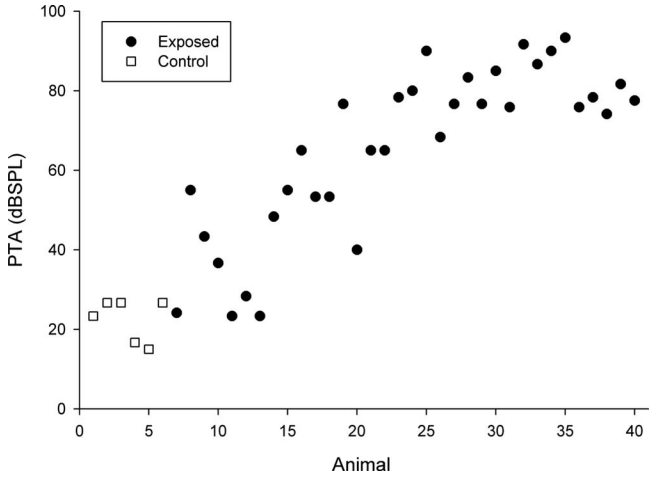


FIG. 1. PTA thresholds computed from 4, 8, and 16 kHz for each animal in this study. The open symbols represent the animals that did not receive noise exposure. The filled symbols are PTAs from noise-exposed animals.

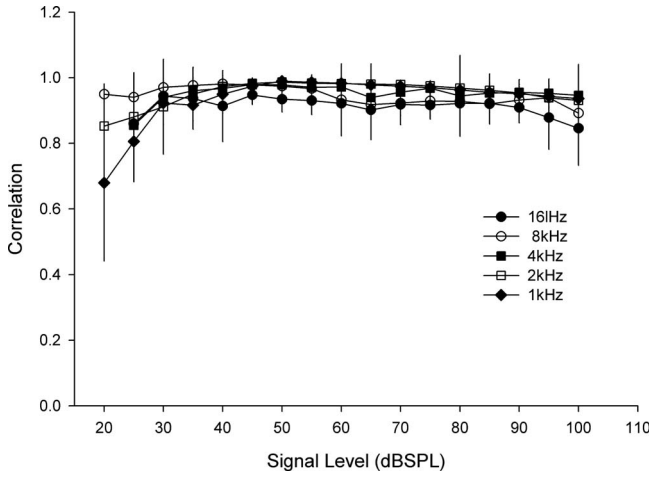


FIG. 2. Symbols represent the mean correlation computed from all of the animals between the analytic CAP and the physiologic CAPs. The error bars indicate ±1 standard deviation.

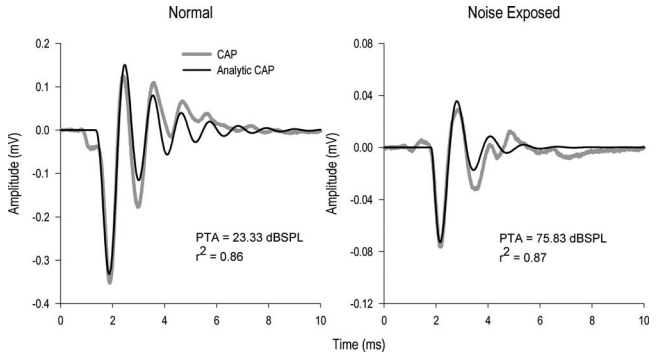


FIG. 3. CAPs from a control (left panel) and noise-exposed animal (right panel). Gray lines are the physiologic recorded CAPs, and the black lines are the analytic CAPs. PTA is the average CAP threshold at 4, 8, and 16 kHz. The fits between the physiologic CAPs and analytic CAPs are indicated by R^2.

ear trend with signal level between the two animals. However, the noise-exposed animal's delay was shifted to longer delays.

Other changes in the parameters of $P(t)$ and $U(t)$ occurred, although they were not immediately apparent in the CAP waveforms of Fig. 3. Beta was smaller in the noise-exposed animal and, in contrast with the control animal, decreased with increasing signal level. K, the decay parameter of $U(t)$, remained constant for many of the signal levels but then decreased at the highest level in the control animal. No trend in the K-level function was found for this noise-exposed animal. In the normal animal, F, or the frequency of $U(t)$, rose quickly from ~800 to ~950 Hz and then remained constant as the signal level increased. A similar trend was noted for the noise-exposed animal although the overall value of F was reduced.

D. Parameter *N*

The influence of the signal level on N is illustrated in Fig. 5. For the control group (dashed lines), the growth functions were similar among animals. Given the log-log axis, N grew linearly at low signal levels and saturated at high signal levels. By contrast, the noise-exposed animals (solid lines) were less homogeneous than the control animals. At high frequencies (16, 8, and 4 kHz), the growth functions shifted to higher sound pressures, and in some cases (16 and 8 kHz),

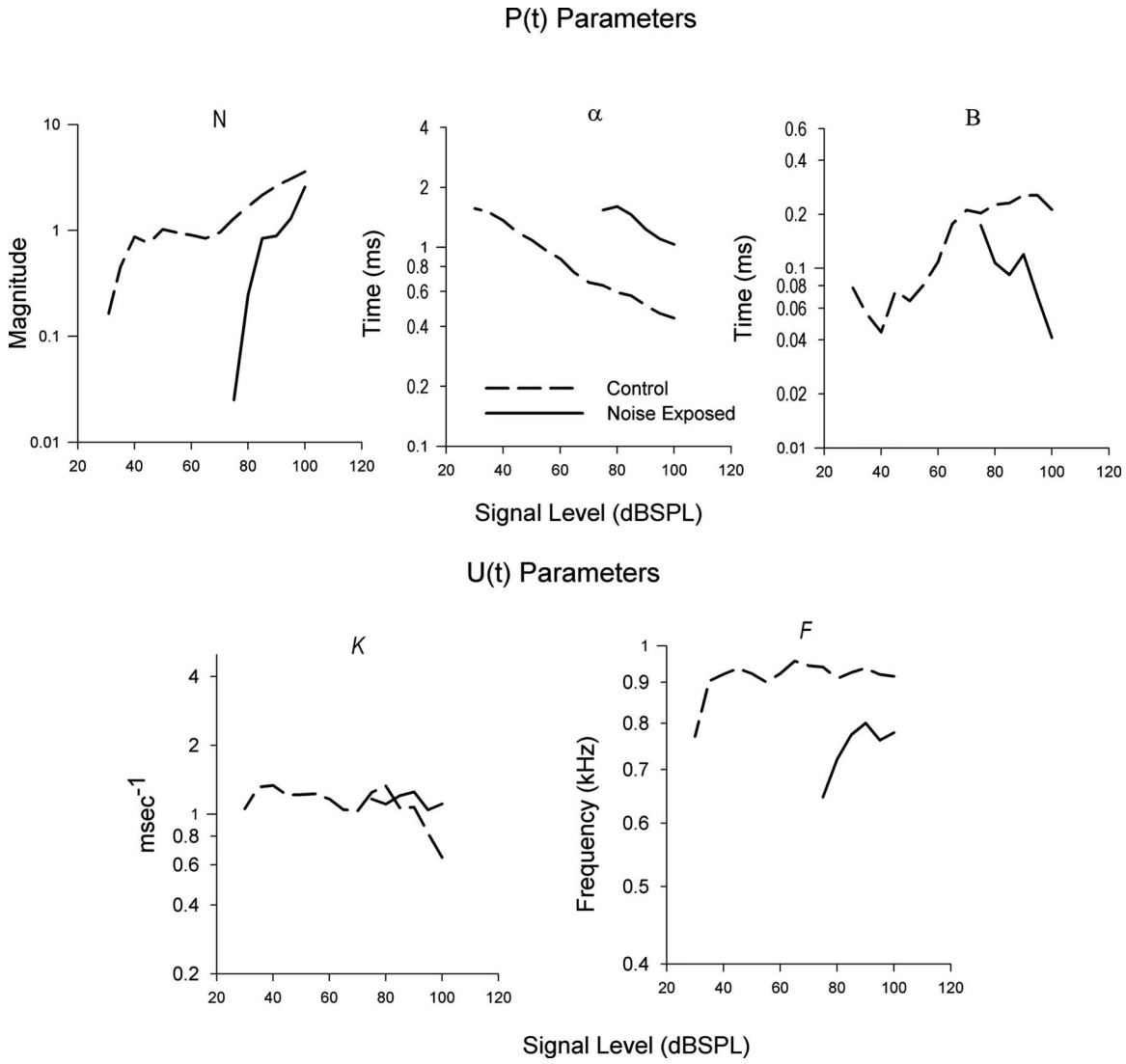


FIG. 4. The change with signal of the parameter values from Eqs. (1) and (2) [$P(t)$, top panels] and Eq. (3) [$U(t)$, bottom panels] for the two animals presented in Fig. 3. The stimulus was 16 kHz. Dashed lines indicate the control animal, and the solid lines indicate the noise-exposed animal.

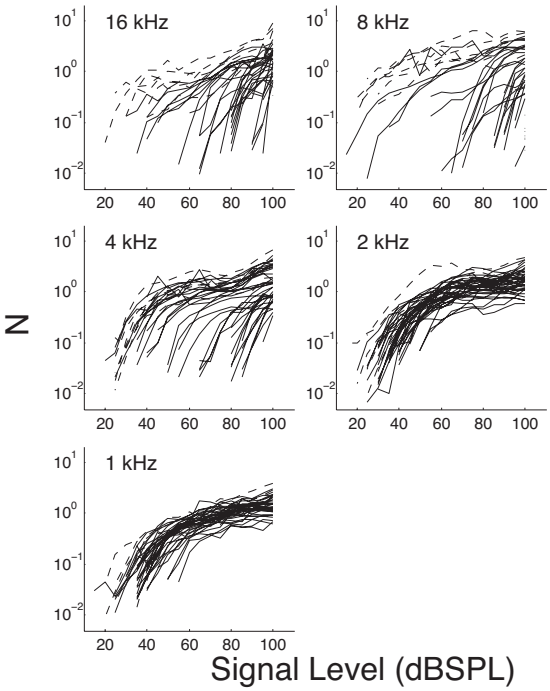


FIG. 5. The change in N with signal level for the control group animals (dashed lines) and noise-exposed animals (solid lines). Each line represents one animal's N-level function.

the functions steepened. For low frequencies (2 and 1 kHz), N changed similarly for both the control and noise-exposed groups.

Figure 6 illustrates the solution to Eq. (4) using each animal's coefficients. Equation (4) fitted the data well; R^2 ranged from a low of 0.86 at 16 kHz to a high of 0.95 at 1 kHz with a small variance among animals. The good fit of Eq. (4) to the data indicates that some of the variability in the N-level functions in the noise-exposed animals can be accounted for by the hearing threshold. For example, the best fit to the 16 and 8 kHz data required the intercept (C_0) and linear term (C_1) to vary as a function of PTA. As shown in Table I (average value of the coefficients), C_0 decreased (i.e., a negative B_0) and C_1 increased (i.e., a positive B_1) with an increase in PTA. For both frequencies, the decrease in C_0 with PTA indicated that the N-level functions shifted along the signal level axis in animals with elevated thresholds. Moreover, the greater the threshold elevation, the greater the shift. The increase in C_1 with PTA indicated that as hearing loss became more severe, the slope of the N-level function steepened. By contrast, at 4 kHz, only the intercept (C_0) decreased as a function of PTA. Thus for many of the animals, the N-level growth function remained in the same shape even as hearing loss became more severe. The N-level function simply shifted along the signal level axis in an amount proportional to PTA. For the lower frequencies (2 and 1 kHz), PTA did not significantly contribute to the polynomial description of the N-level function. Thus, a third-order polynomial equation with constant coefficients described these N-signal level functions.

E. Parameter α

The delay parameter of $P(t)$, α, changed with signal level and the shape of the α-level functions varied with frequency (Fig. 7). In both the control and noise-exposed groups α decreased with increasing signal level for the 4, 8, and 16 kHz signals. At 1 and 2 kHz, a slight curvature in the α-level function was present. Similar to N, a greater variation in the growth functions occurred for the noise-exposed animals than the control animals. In the noise-exposed group, the functions showed longer delays in the high frequencies (16, 8, and 4 kHz) and less curvature in the functions for the low frequencies as compared to the control animals.

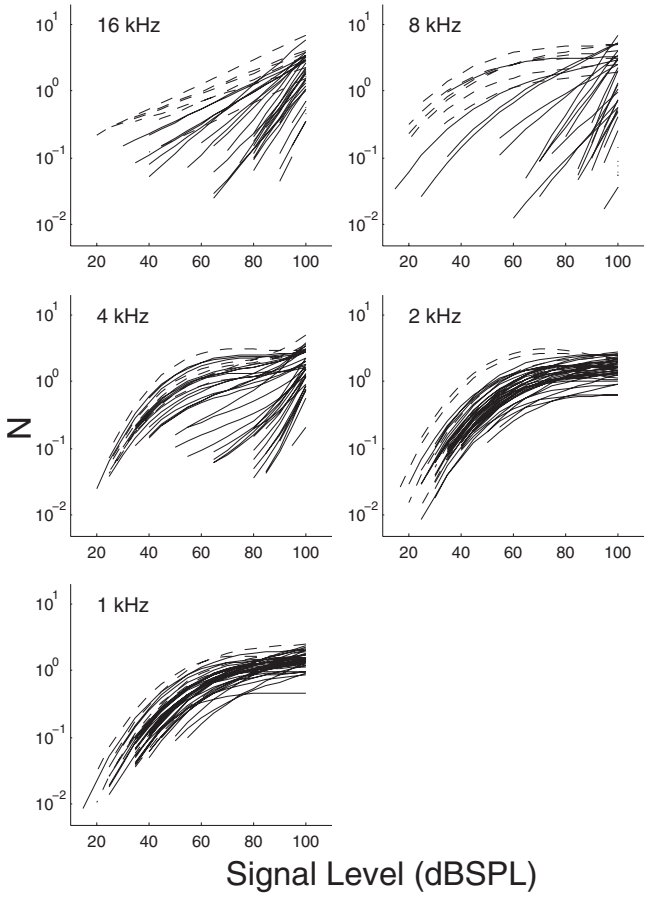


FIG. 6. Estimated N-level functions using Eq. (4) with the coefficients from each individual animal. Control animals are indicated by dashed lines, and noise-exposed animals are represented by solid lines.

Much of the variation in the noise-exposed group could be described by Eq. (4). Figure 8 shows the predicted α-level functions. The fit of Eq. (4) to each animal's data was slightly higher than N, with an R^2 that ranged from a low of 0.89 at 8 kHz to a high of 0.97 at 1 kHz. The variation among animals was small (SD of R^2 ranged from 0.02 to 0.12). Table II shows that at 4, 8, and 16 kHz the best fit to the data was a linear function (on a log-log scale) with an intercept (C_0) that varied with PTA, indicating that the α-level functions shifted to larger delays as PTA increased (positive B_0). By contrast, at 1 and 2 kHz, the linear term (C_1) of the concave-downward quadratic functions increased with PTA (positive B_1), demonstrating that the slope of the α-level function decreased as PTA increased.

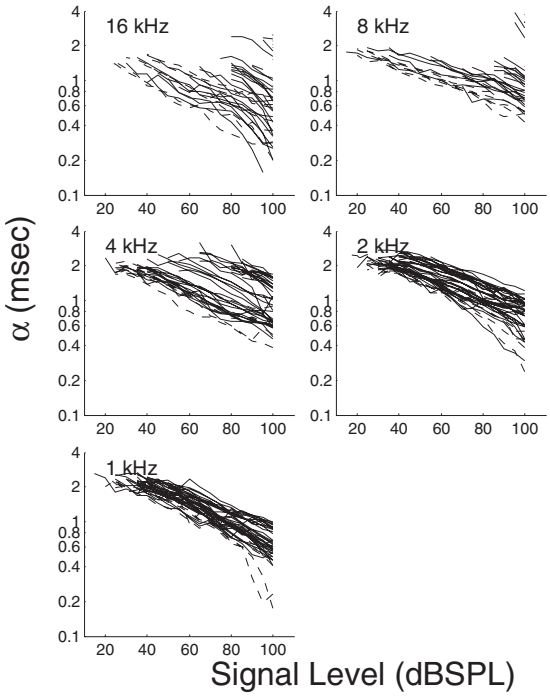


FIG. 7. The change in α as a function of signal level for the control (dashed lines) and noise-exposed animals (solid lines).

F. Parameters β, K, and F

Equation (4) provided a poorer description of the influence of hearing threshold on the growth functions for β, K, and F compared to those for the other parameters. R^2 ranged from a low of 0.21 for β at 1 kHz to a high value of 0.64 for F at 1 kHz. The average standard deviation in the fit of Eq. (4) to the parameter-level function among animals was larger than N and α. In some conditions the standard deviation was as large as the average R^2, indicating wide variation in the ability of Eq. (4) to describe the parameter-level functions from the animals. Thus, instead of describing the relation between the hearing threshold and the growth functions, we examined if the parameters from normal-hearing animals differed from hearing-impaired animals. The animals were divided into two groups; those with PTAs $\leqslant$ 30 dBSPL were considered to have normal hearing and those with PTAs $>$ 30 dBSPL were to have hearing loss. This cut point was determined from our database of CAP thresholds in normal animals from our laboratory. The parameters were evaluated at two signal levels; 100 dBSPL and signal level at the threshold. Independent t-tests were computed between the normal-hearing and hearing-impaired animals for each frequency. The type I error was controlled by dividing 0.05 by five comparisons, thus providing a conservative p value for statistical significance (i.e., 0.01). Levene's test of homogeneity was used to examine the assumption of equal variability, and if significant, individual group variances and an adjusted degree of freedom were used for the t-test (SPSS version 15).

Beta was larger in the hearing-impaired animals than in the normal-hearing animals for the 4, 8, and 16 kHz stimuli (Fig. 9) at both signal levels, reaching statistical significance only at the threshold signal level. At 1 and 2 kHz, there was no difference between the two groups at either signal level. For the parameters F and K there were some minor effects of hearing loss. At 16 kHz only, F increased from an average of 879 Hz in normal-hearing animals to 1077 Hz in impaired animals (p=0.004). The parameter K was larger in the hearing-impaired animals as compared to the normal-hearing animals at 1 kHz, but only for the threshold signal level (p =0.004). The difference was 0.355 ms.

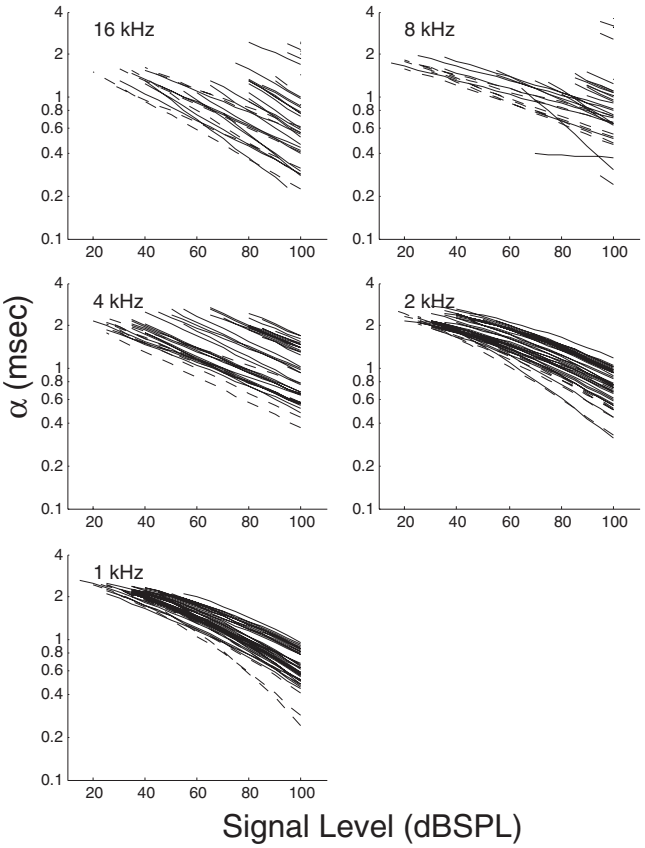


FIG. 8. Predicted α-level functions from Eq. (4) using each animal's coefficients. Dashed and solid lines indicate control and noise exposed animals, respectively.

IV. DISCUSSION

A. Convolution model

The first goal of this investigation was to determine if the analytic CAP fitted well to the physiologic CAPs in

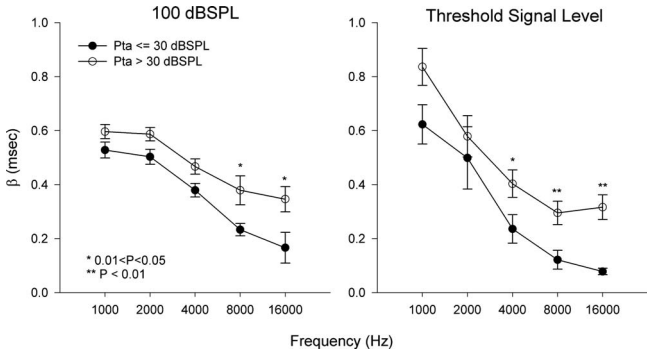


FIG. 9. Parameter β at 100 dBSPL (left panel) and at the threshold signal level (right panel). Filled circles are for animals with PTAs $\leqslant$ 30 dBSPL, and open circles are for animals with PTAs $>$ 30 dBSPL. The error bars represent the standard error of the mean.

noise-damaged animals and to examine the influence of signal level and hearing threshold. The analytic CAP fitted the physiologic CAPs with correlation values that were generally larger than 0.90 across most of the signal levels. Moreover, the fit of the analytic CAP did not vary with the amount of hearing loss. This indicates that on average, the analytic CAP described a large proportion (88% R^2, collapsed across levels and frequencies) of the variance in the CAP waveform morphology, independent of the amount of hearing loss. Although subjective, we felt that the correlations were large enough to consider the analytic CAP a good description of the physiologic CAP.

B. Parameter *N*

Parameter *N* grew nonlinearly with signal level (log-log scale) for the majority of frequencies (1, 2, 4, and 8 kHz). The functions grew steeply at low and high signal levels and were compressive at middle signal levels. This growth pattern is reminiscent of the nonlinear growth pattern of the basilar membrane with signal level (Robles and Ruggero, 2001). For 1 and 2 kHz tone burst stimuli, a significant third-order polynomial equation was required to fit the growth functions. The coefficients did not vary with the degree of hearing loss, indicating that high frequency hearing loss did not alter the low frequency *N*-level functions. At 4 kHz, the growth functions were altered by noise exposure. The best fit to the data was a third-order polynomial equation with an intercept that varied with the degree of hearing loss. This indicates that the *N*-level functions were not altered in shape but simply shifted horizontally. This suggests that after the signal level reaches the threshold, auditory nerve fibers contributing to the CAP respond with a growth function similar to that seen in animals with normal auditory sensitivity. That is, the change in the growth function appears to be caused by a simple lack of gain. The growth functions for both the 16 and 8 kHz stimuli also required coefficients that changed with the degree of hearing loss. Both the intercept and linear terms were related to the auditory threshold. The intercept decreased with hearing threshold, representing a shift along the signal level axis, and the linear term increased proportionally with animals' PTAs, indicating that the greater the hearing loss the steeper the *N*-level functions. A steeper slope is consistent with some reports on the input-output function of the amplitude of the N1 of the CAP in noise-exposed animals (Elberling and Salomon, 1976; Salvi *et al.*, 1983; Wang and Dallos, 1972) and humans with sensorineural hearing loss (Yoshie and Ohashi, 1969; Portmann *et al.*, 1973). However, these reports are by no means conclusive, as other investigators reported no change in the slope or a decrease in the slope in the N1-amplitude functions (Popelar *et al.*, 1987; Dolan and Mills, 1989).

The increase in the slope of the high frequency (i.e., 8 and 16 kHz) *N*-level functions in hearing-impaired animals suggests a recruitmentlike phenomenon. This could result from the broadening of the basilar membrane tuning following noise exposure, leading to the excitation of fibers from remote cochlear regions. Other mechanisms for recruitment, as discussed by Heinz *et al.* (2005), include an increase in the discharge rate of auditory nerve fibers or a compression in the threshold distribution of auditory nerve fibers. However, analyzing a population of auditory nerve fibers from noise-exposed cats, Heinz *et al.* (2005) found no evidence for an increase in the discharge rate of individual fibers or change in the threshold distribution of auditory nerve fibers. They did report that a spread of excitation could occur, but this was limited to high signal levels. This is consistent with our results where animals with the largest hearing losses tended to have steeper slopes. Another possibility, however, is that the *N*-level functions at 16 kHz were actually nonlinear and compressive in a shape similar to 4 kHz. In the statistical fitting procedure for the normal animals, the linear fit may have been dominated by the compressive nature of the function, resulting in a shallow slope. For animals with high thresholds, the nonlinear compressive curve would shift to the right due to elevated thresholds. A shift to the right would decrease the influence of the compressive portion of the function and would highlight the linear (low-level) portion. The statistical fitting would now be influenced by the steep linear portion of the function and would result in an erroneous steep slope. To examine this possibility we examined the *N*-level functions at 8 kHz as a function of sensation level (SL), i.e., decibel above the threshold. Figure 10 illustrates the average (±1SD) *N*-level function from the control animals along with the data from the animals exposed to noise. For signal levels ranging from 0 to 20 dB SL, some of the animals show steeper growth functions than the normal-hearing animals, supporting the statistical analysis. Future studies with more signal levels, however, would help clarify the distinction between actual recruitment and a simple shift along the threshold axis.

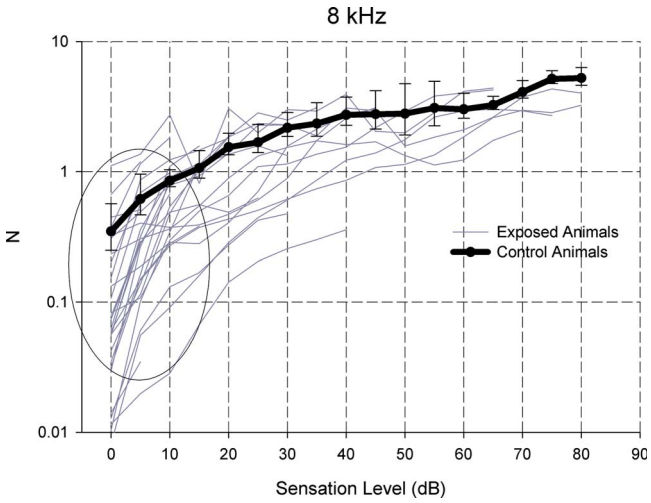


FIG. 10. (Color online) *N*-level values at 8 kHz plotted as a function of sensation level (i.e., dB above threshold). The solid thick line is the average *N* value (±1SD) computed across the control animals, and the gray lines are the data from each of the noise-exposed animals. The circle illustrates the region where animals show steeper growth functions than the control animals.

C. Parameter α

The delay parameter of $P(t)$, α, decreased as the signal level increased. For the high frequency stimuli (4, 8, and

16 kHz), α decreased linearly (log-log axis). To compare the change in α with signal level to previous studies, we fit a linear regression line to the nonlinear growth functions described by Eq. (4) and Table II. The slope of these functions were similar to those reported by Burkard *et al.* (1993) who reported a slope of −0.0117 ms/dB at 8 kHz, and our data indicates (from Table II) a slope of −0.0106 ms/dB.

Noise exposure altered the growth functions by shifting the intercepts of the functions by an amount proportional to the amount of hearing loss. It is interesting to note that the coefficient A_1 in Table II ($A_1=-0.010$) is approximately the negative of the coefficient B_0 ($B_0=0.011$). This suggests a relation, and perhaps a common mechanism, between the change in the latency of the CAP with signal level and the shift in the latency of the CAP with the degree of hearing loss (i.e., slope relating C_0 to PTA). Heil and Neubauer (2001) proposed that auditory nerve fibers are triggered in response to the temporal integral of the pressure envelope of an acoustic signal. The integration of pressure by auditory nerve fibers may account for the relation between the latency of the CAP with signal level and the amount of the shift in the α-level function with threshold elevation. If the envelope of a signal is given as $P_a(t)=P_s \sin^2(\pi t/2t_r)$, where P_s is the peak pressure of the stimulus, t is the time, and t_r is the rise time of the stimulus, then the threshold, θ, is defined as $\theta = \int_0^D P_a(t)dt$, where D is the delay of the CAP. Assuming that auditory nerve fibers respond at the peak of the stimulus when the signal level is at the threshold, then, as shown in the Appendix, the delay of the CAP $\approx c(P_{\text{th}}/P_s)^{1/3}$, where P_{th} is the sound pressure at the threshold, P_s is the stimulus sound pressure, and $c=0.85$. This equation shows that the delay of the CAP with signal level is a family of curves shifted by the pressure at the threshold. On a log-log scale the delay-level function would decrease with a slope of 1/3 for both normal-hearing and hearing-impaired animals, but the curves would be shifted to longer delays in the hearing-impaired animals. The shift would be related to the degree of hearing loss, and a plot of the shift as a function of pressure at threshold would yield a slope of 1/3. This suggests that temporal integration could account for the similarity between the slope of the α-level function (A_1) and the shift of the α-level function with PTA (B_0).

The data from our study can be used to obtain a function similar to that proposed by the temporal integration of the sound pressure model. As shown in the Appendix, $\alpha \approx a(P_{\text{th}}/P_s)^{1/5}$, where a is ~2.0. The slope of the α-level function is equal in magnitude to the shift of the α-level functions with PTA. However, the slope is 1/5 instead of 1/3. Thus, although temporal integration may account for the equal magnitude of the slope of the α-level function and of the C_0-PTA function from our data, it does not account for the actual slopes. It is possible that an additional component to the temporal integration model may be necessary to modify the exponent.

At 1 and 2 kHz, there was a significant quadratic component to the α-level function, and as PTA increased, the linear term of the quadratic function increased. This caused the concavity of the α-level function to decrease with increasing auditory threshold. A likely explanation is that the CAP to low frequencies includes a larger cochlear traveling wave delay component compared to the CAPs to high frequencies, and responses from animals with high thresholds have an initial delay due to hair cell damage in the basal end of the cochlea. That is, at high signal levels the latency of the 2 kHz CAP in normal animals is similar to that at 16 kHz, indicating a common basal cochlear place of origin. As signal level decreases, the 2 kHz response shifts to its "best frequency" region in the cochlea. In animals with high thresholds, there is damage to the basal part of the cochlea, causing an initial apical shift in α to a region of functioning hair cells and neurons. The subsequent shift in α with signal level to the 2 kHz place will therefore be less than that in normal-hearing animals.

The literature regarding the influence of hearing loss on the latency-level function of N1 is conflicting. In guinea pigs with kanamycin induced hearing loss, Wang and Dallos (1972) showed no effect of the hearing loss on the latency-intensity function of N1. Similarly, Berlin and Gondra (1976) reported normal N1 functions in humans with sensorineural hearing loss of 25–55 dB. In contrast, Yoshie and Ohashi (1969) reported a parallel shift in the latency-level functions in damaged ears, which is consistent both with our results and with Elberling and Salomon's (1976) data from subjects who had a large degree of high frequency hearing loss. It is possible that the contrasting literature is due to the type of hair cells damaged, and the extent of hearing loss. The animals from Wang and Dallos' (1972) study had missing outer hair cells (OHCs) and normal inner hair cells (IHCs). In Berlin and Gondra's (1976) patients, the hearing loss was believed to be due to noise exposure, and with the limited extent of hearing loss reported, it is likely that this influenced only OHCs. Remaining IHCs could be stimulated by stereocilia being driven by fluid coupling in the organ of Corti (Nowotny and Gummer, 2006). Thus the latency of the response would be normal. In contrast, if both OHCs and IHCs were damaged, as we believe in our study because of the long duration noise exposures, an additional elevation in threshold would occur. More time would be required to integrate neurotransmitters before auditory nerve fibers would respond and, along with a shift to remote cochlear regions, could result in a shift in the latency-level function.

D. Parameters β, F, and K

The influence of the hearing threshold on the growth of β with signal level was not accurately described by the power function in Eq.(4). This could reflect a limitation in the power function chosen to fit the growth function or it may indicate that changes in β with signal level are not related to the degree of hearing loss. Beta was larger at the high frequencies in animals with hearing loss as opposed to normal-hearing animals. The trend was apparent at 100 dBSPL but only reached statistical significance for signal levels at the threshold. This indicates that the width of $P(t)$ was larger in impaired animals than in normal animals and suggests a decrease in the synchrony of discharges from auditory nerve fibers contributing to the CAP. A decrease in synchrony could result from a change in the location of the

fibers contributing to the CAP. That is, given that the latency of auditory nerve fibers is $\approx 1/\mathrm{CF}$ (Kiang, 1965), where CF is a nerve fiber's characteristic frequency, the synchrony among apical fibers is different from that among basal fibers. If the fibers that are dominating the CAP in impaired ears are located more apically than normal ears, then less synchronous responses among auditory fibers could occur and lead to an increase in β.

Similar to β, the fit of Eq. (4) to the parameters of $U(t)$ was quite variable among animals. However, some effects of hearing loss on the $U(t)$ parameters occurred when evaluated at two different signal levels. Because $U(t)$ is a volume conducted extracellular triphasic waveform (Stegeman *et al.*, 1979) that is related to the intracellular potential of a nerve fiber (Clark and Plonsey, 1968), various explanations could account for the change in $U(t)$. Speculatively, these include a change in the distance between the source of the CAP and the recording site, alteration in the homogeneity of the conductive medium due to mixing of cochlear fluids after damage, and, perhaps, changes to ion channels in surviving nerve fibers causing an alteration in the action potential. More detailed physiologic experiments will be needed to understand the changes in $U(t)$ in pathologic ears.

E. Clinical implications

The research presented in this paper is part of an ongoing research to develop clinical diagnostic procedures that identify the locus of auditory damage and describe the pathophysiology underlying sensory-neural hearing loss in humans. As future stem cell and genetic treatments are developed, it will be important to have clinical techniques that identify the locus and extent of insult in order to appropriately target therapeutic intervention. The results of this study showed that the convolution model of the CAP fitted well the CAPs from both normal-hearing and noise-exposed animals, and a quantitative relation between the signal level and the hearing threshold was obtained. This was in agreement with the findings of Lichtenhan and Chertoff (2008), which showed that the convolution model could also be fitted to CAPs recorded from humans using an electrode placed on the eardrum both before and after noise exposure. This is encouraging as a first step in translating our animal findings to human application. Interestingly, some of the parameters changed similarly between the two studies. For example, β increased and N decreased at low signal levels after noise exposure in both studies. However, the parameter α decreased after noise exposure in the human subjects, whereas it increased in the animal subjects. This may be due to the different exposure frequencies used in the two studies, 2 kHz versus 8 kHz narrow-band noise, and differences in the mechanisms of temporary and permanent hearing loss. For a continued development of the parameters into a clinical tool, future research will be required to determine the relation between the changes in the parameters and the changes in the anatomy of the auditory nerve, as well as attempt to determine if the parameters can distinguish cochlear from neural damage.

V. CONCLUSIONS

The high correlations between the analytic CAP and the physiologic recorded CAPs indicate that the convolution model of the CAP provides an adequate description of CAP waveform morphology.

The growth functions of the CAP parameters N and α are well described by a polynomial equation with coefficients that depend on the hearing threshold. In some cases the variation of the coefficient with the hearing threshold indicated that noise exposure resulted in a simple lack of gain, whereas in other cases the slopes of the functions were altered.

The growth functions of the parameters β, F, and K were not adequately explained by a polynomial function with coefficients that varied with the hearing threshold. However, the parameters did change as a result of noise exposure and were suggestive of a decrease in synchrony in discharges among auditory nerve fibers and perhaps changes in either the source or the volume conduction properties of the medium from which the CAP is recorded.

ACKNOWLEDGMENTS

This research was funded by the National Institute on Deafness and other Communication Disorders at the National Institutes of Health (Grant No. RO1 DC02117). The authors thank Byron Gajewski for his assistance on the statistical analysis and David Lerner for his consultation with the mathematics.

APPENDIX

1. Temporal integration of sound pressure and CAP latency

The purpose of this Appendix is to illustrate the following two points. First is to show that if the latency of the CAP depends on the temporal integration of sound pressure (i.e., similar to auditory nerve fibers, Heil and Neubauer, 2001), then the latency-intensity function can be approximated by a power function. Second is to show that the slope of this function on a log-log scale is the same as the slope relating CAP latency to elevations in auditory threshold.

Given that the envelope of our signal is $P_a(t) = P_s \sin^2(\pi t/2t_r)$, where P_s is the peak pressure of the stimulus and t_r is the rise time of the stimulus, then the threshold is defined as

$$\theta = \int_0^D P_a(t)dt, \tag{A1}$$

where D is the delay from the stimulus onset. Substituting in the functional form of $P(t)$ from above and solving the integral yield

$$\theta = P_s\left[\frac{D}{2} - \frac{t_r}{2\pi}\sin\left(\frac{\pi D}{t_r}\right)\right], \tag{A2}$$

with units of Pa s. In our data, the threshold is defined in pascals. If we assume that at threshold the auditory nerve responds at the peak of the stimulus, then the delay of the

CAP from the stimulus onset due to temporal integration would equal the rise time (t_r). Therefore, if $D=t_r$, then

$$\theta = P_{\text{th}}\left[\frac{t_r}{2} - \frac{t_r}{2\pi}\sin(\pi)\right] \tag{A3}$$

or

$$\theta = \frac{P_{\text{th}}D}{2}, \tag{A4}$$

where P_{th} is the pressure at the threshold. The rise time of the signals in this study was 1 ms; therefore $\theta=0.5P_{\text{th}}$. Using this relation and rearranging Eq. (A2) allow us to examine the relation between the delay of the CAP, threshold, and stimulus pressure. That is,

$$\frac{P_{\text{th}}}{P_s} = 2\left[\frac{D}{2} - \frac{t_r}{2\pi}\sin\left(\frac{\pi D}{t_r}\right)\right]. \tag{A5}$$

To show that this equation provides a function that is approximately linear on a log-log scale, we let $k=t_r/\pi$, and rearranging we find

$$D - k\sin\left(\frac{D}{k}\right) = \frac{P_{\text{th}}}{P_s}. \tag{A6}$$

Expanding Eq. (A6) in a Taylor series in D/k and retaining the lowest order give

$$D \approx c\left(\frac{P_{\text{th}}}{P_s}\right)^{1/3}, \tag{A7}$$

where $c=0.85$. Equation (A7) shows that the CAP delay is a family of curves that on a log-log axis would have the same slopes of $-1/3$ but shifted along the delay axis. A plot of the shift of the functions with a change in threshold would yield a slope of $1/3$.

2. CAP latency

A similar function as B7 can be derived from the HLM analysis of α. From Eq. (4) and $\alpha=Z$,

$$\log_{10}\alpha(x,y) = C_0 + C_1 x, \tag{A8}$$

where

$$C_0 = A_0 + B_0 y,$$

$$C_1 = A_1.$$

x is the signal level in dBSPL and y is PTA in dBSPL. Thus,

$$\alpha = 10^{A_0 + B_0 y + A_1 x}. \tag{A9}$$

Now $\text{SPL}=20\log 10(P_s/P_0)$ and $\text{PTA}=20\log 10(P_{\text{th}}/P_0)$, where P_s, P_{th}, and P_0 are pressures in pascals for the stimulus, stimulus level at threshold, and reference pressure (0.000 02 Pa), respectively. Therefore,

$$\alpha = 10^{A_0}\left(\frac{P_{\text{th}}}{P_0}\right)^{B_0(20)}\left(\frac{P_s}{P_0}\right)^{A_1(20)}, \tag{A10}$$

and at 16 kHz, $A_0=0.206$, $B_0=0.011$, and $A_1=-0.010$ (Table II). Equation (A10) can be rewritten as

$$\alpha = a\left(\frac{P_{\text{th}}^{0.22}}{P_s^{0.20}}\right) \tag{A11}$$

or

$$\alpha \approx a\left(\frac{P_{\text{th}}}{P_s}\right)^{1/5}, \tag{A12}$$

where $a=1.9952$. Similar to Eq. (A7), on a log-log axis, α would show that the magnitude of the slope relating α to the signal level would equal the slope relating α to the threshold. However, the slopes obtained from the data are smaller than predicted by a temporal integration of the sound pressure model for CAP latency.

Berlin, C. I., and Gondra, M. I. (**1976**). "Extratympanic clinical electrocochleography with clicks," *Electrocochleography*, edited by R. Ruben, C. Elberling, and G. Salomon (University Park Press, Baltimore, MD), pp. 457–469.

Burkard, R., Voigt, H. F., and Smith, R. L. (**1993**). "A comparison of N1 of the whole nerve action potential and wave I of the brain-stem auditory evoked response in Mongolian gerbil," J. Acoust. Soc. Am. **93**, 2069–2076.

Chertoff, M. E. (**2004**). "Analytic treatment of the compound action potential: Estimating the summed post-stimulus time histogram and unit response," J. Acoust. Soc. Am. **116**, 3022–3030.

Chertoff, M. E., Yi, X., and Lichtenhan, J. T. (**2003**). "Influence of hearing sensitivity on mechano-electric transduction," J. Acoust. Soc. Am. **114**, 3251–3263.

Clark, J., and Plonsey, R. (**1968**). "The extracellular potential field of the single active nerve fiber in a volume conductor," Biophys. J. **8**, 842–864.

de Boer, E. (**1975**). "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**, 1030–1045.

Dolan, T. G., and Mills, J. H. (**1989**). "Recoveries of whole-nerve AP thresholds, amplitudes and tuning curves in gerbils following noise exposure," Hear. Res. **37**, 193–202.

Elberling, C. (**1976**). "Deconvolution of action potentials recorded form the ear canal in man," *Disorders of Auditory Function*, edited by S. D. G. Stevens (Academic, London), Vol. II, pp. 109–117.

Elberling, C., and Salomon, G. (**1976**). "Action potentials from pathological ears compared to potentials generated by a computer model," *Electrocochleography*, edited by R. Ruben, C. Elberling, and G. Salomon, pp. 439–455.

Goldstein, M. H., and Kiang, N. Y. (**1958**). "Synchrony of neural activity in electric responses evoked by transient acoustic stimuli," J. Acoust. Soc. Am. **30**, 107–114.

Hall, R. D. (**1990**). "Estimation of surviving spiral ganglion cells in the deaf rat using the electrically evoked auditory brainstem response," Hear. Res. **45**, 123–136.

Heil, P., and Neubauer, H. (**2001**). "Temporal integration of sound pressure determines thresholds of auditory-nerve fibers," J. Neurosci. **21**, 7404–7415.

Heinz, M. G., Issa, J. B., and Young, E. D. (**2005**). "Auditory-nerve rate responses are inconsistent with common hypothesis for the neural correlates of loudness recruitment," JARO **6**, 91–105.

Kiang, N. Y. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA), pp. 25–30.

Lichtenhan, J. T., and Chertoff, M. E. (**2008**). "Temporary hearing loss influences post-stimulus time histogram and single neuron action potential estimates from human compound action potentials," J. Acoust. Soc. Am. **123**, 2200–2212.

McMahon, C. M., and Patuzzi, R. B. (**2002**). "The origin of the 900 Hz spectral peak in spontaneous and sound-evoked round-window electrical activity," Hear. Res. **173**, 134–142.

Miller, A. M., Abbas, P. J., Robinson, B. K., Rubinstein, J. T., and Matsuoka, A. J. (**1999**). "Electrically evoked single-fiber action potentials from cat: responses to monopolar, monophasic stimulation," Hear. Res. **130**, 197–218.

Morest, D. K., Kim, J., Potashner, S. J., and Bohne, B. A. (**1998**). "Long-term degeneration in the cochlear nerve and cochlear nucleus of the adult chinchilla following overstimulation," Microsc. Res. Tech. **41**, 205–216.

Nowotny, M., and Gummer, A. W. (**2006**). "Nanomechanics of the subtec-

torial space caused by electromechanics of cochlear outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **103**, 2120–2125.
Popelar, J., Syka, J., and Berndt, H. (**1987**). "Effect of noise on auditory evoked responses in awake guinea pigs," Hear. Res. **26**, 239–247.
Portmann, M., Aran, J., and Lagourgue, P. (**1973**). "Testing for recruitment by electrocochleography," Ann. Otol. Rhinol. Laryngol. **82**, 36–43.
Raudenbush, S. W., and Bryk, A. S. (**2002**). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. (Sage, Newbury Park, CA).
Raudenbush, S. W., Bryk, A. S., and Congdon, R. (**2000**). HLM6, hierarchical linear and nonlinear modeling, Scientific Software International (SSI).
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Salvi, R. J., Henderson, D., Hamernik, R., and Ahroon, W. A. (**1983**). "Neural correlates of sensorineural hearing loss," Ear Hear. **4**, 115–129.
Schuknecht, H. F. (**1988**). "Retrograde cochlear neuronal degeneration in human subjects," Acta Oto-Laryngol., Suppl. **450**, 1–20.
Stegeman, D. F., de Weerd, J. P. C., and Eijkman, E. G. J. (**1979**). "A volume conductor study of compound action potentials of nerves in situ: The forward problem," Biol. Cybern. **33**, 97–111.
Versnel, H., Prijs, V. F., and Schoonhoven, R. (**1992b**). "Round-window recorded potential of single-fibre discharge (unit response) in normal and noise-damaged cochleas," Hear. Res. **59**, 157–170.
Versnel, H., Schoonhoven, R., and Prijs, V. F. (**1992a**). "Single-fibre and whole-nerve responses to clicks as a function of sound intensity in the guinea pig," Hear. Res. **59**, 138–156.
Wang, B. (**1979**). "The relation between the compound action potential and unit discharges of the auditory nerve," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.
Wang, B., and Dallos, P. (**1972**). "Latency of whole-nerve action potentials: Influence of hair-cell normalcy," J. Acoust. Soc. Am. **52**, 1677–1686.
Yoshie, N., and Ohashi, T. (**1969**). "Clinical use of cochlear nerve action potential responses in man for differential diagnosis of hearing losses," Acta Oto-Laryngol., Suppl. **252**, 71–87.

A cascade autocorrelation model of pitch perception

Emili Balaguer-Ballester[a] and Susan L. Denham[b]
Centre for Theoretical and Computational Neuroscience, University of Plymouth, Devon PL4 8AA, United Kingdom

Ray Meddis[c]
Department of Psychology, Essex University, Colchester CO4 3SQ, United Kingdom

(Received 24 September 2007; revised 23 June 2008; accepted 25 June 2008)

Autocorrelation algorithms, in combination with computational models of the auditory periphery, have been successfully used to predict the pitch of a wide range of complex stimuli. However, new stimuli are frequently offered as counterexamples to the viability of this approach. This study addresses the issue of whether in the light of these challenges the predictive power of autocorrelation can be preserved by changes to the peripheral model and the computational algorithm. An existing model is extended by the addition of a low-pass filter of the summary integration of the individual within-channel autocorrelations. Other recent developments are also incorporated, including nonlinear processing on the basilar membrane and the use of integration time constants that are proportional to the autocorrelation lags. The modified and extended model predicts with reasonable success the pitches of a range of stimuli that have proved problematic for earlier implementations of the autocorrelation principle. The evaluation stimuli include short tone sequences, click trains consisting of alternating interclick intervals, click trains consisting of mixtures of regular and irregular intervals, shuffled click trains, and transposed tones.
 [DOI: 10.1121/1.2967829]

PACS number(s): 43.66.Ba, 43.66.Hg [BCM] Pages: 2186–2195

I. INTRODUCTION

Autocorrelation algorithms have given a useful account of a wide range of auditory pitch phenomena (e.g., Licklider, 1951; Slaney and Lyon, 1990; Meddis and Hewitt, 1991; Cariani and Delgutte, 1996a, 1996b). Their success has generated a large number of studies designed to test the limits of these theories. Some studies have identified weaknesses in the detail of early formulations but have also suggested simple remedies that leave the basic periodicity principles intact (e.g., Wiegrebe, 2001; Pressnitzer *et al.*, 2001; Bernstein and Oxenham, 2005) while other studies, to be discussed below, remain as challenges to the underlying idea of the use of autocorrelation. This report will argue that autocorrelation can continue to provide a satisfactory account if an existing model (Meddis and O'Mard, 1997) is changed to incorporate a more sophisticated model of the auditory periphery and is expanded by adding a stage with a longer integration time constant than used hitherto. This model combines Licklider's (1951) original idea of an autocorrelation function (ACF) of auditory nerve (AN) activity with the idea of generating a summary autocorrelation function (SACF) based on an aggregation of the individual ACFs across all fibers.

Licklider (1951) originally suggested using a running autocorrelation with a narrow temporal window shifted along the time axis. This window is represented by a low-pass filter having a time constant of only 2 or 3 ms. Most explorations have used time constants in this region. However, Wiegrebe (2001) provided evidence that multiple time constants are required, each proportional to the autocorrelation lag being evaluated. A value of twice the lag appeared to give a satisfactory account of the audibility of the oscillations in pitch strength in response to repeated-interval noises. According to this rule, time constants between 40 and 4 ms would be appropriate for pitches in the range 50–500 Hz. This suggestion will be adopted along with his other suggestion that some stimuli require even longer integration times. The proposed solution is to extend the existing model by passing the running SACF through a low-pass filter to produce a new function, which will be given the acronym LP-SACF.

The necessity for longer time scales has been evident since the publication of a demonstration by Hall and Peters (1981) that harmonically related pure tones can generate a pitch percept even when presented nonsimultaneously. They presented three harmonically related pure tones in rapid succession and obtained a pitch sensation consistent with that obtained when the three were presented simultaneously, but only when noise was present. The offset of the first and the onset of the third tone were separated by 60 ms and yet pitch integration still occurred. This demonstration will be used in the first evaluation of the extended model.

A new class of pitch stimuli has recently emerged, which appears to offer a more fundamental challenge to autocorrelation as a general explanation. These stimuli are high-pass filtered irregular click trains in which the distortion products have been masked with background noise; nevertheless, they can be ordered on a pitch-height scale (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001). The responses of listeners to these stimuli cannot be predicted by

[a]Electronic mail: emili.balaguer-ballester@plymouth.ac.uk
[b]Electronic mail: s.denham@plymouth.ac.uk
[c]Electronic mail: rmeddis@essex.ac.uk

any current autocorrelation algorithms and this has resulted in calls to search beyond autocorrelation for alternative theories. However, it will be shown that a radical change of direction is not required and that the extended model can be used to predict these data.

The use of short time constants to compute the running ACFs has led to an acknowledged weakness when the stimulus contains fluctuations whose durations are longer than the time constant used in the calculations. The work-around used in earlier studies to deal with this problem involved inspecting the SACF at the end of a complete pitch period (e.g., Meddis and Hewitt, 1991). While this worked with periodic stimuli, it is clearly only a short-term expedient and not an elegant long-term solution to the problem. The challenge of aperiodic click stimuli makes the problem more pressing because it is not possible to specify an ideal time at which to sample the SACF. The new extended model removes the problem by using a longer integration window, thus stabilizing the output function.

II. THE MODEL

The original model has been fully described and extensively evaluated in previous publications (Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). It consists of three stages: (1) an auditory model to simulate AN spiking probabilities, (2) an autocorrelation algorithm, and (3) an algorithm for making predictions of the pitches heard by listeners and their pitch discrimination abilities.

A. Auditory model

An auditory model simulates the generation of spikes in AN fibers by simulating the processing stages between stimulus reception and the inner hair cell (IHC) AN synapse. Successive processing stages are (1) stimulus input, (2) stapes response, (3) multichannel basilar membrane (BM) response, (4) IHC receptor potential, (5) IHC/AN synapse transmitter vesicle release rates, and (6) AN spiking probabilities. Computational models of the auditory periphery continue to evolve but the version used here has been fully described in a recent study of AN first spike latencies (Meddis, 2006). All formulas and parameters are given in the Appendix to that report; and used but unchanged here; except the transmitter release permeability in the IHC presynapse, which takes the value indicated in the work of Sumner *et al.* (2002). The input to the model is the stimulus used in the corresponding psychophysical experiment. The output is a stream of AN spiking probabilities. The implementation used here has 60 channels with best frequencies (BFs) ranging from 100 to 10 000 Hz along a logarithmic scale. The auditory model was evaluated at an integration period of 1/44 100 s, except for Evaluation 4, which used a period one-quarter of that duration.

Figure 1(a) shows the response of the auditory model to a 100 ms harmonic complex consisting of the third to sixth harmonics of 100 Hz. The lower panel in Fig. 1(a) shows the AN spiking probabilities, $p(t,k)$, arranged as a channel x time matrix. This is the input to the next stage, the autocorrelation algorithm.

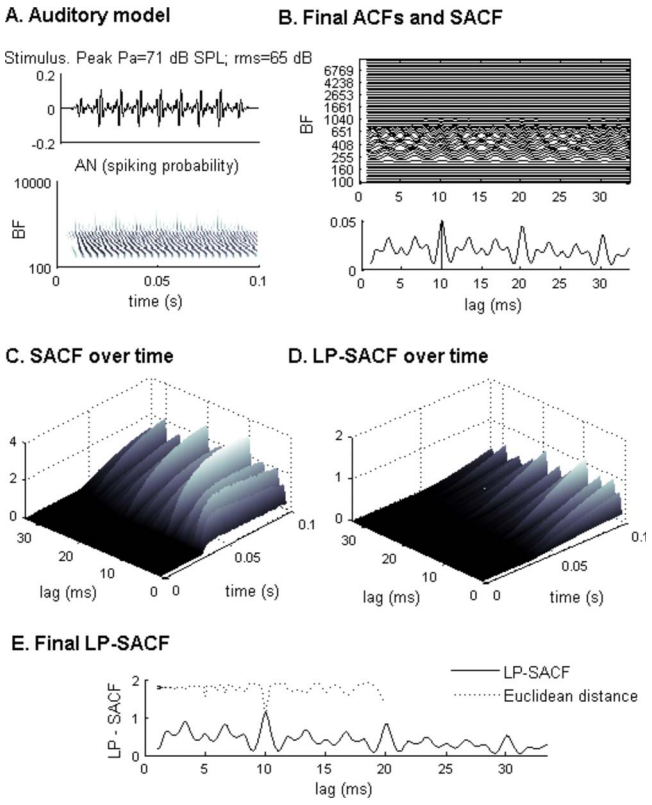


FIG. 1. (Color online) Response of the model to a 100 ms complex stimulus consisting of harmonics 3, 4, 5, and 6 of a 100 Hz fundamental. (A) Stimulus waveform (upper panel) and AN spiking probabilities generated by an auditory model (lower panel). (B) Multichannel ACFs (upper panel) and SACF (lower panel) after 100 ms. (C) Evolution of the SACF over time. (D) Evolution of the LP-SACF over time. (E) LP-SACF at the end of the stimulus (continuous line). The Euclidean-distance function is shown as a dotted line. The minimum of this function (10 ms) predicts the pitch.

B. Autocorrelation algorithm

The algorithm for computing the SACF function is implemented as described in the work of Meddis and O'Mard (1997). The computation of the individual running autocorrelations in each channel, $h(t,l,k)$, is based on the spike probabilities, $p(t,k)$,

$$h(t,l,k) = p(t,k) \cdot p(t-l,k) \cdot \frac{\Delta t}{\tau(l)} + h(t-\Delta t,l,k) \cdot e^{-\Delta t/\tau(l)}, \tag{1}$$

where t is time, l is the autocorrelation lag, Δt is the sampling interval, and k is the channel number. $\tau(l)$ is a time constant specific to a given lag and is set to $2l$ (Wiegrebe, 2001). This equation is the same as Eq. (1) in the work of Meddis and O'Mard (1997) but expressed as a recursive function. 191 lags were used in this study. They were linearly spaced between 1/30 and 1/1000 s and were the same for all BFs.

The ACFs in Eq. (1) are summed across channels to create the SACF

$$S(t,l) = \sum_{k=1}^{N} h(t,l,k). \tag{2}$$

Figure 1(b) shows the running ACFs and the running SACF as they appear at the end of the stimulus presentation while Fig. 1(c) shows how the SACF changes over time.

At this point, a new stage is added. The running SACF is passed through a low-pass filter implemented as an exponentially decaying average,

$$P(t,l) = S(t,l) + P(t - \Delta t, l) \cdot e^{-\Delta t/\lambda}, \quad (3)$$

where λ is the time constant of the filter. $P(t,l)$ is the LP-SACF. Figure 1(d) shows how the LP-SACF changes over time. λ was set to 120 ms throughout the study except for evaluation 4 which was an order of magnitude larger. A detailed study of the optimum value was not attempted. This value was chosen simply because it was the minimum time constant large enough to yield acceptable results for the data of Hall and Peters (1981) described in the first evaluation below. It is consistent with the minimum estimate of integration period of 210 ms proposed by Grose *et al.* (2002) given that an exponential decay function with a time constant of 120 ms will have decayed to 17% of its starting value after 210 ms.

Since the original model was published, a number of detailed improvements have been suggested by different authors. We have adopted Wiegrebe's suggestion that the time constant of integration of the SACF should be linked to the individual lag. On the other hand, Pressnitzer *et al.* (2001) have suggested a weighting function that reduces the magnitude of the SACF as a function of lag and Bernstein and Oxenham (2005) and Denham (2005) have suggested schemes for omitting some lags from the individual channel ACFs. Notwithstanding the agreed merits of these two suggestions, they have not been adopted here in order to simplify the discussion of the contribution of the other changes introduced in this study and because they do not alter the conclusions to be drawn in this report.

C. Pitch predictions

The simplest method for predicting pitch uses the reciprocal of the lag associated with the highest peak in the SACF, and the same argument applies to the new LP-SACF. This often leads to unambiguous and accurate predictions. However, there are some situations when it cannot be used; these include small pitch shifts caused by mistuned harmonics and predictions of the ability to discriminate the pitch of two stimuli (see Meddis and Hewitt, 1991 and Meddis and O'Mard, 1997 for details). On these occasions, it is better to follow the method used in the experimental procedure. This typically involves searching for a best pitch match by adjusting the pitch of a second tone until it matches with the pitch of a reference tone. This is the approach adopted here.

To implement the matching procedure, the LP-SACF is generated for a range of periodic comparison stimuli. These are then compared, one by one, with the LP-SACF of the test stimulus used in the study to be simulated. This is achieved by computing the Euclidean distance between the LP-SACF of the comparison tone and the LP-SACF of the test stimulus. The fundamental frequency of the comparison stimulus with the smallest associated Euclidean distance is then chosen as the predicted "best-match" pitch [Figure 1(e)].[1]

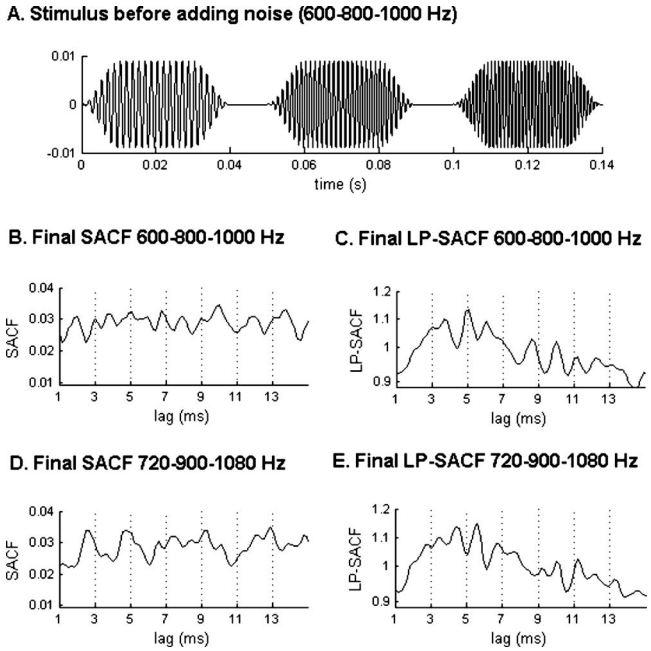


FIG. 2. Predicted virtual pitch evoked by successive tones. (A) The stimulus waveform for a sequence of 40 ms pure tones of 600, 800, and 1000 Hz separated by a silent period of 10 ms (Hall and Peters, 1981), before the addition of noise. (B) The average final SACF in response to five presentations of the 600–800–1000 Hz tone sequence in the presence of noise. (C) The average final LP-SACF in response to five presentations of the 600–800–1000 Hz tone sequence in the presence of noise. (D) The average final SACF in response to five presentations of a 720–900–1080 Hz tone sequence in the presence of noise. (E) The average final LP-SACF in response to five presentations of a 720–900–1080 Hz tone sequence in the presence of noise.

III. EVALUATIONS

A. Evaluation 1

The need for a window of pitch integration longer than 3 ms is clear from a demonstration by Hall and Peters (1981). They showed that nonsimultaneous tones presented in background noise can combine to create a virtual pitch that can be matched to other pitch-evoking stimuli. Other more recent studies using nonsimultaneous components such as temporal fringes (Carlyon, 1996; Micheyl and Carlyon, 1998), mistuned delayed harmonics (Ciocca and Darwin, 1999; Gockel *et al.*, 2005), and stimuli with interpolated silences (Plack and White, 2000) have provided further evidence that substantial temporal integration is required to explain how pitch perception aggregates stimulus information across silent intervals.

Hall and Peters' (1981) stimulus is illustrated in Fig. 2(a). It consists of three tones played successively against a white noise background. Each 50 dB sound pressure level (SPL) tone lasts 40 ms and is separated from the following tone by a gap of 10 ms. In this particular example, the frequencies used are 600, 800, and 1000 Hz. The noise was white noise at a level of 65 dB SPL rms. The authors specify that the tones were 6 dB *above* masked threshold. The individual tones of the sequence were still audible, but listeners

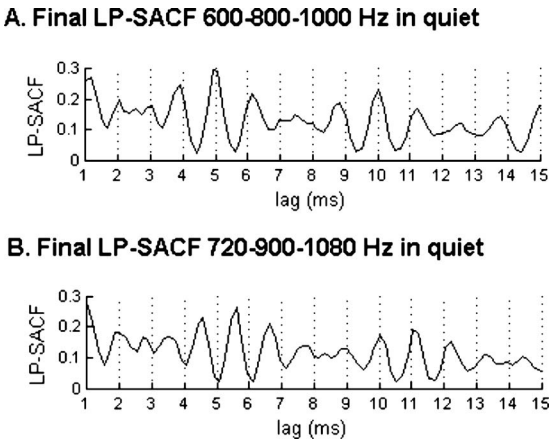


FIG. 3. LP-SACF response to short tone sequences in quiet. (A) Final LP-SACF for the 600–800–1000 Hz tone sequence in quiet. (B) Final LP-SACF for the 720–900–1080 Hz tone sequence in quiet.

were instructed to attend to the lowest of the perceived pitches. It was found that listeners matched the lowest perceived pitch to a 200 Hz pure tone. To further illustrate the effect, Hall and Peters (1981) compared this stimulus with another version where the component frequencies were 720, 900, and 1080 Hz. This stimulus was matched by their subjects to a pure tone with a lower pitch (180 Hz) even though the component frequencies were higher in the second stimulus.

The waveform in Fig. 2(a) shows the stimulus before the noise was added. The SACF for the stimuli in the presence of the noise [Figs. 2(b) and 2(d)] is unable to predict the low virtual pitch. Figures 2(c) and 2(e) show the LP-SACFs for the two stimuli in the presence of the noise. These are both based on an average of five trials to reduce variability created by the noise background. Due to the loudness of the added noise, we based the predictions in this evaluation on the highest LP-SACF peak rather than on the Euclidean metrics. The LP-SACF for the 600–800–100 Hz stimulus shows a broad peak around a lag of 5 ms corresponding to a pitch in the region of 200 Hz. An examination of the fine structure of the LP-SACFs reveals a shift to the right (longer lags) in the case of the 720–900–1080 Hz stimulus, indicating a lower predicted pitch around 5.6 ms (180 Hz).

The response of the computer model shows that low-pass-filtering the SACF to form the LP-SACF was successful in integrating periodicity information across the whole stimulus. This was not achieved with the SACF alone because of its short time constants. For example, in the model the time constant associated with the 5 ms lag (corresponding to a periodicity of 200 Hz) is only 10 ms, which is considerably shorter than the time interval between the first and the last tones.

Hall and Peters (1981) found that the low virtual pitch was not heard when the tones were presented in quiet. It is therefore of some interest to examine the model response to the sequence of tones without the background noise. Figure 3 shows the LP-SACF output for the two stimuli presented in quiet. Major peaks are present at lags of 5 and 5.6 ms, respectively. This would appear to predict corresponding pitches of 200 and 180 Hz, pitches that were not reported by subjects in the experiment. However, in quiet, the LP-SACF contains a series of peaks, potentially indicating a much higher pitch. The listeners in Hall and Peters' (1981) experiment never matched the stimulus to a single pure tone with low pitch but always preferred to match these stimuli to a higher pitch in the region of individual component tones. This suggests caution against the common tendency to identify a single large peak in an ACF with the predicted pitch and indicates the need to take the whole of the function into account.

In pitch experiments it is not unusual to find that subjects listen analytically and base their judgements on individual tone components rather than the total stimulus complex. Normally, subjects are discouraged from matching the pitch of one of the stimulus components. A comparison between the model responses in Figs. 2 and 3 suggests that noise has the effect of reducing the relative size of the more rapid oscillations in the LP-SACF corresponding to the component tones and makes it more likely that a judgement of the lowest perceived pitch will be based on the peaks in the range of the virtual pitch. Hall and Peters (1981) also suggested that the transient amplitude changes (audible when the tones are presented in a quiet background) may trigger some other mechanism that emphasizes the separate nature of the tones and tips the perceptual balance in favor of analytic hearing. Unfortunately, the computer model presented here does not contain any mechanism at this level of sophistication.

In conclusion, the introduction of a low-pass filter to further process the SACF into the LP-SACF was successful in integrating tone sequences presented successively in noise to generate a pitch prediction that agreed with the lowest perceived pitch reported by listeners. However, the demonstration indicated an unsolved problem concerning how the balance of synthetic and analytic listening is established in general for sounds presented in quiet.

B. Evaluation 2

When a regular click train is played, subjects hear a pitch whose frequency is the reciprocal of the time interval between successive clicks. Kaernbach and Demany (1998) modified this stimulus by placing a single additional click at a random time between each regular click. The stimuli were high-pass filtered (in this case at 6 kHz) to remove all resolved harmonics. The modified stimulus with the interpolated random clicks has only a weak pitch compared to the regular interval train. Indeed, their subjects found it difficult to discriminate between it and a completely random click train with the same overall click rate. The interpolation of random clicks had greatly reduced the sensation of pitch even though regular intervals were present between every second click.

This result is not expected if a prediction is based on an autocorrelation of the click train itself. A simple autocorrelation of the stimulus waveform shows a clear peak at the regular interval for both the simple click train and the one

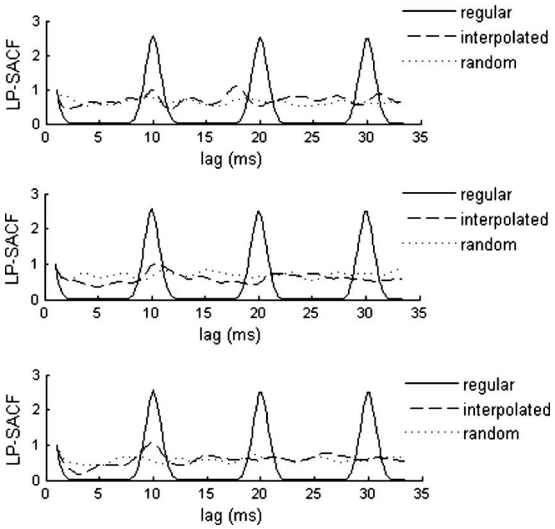


FIG. 4. The effect of the LP-SACF on adding randomly interpolated clicks (dashed line) to a regular click train (solid line). The results can be compared with those for a completely random click train (dotted line). The click trains have been made as explained in the text. The LP-SACFs are shown for three different stimulus samples.

with the interpolated clicks. This approach to autocorrelation would lead to the prediction that both stimuli should have a similarly clear pitch.

However, different predictions are made when the autocorrelation is based on the output of the auditory model. To demonstrate this, the model was evaluated using three different stimuli: regular, interpolated, and random. The click rate in the regular click train was 100 clicks/s. The random click train had a mean rate of 200 clicks/s and was constrained so that any run of three consecutive clicks with above average intervals was deleted and replaced. The interpolated click train contained repeated *ABX* sequences containing a fixed interval ($A+B=10$ ms) in which an interpolated click was randomly placed between each pair of regular clicks. The *AB* pair was then followed by another random interval (*X*) with an average duration of 10 ms. The overall click rate of this stimulus was 150 clicks/s. All clicks were bandpass filtered between 6 and 10 kHz using a cascade of six first-order Butterworth filters. They were presented to the model at a spectrum level of 55 dB SPL in a low-pass filtered white noise background at a spectrum level of 30 dB SPL. The duration of all stimuli was 500 ms.

The responses of the model to these stimuli are shown in Fig. 4. Because these stochastic stimuli change from trial to trial, three sets of results are given. The click train with random intervals produces a LP-SACF (dotted line) with no reliable structure. The regular click train, on the other hand, gives a regular multipeak LP-SACF (continuous line) with the first major peak at 10 ms, as expected. When the interpolated clicks are added, the LP-SACF (dashed line) shows a reliable but small peak at 10 ms but regularly spaced peaks at 20 and 30 ms are typically not present. On visual inspection, the LP-SACF for this interpolated click stimulus looks much more like the LP-SACF for the random click stimulus than that for the regular click train. In summary, the model predicts an unambiguous pitch for the regular click train but a relatively weak pitch for the interpolated click train. It also suggests that the interpolated click train will be discriminated from the random click train only with difficulty. This prediction is consistent with the reported observation that the interpolated sequence does not sound regular and is perceived by listeners as similar to a random click train. Nevertheless, the interpolated click train does produce a reliable small peak at 10 ms and this is consistent with the observation that, in the long run, listeners are better than chance at identifying it as more tonal than the completely random stimulus (Kaernbach and Bering, 2001).

In summary, the use of an auditory model as the input to the autocorrelation algorithm gives a good account of the main finding that the addition of randomly interpolated clicks degrades the tonality of a regular click train. This contrasts with the predictions of Kaernbach and Demany (1998) based on the autocorrelation of the stimulus waveform where little degradation is expected. Clearly, autocorrelation of the *stimulus waveform* is not a viable predictor of pitch perception in all cases but its shortcomings do not apply when the output of a model of the auditory periphery is used as the input to the autocorrelation calculations.

C. Evaluation 3

The success of Evaluation 2 raises the question of why the auditory model succeeds when an autocorrelation of the stimulus waveform fails. Pressnitzer *et al.* (2002, 2004) drew attention to the important changes that can take place in the representation of the stimulus as a consequence of nonlinearities inherent in auditory peripheral processing. Specifically, they suggested that the half-wave rectification that takes place in the electrical response of the IHC cell can introduce previously absent spectral components if the stimulus frequency components are not resolved by the peripheral auditory system. To demonstrate this, they applied autocorrelation at different stages in the peripheral processing and showed that changes took place following the half-wave rectification that were critical for distinguishing the pitch characteristics of two carefully chosen click-train stimuli (described below). These two click trains had the same average click rate but different pitches. They were able to show that autocorrelation of the output following half-wave rectification resulted in profiles that could be related to the pitches heard by listeners.

The peripheral model used in the work of Pressnitzer *et al.* (2002) was highly schematic, used long-term autocorrelation, and employed only linear peripheral filters. The model evaluated here uses nonlinear peripheral filtering to represent the response of the BM and has a considerably more complex representation of the generation of the IHC response. It also features short-term autocorrelation computations followed by a low-pass filtering stage and this contrasts with their approach using long-term autocorrelation. It will be useful to demonstrate that the findings of Pressnitzer *et al.* (2002) remain valid with the new model, because their insight is fundamental to understanding the success of the autocorrelation model in explaining the perceived pitch of many high-pass click-train stimuli.

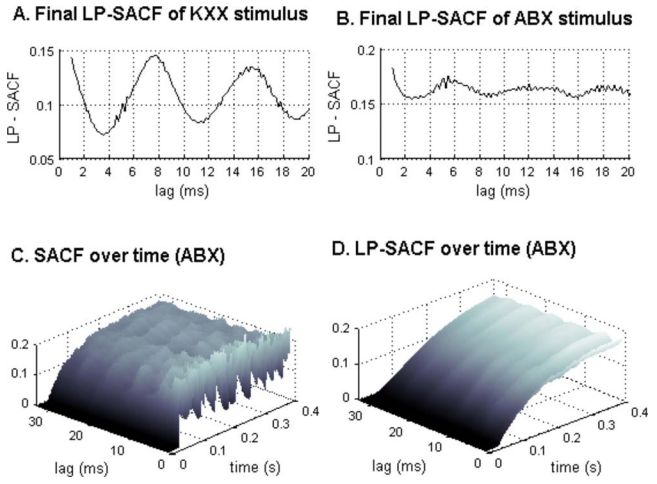


FIG. 5. (Color online) Model responses for the *KXX* and *ABX* click trains, with $K=5$ ms (see text). (A) Final LP-SACF for the *KXX* stimulus. (B) Final LP-SACF for the *ABX* stimulus. (C) Evolution of the SACF across time for the *ABX* stimulus. (D) Evolution of the LP-SACF across time for the same *ABX* stimulus.

The two click trains used by Pressnitzer *et al.* (2002, 2004) had mixtures of regular and random interclick intervals. The first stimulus (*KXX*) contained a single interval of fixed duration (K) followed by two intervals of random duration (mean duration, $K/2$). The second stimulus (*ABX*) consisted of trains of three random intervals with the constraint that the duration of the first pair of intervals summed to K and the third interval had a mean duration of K. Both click trains have the same average click rate of $3/(2K)$ clicks/s. Also, both click trains contain an interval of duration K. For the *KXX* stimulus, this interval occurs between the first and second clicks while for the *ABX* stimulus, the interval occurs between the first and the third clicks. Despite having identical mean click rates, the stimuli are reliably heard to have different pitches; *ABX* has a higher pitch than *KXX*.

These stimuli, *KXX* and *ABX* ($K=5$ ms), were presented to the model at a spectrum level of 60 dB SPL for a duration of 400 ms with onset and offset ramps of 5 ms. Click trains were high-pass filtered at 3 kHz. The response of the model is shown in Fig. 5 where it can be seen that the LP-SACFs for the two stimuli are clearly different. The LP-SACF of the *KXX* stimulus has a major peak at around 7.5 ms repeating at multiples of 7.5 ms indicating a predicted pitch of 133 Hz. The same prediction was made using the Euclidean-distance metric (not shown). In contrast, the *ABX* stimulus has its repeating peak starting at around 5.5 ms (predicted pitch 182 Hz). These results are consistent with the psychophysical observation of a lower pitch for the *KXX* stimuli. This result replicates the findings of Pressnitzer *et al.* (2004) and confirms that the operation of the present model is consistent with their principles.

The lower panels in Fig. 5 show the evolution of the SACF and the LP-SACF during one of the two stimuli (*ABX*). The development of the major autocorrelation peaks can be seen in the LP-SACF. However, the peaks are only intermittently represented in the SACF. The LP-SACF by virtue of its longer time window is able to average these peaks and produce a more stable representation. Hence, a long-term autocorrelation as used by Pressnitzer *et al.* (2002) is not necessary.

D. Evaluation 4

Carlyon *et al.* (2002) studied the pitch of a click train with alternating intervals of 4 and 6 ms that was bandpass filtered between 3.9 and 5.4 kHz to remove any resolved frequency components. They found that this stimulus generated pitch matches in the region of 4.5–7 ms when presented at a level of 54 dB SPL against a background of pink noise. The geometric mean of all matches was 5.7 ms. The regular intervals between the clicks of 4 and 6 ms might have been expected to give rise to pitch matches corresponding to either or both of these values. The second-order interval of 10 ms might also be a candidate. However, none of these were regularly reported by listeners. The authors performed a SACF analysis based on the method used by Meddis and O'Mard (1997). Their computations indicated likely pitch matches at 4, 6, and 10 ms. The authors concluded that autocorrelation analysis was unable to explain the results.

Carlyon *et al.* (2008) demonstrated that the combined AN responses to the 6 ms interval, measured as compound action potentials (CAPs), were stronger than for the 4 ms interval. Therefore, they suggested that a population of more central neurons, which respond only when their inputs exceed a fixed threshold value, would respond preferentially to the 6 ms intervals, resulting in a total average response closer to this interval, which would explain the listeners' preference for matching a pitch very close to the longer first-order interclick interval.

This raises the question of whether the AN spiking probabilities generated by the auditory peripheral model could reflect in some indirect way the above findings. If the responses were slightly stronger for the 6 ms intervals than for the 4 ms intervals, then a long-term periodicity analysis might also illustrate the dominance of the 6 ms period (on average within frequency channels). We repeated the autocorrelation analysis using the current model and obtained a more favorable outcome. Particular care was taken to reproduce the experimental stimulus exactly by filtering the clicks appropriately and adding background pink noise. Care was also taken to use the same pitch-match comparison stimuli as used in the original psychophysical experiment. The experiment was simulated using click trains with alternating intervals of 6 and 4 ms. The clicks were presented at an overall level of 78 dB SPL for 400 ms. They were bandpass filtered with cut-off frequencies of 3900 and 5300 Hz, generated using an eight-order Butterworth filter yielding an attenuation of 24 dB at half an octave above and below the cut-off frequencies. The background pink noise had the characteristics indicated in the psychophysical experiment. The stimulus was gated on and off with 50 ms raised-cosine ramps.

Comparison stimuli were generated according to the author's description: 29 isochronous pulse trains, with periods ranging from 2 to 14 ms in steps of 7%, with the period rounded to the nearest 0.1 ms. These were filtered in the same way as the test stimuli and LP-SACFs were generated

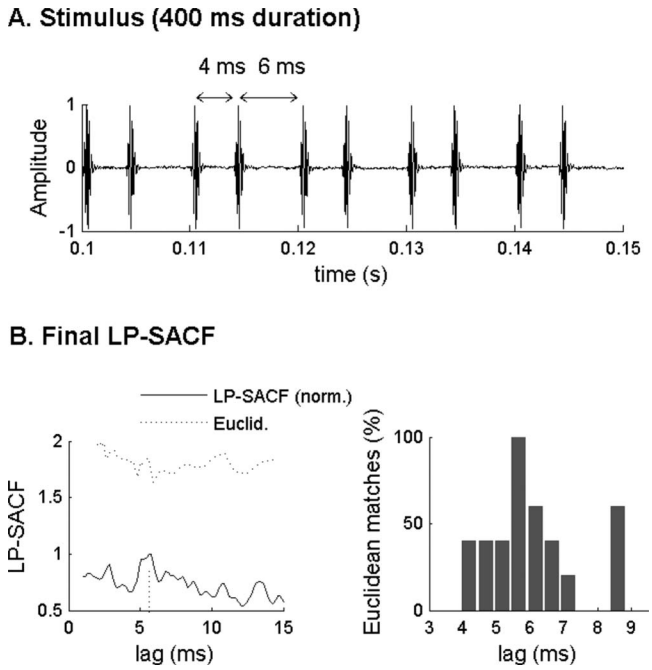


FIG. 6. Model response to the 6–4 ms alternating click train (Carlyon *et al.*, 2002) with bandpass filtering and pink noise included (see text). (A) Stimulus waveform (amplified for better visualization). (B) The left plot shows the final LP-SACF normalized response (solid line) and the corresponding Euclidean distance (dotted line). The maximum peak occurs at 5.76 ms and the minimum of the Euclidean distance occurs at 5.91 ms. The right plot shows a histogram of the Euclidean-distance matching for 20 different realizations of the background noise.

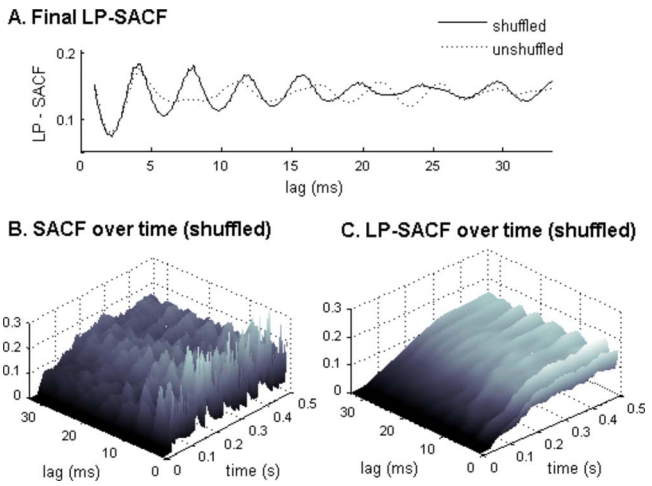


FIG. 7. (Color online) Response of the model to the *KX* click trains (*K* =4 ms). (A) Final LP-SACF for the shuffled (solid line) and unshuffled (dotted line) click trains. (B) Evolution of the SACF over time for the shuffled click train. (C) Evolution of the LP-SACF for the same shuffled stimulus.

to act as comparison templates. The best-match comparison pulse train was chosen on the basis of the smallest Euclidean distance between the test stimulus LP-SACF and that of the comparison stimulus.

The LP-SACF derived from the model response is shown in Fig. 6(b). It has a maximum peak at 5.76 ms, in agreement with the perceptual data (Carlyon *et al.*, 2002, 2008). The dotted line above the LP-SACF shows the Euclidean distances between the range of templates and the test stimulus. For the Euclidean-distance measure, the *x*-axis represents the interclick interval so that the best pitch match (smallest Euclidean distance) is at 5.91 ms. This contradicts our intuition that matches would be most likely to occur at both 4 and 6 ms, the intervals between the successive clicks. In this respect, the result agrees with the main findings of the original experiment. The distribution of matches for several realizations of the background noise [right plot in Fig. 6(b)] is not clearly unimodal, in contrast with the results obtained Carlyon *et al.* (2002, 2008). Nevertheless, the mean value of the predictions is robust for large numbers of stimulus realizations. This result further supports the claim that the *current* model is qualitatively consistent with the experimental results in that it systematically predicts pitch matches closer to the longest first-order interclick interval. As a result we argue that an autocorrelation account of pitch perception is not contradicted in a fundamental way, by recent results (Carlyon *et al.*, 2008). However, more research is needed to accurately model CAP responses of the AN to this stimulus (Carlyon *et al.*, 2008).

E. Evaluation 5

Yost *et al.* (2005) experimented with *KX* click trains where a regular interval (duration *K* ms) was alternated with a random interval (*X*). The duration of the random interval was uniformly distributed between 0 and 2*K* ms. This stimulus was then used to generate a second stimulus by randomly reordering all the interclick intervals. The "shuffled" click train contained exactly the same interclick intervals as the first "unshuffled" click train; only the sequence of the intervals was different. Surprisingly, the randomly shuffled click trains were typically heard to have a *greater* pitch strength than the unshuffled click trains, even though they were less regular as a result of the shuffling. When Yost *et al.* (2005) computed the SACF using an earlier version of the model (Meddis and Hewitt, 1991), there was little to indicate that the shuffled click train would be judged to have a greater pitch strength. The same was true of the autocorrelation of the stimulus waveform. They concluded that current autocorrelation models based on the long-term ACFs *cannot* account for the data of this study. Here, it will be shown that their results are indeed consistent with an autocorrelation analysis if an appropriate low-pass filtering is applied to the SACF.

The authors did note that the shuffled click trains contain longer runs of consecutive regular intervals than the unshuffled click trains. This is because an unshuffled click train, by definition, can never have two consecutive *K* intervals. In their view, the longer consecutive runs of the fixed interval are the key to understanding the phenomenon. This suggestion led us to expect that the new ACF with the low-pass filtering would reflect this long-term property of the stimulus

Figure 7 shows the LP-SACF of two (*KX*) click trains (*K*=4 ms), one unshuffled and the other shuffled. The random nature of these click trains means that these patterns will change from stimulus to stimulus but the examples given are typical. Unshuffled *KX* stimuli always have a single strong peak at *K* ms. The absence of secondary strong

peaks at multiples of K ms reflects the fact that such intervals occur only rarely in the stimulus. On the other hand, the LP-SACFs of the shuffled click trains have a number of peaks at multiples of K ms. The regularly spaced multiple peaks are caused by runs of consecutive regular intervals that occur by chance. Figure 7(b) compares the SACF with the LP-SACF over time. The obvious repeating peaks are not easily visualized in the SACF but are clearly present in the LP-SACF

The height of the first peak is approximately the same for both LP-SACF functions. This is not surprising as both shuffled and unshuffled stimuli have the same number of fixed-duration intervals. The height of the first peak of the SACF has often been taken to predict the strength of the pitch percept (Yost, 1996; Patterson *et al.*, 1996). In this case, it does not appear to be a useful guide for predicting which of these two stimuli will be perceived as more tonal.

The main difference between the two functions is the presence of a repeating series of equally spaced LP-SACF peaks in the case of the shuffled click train that are absent for the unshuffled click train. This repetition of equally spaced peaks is characteristic of the LP-SACFs of stimuli with a generally acknowledged clear pitch, such as harmonic tone complexes. If we accept the reasonable proposition that the presence of these additional regularly spaced peaks contributes to the overall tonality of click trains, we can conclude that the result is consistent with an autocorrelation approach to pitch perception.

F. Evaluation 6

Oxenham *et al.* (2004) used "transposed stimuli" to further explore the arguments surrounding periodicity theories of pitch. These transposed stimuli are high-frequency carrier tones multiplied by a half-wave rectified low-frequency sinusoid. Essentially, these are pulses of high-frequency tones. The simplest example is a 4 kHz carrier tone pulsed at 100 Hz [see Fig. 8(a), upper panel]. The stimulus was presented at a level of 77 phons in a white noise background, low-pass filtered at 600 Hz and at a level of 27 dB below the overall level of the tones.

This stimulus gives rise to a weak pitch sensation [see Fig. 2(a) in the work of Oxenham *et al.*, (2004)]. Figure 8(a) (lower plot) illustrates the LP-SACF for this stimulus (solid line). It has broad peaks around 10, 20, and 30 ms. However, these are clearly less prominent than those in the LP-SACF of a similar stimulus but in which the carrier frequency is 100 Hz (dashed line), which has a clear pitch. This result suggests that the pitch sensation of the transposed tone is weak, which is consistent with the findings of Oxenham *et al.* (2004).

A more complex stimulus is the combination of three carrier tones with frequencies of 4, 6.35, and 10.08 kHz modulated at 300, 400, and 500 Hz, respectively.[2] This stimulus was presented at an overall level of 65 dB SPL, in a background pink noise bandpass filtered (31.2–1000 Hz) at a similar level to that in the previous experiment.

If the auditory system aggregates periodicity information, one might expect that this would also be heard as a

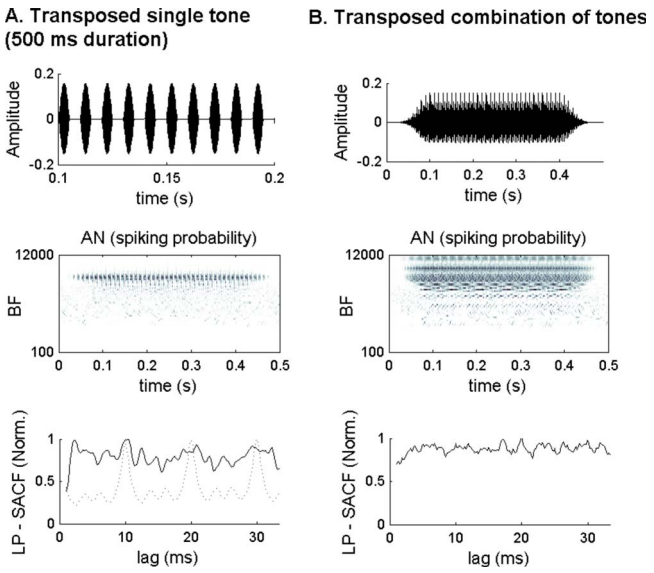


FIG. 8. (Color online) Autocorrelation analysis for 500 ms transposed stimuli (Oxenham *et al.*, 2004). (A) The upper panel shows the waveform of a single transposed pure tone of 4 kHz modulated at 100 Hz (amplified for better visualization). The middle panel shows the corresponding AN spiking probabilities (before the addition of noise). The lower panel shows the final LP-SACF. (B) Waveform of the sum of three transposed pure tones of 4, 6.35, and 10.08 kHz modulated at 300, 400, and 500 Hz, respectively. The middle panel shows the corresponding AN spiking probabilities (before the addition of noise). The lower panel shows the final LP-SACF.

weak 100 Hz pitch by analogy with a stimulus consisting of three pure tones at 300, 400, and 500 Hz. For this stimulus, however, the LP-SACF has an almost totally random structure [Fig. 8(b), lower plot]. Small peaks are present at 20 and 30 ms but the pattern is much less strongly modulated than the LP-SACF for the 100 Hz modulated carrier described above. This result agrees with the observations of Oxenham *et al.* (2004) who found that subjects performed even more poorly with multiple transposed tones than with a single transposed tone in a pitch discrimination task. Only one out of four of their subjects was able to discriminate the virtual pitch for these stimuli and that subject was not able to make pitch matches to the missing $F0$.

The experimenters analyzed their stimuli using an autocorrelation model very similar to the model studied in this report and obtained a different result from ours; the pure and transposed three-tone harmonic complexes produced very similar SACFs in their analysis. Both stimuli showed a distinct peak at a time interval corresponding to the reciprocal of the $F0$. Thus, the model correctly predicted that the $F0$ would be perceived in the case of the pure tones, but incorrectly predicted a similar pitch percept in the case of the transposed tones.

One of the several possible factors that could explain the difference between their analysis and that given in Fig. 8(b) are the changes to the auditory model at the level of the BM that have taken place since they performed their study. Oxenham *et al.* (2004) used a bank of linear gammatone bandpass filters to simulate cochlear filtering. The current model uses nonlinear filters to simulate the compression that takes place on the BM and this has consequences for the shape of the filters as a function of signal level.

However, the nonlinear response of the BM is such that the width of the filters increases at higher signal levels. At high signal levels, the frequency components will not necessarily be resolved as a result of the wider filters. Figure 8(b) (middle row) shows the pattern of AN responses across channels predicted by the model (before the addition of noise). The narrow bands implied by the term "resolved" are not clearly visible. It may be true that the BFs corresponding to the carrier frequencies contain only single-frequency activity but the intermediate channels are being excited by more than one frequency. The computation of the SACF involves a mandatory summation across *all* channels and therefore the intermediate channels will be well represented. Dreyer and Delgutte (2006) examined the AN response in cats to transposed tones and found that phase locking to transposed tones degraded substantially as signal levels were raised above threshold. SACFs resulting from stimuli containing unresolved components are typically flatter because the envelope of the stimulus has a greater influence than the individual sinusoidal components. This could be the case here. However, from our simulations, it is not conclusive that the within-channel interactions of the filters are responsible for the inaudibility of the pitch. Other possible factors include level-dependent compression, and possible saturation effects at signal levels such as the ones used in this experiment.

Nevertheless, at very low signal levels, our model replicates the results of the linear model used in the work of Oxenham *et al.* (2004) (not shown). Therefore, it is a prediction of the model that the pitch of the transposed complex should be audible at very low levels where the auditory system is functioning linearly and the auditory filters are sharply tuned, as assumed in the analysis of Oxenham *et al.* (2004). Such conditions may exist near threshold or in certain individuals for whatever reason. We note that one of the four subjects in the experimental study (their Fig. 3 subject S7) was able to make successful pitch discriminations at $F0$ for the transposed complex. A speculative explanation for this might be reduced compression in this subject. We do not, however, underestimate the difficulties of carrying out such a test given the need to demonstrate linear responses, narrow filters, and audibility of all components when close to threshold.

In summary, the results of Oxenham *et al.* (2004) are mirrored qualitatively in the response of the model to transposed stimuli and their data do not contradict the autocorrelation approach to modeling pitch perception. The difference in the modeling results could be a consequence of the nonlinear characteristics of the BM response in the new model. However, our conclusion should be qualified by the uncertainty that surrounds any model that purports to represent the exact pattern of action potentials in the human AN. We have no way of checking this and the evidence for such models is always indirect.

IV. DISCUSSION

The question is whether the autocorrelation approach should be completely rejected on the basis of recently published psychophysical studies. The results of experiments using click trains consisting of a mixture of regular and irregular interclick intervals (e.g., Kaernbach and Demany, 1998; Yost 1996) are certainly inconsistent with an approach based on the application of a long-term autocorrelation analysis to the acoustic *waveform* of the stimulus. However, the main tradition of the application of autocorrelation that began with Licklider (1951) has stressed the application of the analysis to the activity of the AN. Such analysis depends on a combination of two separate theories. The first is a theory of how the AN activity is generated; the second concerns the most appropriate method for analyzing this activity. If either theory has shortcomings, this will be reflected in a failure of the model to cope with some of the data.

In this report we have revisited a number of published psychophysical studies whose data had been argued to be inconsistent with an existing autocorrelation model of pitch. Here it has been shown that these difficulties can be reduced if modifications are made both to the peripheral model and to the method of analysis. Modifications to the peripheral model include the addition of nonlinearities in the BM filtering and an improved model of the generation of the IHC receptor potential. Modifications to the analysis method consisted of the introduction of lag-dependent time constants used in computing the ACFs, and an additional stage that integrated the output of the SACF over a longer time window. Together they give a more useful account of the data.

The 2 or 3 ms short time constant originally suggested by Licklider (1951) has proved in the past to be successful for many stimuli, but the pitch characteristics of irregular click-train stimuli cannot be so easily accommodated because the regularities in the stimuli can only be assessed over a longer time period. The solution to the problem can be found in Wiegrebe's (2001) study of the pitch of repeated pulses of noise. He suggested a multistage approach employing a second wider temporal integration window. The present cascade approach is the practical application of this idea. The longer time constant is the major contributor to the ability of the revised model to explain the pitch properties of the irregular click trains described above.

The shuffled click-train data of Yost *et al.* (2005) present another problem concerning the interpretation of the autocorrelation analysis. Their shuffled click trains were judged to be more "tonal" than the unshuffled click trains. This could not be predicted on the basis of the height of the highest peak in the LP-SACF because both shuffled and unshuffled click trains generated peaks of (on average) the same height. The key difference between the two functions was to be found in the pattern of minor peaks. A shuffled click train can be distinguished from the unshuffled version by the presence of a repeating series of equally spaced minor peaks. These peaks are commonly observed in the SACF (and in the LP-SACF) of harmonic tone complexes but often ignored by researchers as redundant. However, these new stimuli indicate that the repeating peaks should contribute to our predictions of the salience of the perceived pitch.

The autocorrelation analysis will be of limited value if it is based on the output of an inadequate peripheral model. Such models remain primitive but are subject to continuous revision. Our insights into the significance of the many

subtleties of peripheral auditory processing are developing in parallel. Of particular interest are the ideas of Pressnitzer *et al.* (2002) concerning the role of the half-wave rectification that occurs in the generation of the receptor potential in the IHC. Here it has been shown that their proposal survives translation to a more sophisticated auditory model and the new cascade method of analyzing the model output.

Another nonlinearity in peripheral processing occurs at the level of the BM and involves compression of the stimulus waveform at signal frequencies close to the filter BF. This compression is less evident at remote frequencies with the consequence that the width of the filter increases with signal level. As a result, the individual frequency components of a stimulus spread their effects more widely over the BM and the excitation pattern changes radically as the level of the stimulus is increased. This is one of the possible factors in explaining the success of our new model in the study of Oxenham *et al.* (2004). More research is needed to understand precisely which aspect of the auditory peripheral model is primarily responsible for this perceptual phenomenon and for the perception of alternating click trains (Carlyon *et al.*, 2008).

Both peripheral models and methods for analyzing their output are continuing to evolve and we must expect increasingly sophisticated accounts of pitch perception to emerge as a consequence. The novel and challenging stimuli described above have an important role to play in this evolution. Nevertheless, we conclude that, for the present, Licklider's (1951) view that pitch perception can be understood in terms of a periodicity analysis of the activity of the AN remains intact.

ACKNOWLEDGMENTS

This work was supported by EmCAP (Emergent Cognition through Active Perception, 2005-2008) a research project in the field of Music Cognition funded by the European Commission (FP6-IST, Contract No. 013123). We thank Dr. Robert P. Carlyon, Professor Brian C. J. Moore, and an anonymous reviewer for their comments and advice. E.B.B. thanks Dr. Martin Coath for his support.

[1]Software used in this study is available on request from the authors.
[2]For this evaluation, the range of BF channels in the auditory model was extended to 12 kHz for all stimuli.

Bernstein, J. G. W., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Cariani, P. A., and Delgutte, B. (**1996a**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698–1716.

Cariani, P. A., and Delgutte, B. (**1996b**). "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch, and the dominance region of pitch," J. Neurophysiol. **76**, 1717–1734.

Carlyon, R. P. (**1996**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P., Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

Carlyon, R. P., Mahendran, S., Deeks, J. M., Long, C. J., Axon, P., Baguley, D., Bleeck, S., and Winter, I. M. (**2008**). "Behavioral and physiological correlates of temporal pitch perception in electric and acoustic hearing," J. Acoust. Soc. Am. **123**, 973–985.

Ciocca, V., and Darwin, C. J. (**1999**). "The integration of nonsimultaneous frequency components into a single virtual pitch," J. Acoust. Soc. Am. **105**, 2421–2430.

Denham, S. L. (**2005**). "Dynamic iterated ripple noise: Further evidence for the importance of temporal processing in auditory perception," BioSystems **79**, 199–206.

Dreyer, A., and Delgutte, B. (**2006**). "Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization," J. Neurophysiol. **96**, 2327–2341.

Gockel, H., Plack, C. J., and Carlyon, R. P. (**2005**). "Reduced contribution of a nonsimultaneous mistuned harmonic to residue pitch," J. Acoust. Soc. Am. **118**, 3783–3793.

Grose, J. H., Hall, J. W., and Buss, E. (**2002**). "Virtual pitch integration for asynchronous harmonics," J. Acoust. Soc. Am. **112**, 2956–2961.

Hall, J. W., and Peters, R. W. (**1981**). "Pitch for nonsimultaneous successive harmonics in quiet and noise," J. Acoust. Soc. Am. **69**, 509–513.

Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanisms involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.

Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," J. Acoust. Soc. Am. **104**, 2298–2306.

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.

Meddis, R. (**2006**). "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," J. Acoust. Soc. Am. **119**, 406–417.

Meddis, R., and Hewitt, M. J. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I. Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Micheyl, C., and Carlyon, R. P. (**1998**). "Effects of temporal fringes on fundamental-frequency discrimination," J. Acoust. Soc. Am. **104**, 3006–3018.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, J. (**1996**). "The relative strength of tone and noise components in iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286–3294.

Plack, C. J., and White, L. (**2000**). "Perceived continuity and pitch perception," J. Acoust. Soc. Am. **108**, 1162–1169.

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

Pressnitzer, D., de Cheveigné, A., and Winter, I. M., (**2002**). "Perceptual pitch shift for sounds with similar waveform autocorrelation," ARLO **3**, 1–6.

Pressnitzer, D., de Cheveigné, A., and Winter, I. M., (**2004**). "Physiological correlates of the perceptual pitch shift for sounds with similar waveform autocorrelation," ARLO **5**, 1–6.

Slaney, M., and Lyon, R. F. (**1990**). "A perceptual pitch detector," International Conference on Acoustics, Speech, and Signal Processing, (ICASSP-90), Albuquerque, USA.

Sumner, C. J., O'Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (**2002**). "A revised model of the inner-hair cell and auditory nerve complex," J. Acoust. Soc. Am. **111**, 2178–2189.

Wiegrebe, L. (**2001**). "Searching for the time constant of neural pitch extraction," J. Acoust. Soc. Am. **109**, 1082–1091.

Yost, W. A. (**1996**). "Pitch of iterated rippled noise," J. Acoust. Soc. Am. **100**, 511–518.

Yost, W. A., Mapes-Riordan, D., Shofner, W., Dye, R., and Sheft, S. (**2005**). "Pitch strength of regular interval click trains with different length "runs" of regular intervals," J. Acoust. Soc. Am. **117**, 3054–3068.

Use of psychometric-function slopes for forward-masked tones to investigate cochlear nonlinearity[a)]

Kim S. Schairer,[b)] Jessica Messersmith, and Walt Jesteadt
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 13 September 2007; revised 8 July 2008; accepted 11 July 2008)

Schairer *et al.* [(2003). "Effects of peripheral nonlinearity on psychometric functions for forward-masked tones," J. Acoust. Soc. Am. **133**, 1560–1573] demonstrated that cochlear nonlinearity is reflected in psychometric-function (PF) slopes for 4 kHz forward-masked tones. The goals of the current study were to use PF slopes to compare the degree of compression between signal frequencies of 0.25 and 4 kHz in listeners with normal hearing (LNH), and between LNH and listeners with cochlear hearing loss (LHL). Forward-masked thresholds were estimated in LNH and LHL using on- and off-frequency maskers and 0.25 and 4 kHz signals in three experiments. PFs were reconstructed from adaptive-procedure data for each subject in each condition. Trends in PF slopes across conditions suggest comparable compression at 0.25 and 4 kHz, and potentially a wider bandwidth of compression in relative frequency at 0.25 kHz. This is consistent with other recent behavioral studies that revise earlier estimates of less compression at lower frequencies. The preliminary results in LHL demonstrate that PF slopes are abnormally steep at frequencies with HL, but are similar to those for LNH at frequencies with NH. Overall, the results are consistent with the notion that PF slopes reflect degree of cochlear nonlinearity and can be used as an additional measure of compression across frequency. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2968686]

PACS number(s): 43.66.Dc, 43.66.Ba [BCM] Pages: 2196–2215

I. INTRODUCTION

The overall goal of the current set of experiments was to expand the results of Schairer *et al.* (2003b), which demonstrated that cochlear nonlinearity is reflected in slopes of psychometric functions (PFs) for forward-masked, 4 kHz tones. PF slopes are used here to investigate cochlear nonlinearity at 0.25 kHz in comparison to 4 kHz in listeners with normal hearing (LNH) and in listeners with cochlear hearing loss (LHL). Sections I A–I D describe how cochlear nonlinearity is reflected in forward masking, how cochlear nonlinearity is reflected in PF slopes, and how PF slopes can be used to investigate compression at low frequencies and in ears with HL.

A. Forward masking and cochlear nonlinearity

Forward masking refers to the condition in which the threshold for a short-duration signal is elevated in the presence of a preceding masker. Forward-masked thresholds have been used to estimate frequency selectivity (see, e.g., Nelson and Freyman, 1984) as well as auditory time constants or temporal resolution (see, e.g., Nelson and Pavlov, 1989; Nelson and Freyman, 1987; and Jesteadt *et al.*, 1982). Many recent studies have used forward masking to assess the amount of cochlear compression or nonlinear basilar-membrane (BM) response growth (see, e.g., Lopez-Poveda *et al.*, 2003; Nelson *et al.*, 2001; Nelson and Schroder, 2004; Oxenham and Plack, 1997; Plack and Oxenham, 1998; Rosengard *et al.*, 2005; Schairer *et al.*, 2003b; and Williams and Bacon, 2005).

Forward masking is thought to reflect either temporal integration or adaptation, or some combination of both processes (Chatterjee, 1999; Oxenham, 2001; Plack and Oxenham, 1998). Temporal integration is conceptualized as an overlap of the internal representations of the signal and the masker that occurs centrally; it can also be thought of as "persistence" of neural activity after the masker offset. Adaptation is the reduction of activity or response to a signal after presentation of a masker. It may occur at different places in the auditory periphery, such as the synapse between the inner hair cells (IHCs) and the eighth nerve, and the involvement of neural adaptation is supported by a recent computer modeling study (Meddis and O'Mard, 2005). Peripheral adaptation cannot entirely account for the observed threshold shift, however, because forward masking can be obtained in individuals with cochlear implants in whom stimulation bypasses the cochlea and IHC-eighth nerve synapse (Chatterjee *et al.*, 2006; Chatterjee, 1999; Shannon, 1990). Thus, forward masking is due not only to peripheral adaptation, but is almost certainly influenced by a retrocochlear process.

[a)]Portions of this work were presented in Schairer, K. S., Messersmith, J., and Jesteadt, W. (2005). "Psychometric-function slopes for forward-masked tones in listeners with cochlear hearing loss," J. Acoust. Soc. Am. **117**, 2599 (Abstract) and Schairer, K. S, and Jesteadt, W. (2003). Evidence of peripheral nonlinearity in psychometric function slopes of forward-masked tones at 250 and 4000 Hz," J. Acoust. Soc. Am. **113**, 2226 (Abstract).

[b)]Present address: University of Wisconsin, 1975 Willow Drive, Madison, WI 53706. Electronic mail: kschairer@wisc.edu

For sinusoidal forward maskers and signals, the amount of forward masking is greater when the masker and signal are at the same frequency (on frequency) than when the masker is at a different typically lower frequency (off frequency). Slopes of growth of masking (GOM), or the change in signal level at threshold for a given change in masker level, can be obtained in conditions in which the signal level is varied to estimate threshold in different fixed masker-level conditions [variable signal (VS)] or conditions in which the masker level is varied in different fixed signal level conditions [variable masker (VM)]. In VS conditions, slopes of GOM are less than 1 dB/dB in on-frequency conditions and are similar at low and high frequencies in LNH (see, e.g., Jesteadt *et al.*, 1982). Signal level at threshold increases at a faster rate for off-frequency than for on-frequency conditions at moderate masker levels (see, e.g., Luscher and Zwislocki, 1949 and Plack and Drga, 2003). Comparable effects can be observed by fixing the signal at a low level in a VM paradigm with different signal delays in different conditions using on- and off-frequency maskers (Nelson and Schroder, 2004; Nelson *et al.*, 2001). The resulting temporal masking curves (TMCs) show the effects of both signal delay and masker frequency. GOM functions obtained with VS procedures are used in the current set of experiments.

BM response growth is nonlinear at characteristic frequency (CF) (see, e.g., Ruggero *et al.*, 1997 and Yates *et al.*, 1990). That is, if a tone with a frequency equal to CF is presented, the amount of BM deflection as a function of stimulus level is linear at low levels and gradually becomes compressive as stimulus level increases. If the recording is made from the same place on the BM, but a lower or higher frequency is presented, the response growth becomes more linear. Response growth also becomes more linear in ears with HL, presumably due to the loss of outer hair cell (OHC) function.

Although forward masking almost certainly has a retrocochlear contribution, it is thought that cochlear nonlinearity is reflected in slopes of GOM and TMC in on-frequency and off-frequency masker conditions. In a VS paradigm in LNH, the shallow GOM in the moderate masker-level range in on-frequency masker conditions is thought to be due to compression of the on-frequency masker (Plack and Oxenham, 1998). In a VM paradigm, the off-frequency masker condition has been used as a linear reference in a ratio of slopes of GOM for on- and off-frequency masker conditions to estimate the degree of compression in LNH and reduced compression in LHL (Oxenham and Plack, 1997). On-frequency TMCs are steeper and off-frequency TMCs are shallower in LNH, whereas the differences in TMC slopes decrease and functions become more parallel across masker frequencies in LHL (Nelson *et al.*, 2001; Plack *et al.*, 2004; Rosengard *et al.*, 2005).

A three-stage model described by Plack and Oxenham (1998) can be used to predict the pattern of thresholds in a VS paradigm at various signal delays and masker levels. The model has a compressive nonlinearity as its first stage, followed by a sliding temporal integration window, and a decision process that compares the level at the output of the window to determine which interval contained the signal.

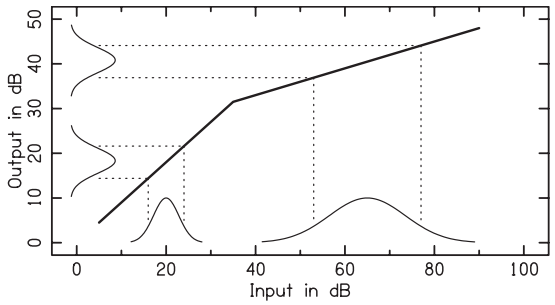


FIG. 1. Relationship between compressive nonlinearity and psychometric-function (PF) slope (from Schairer *et al.*, 2003b). In a forward-masking condition that produces a low signal level at threshold, the signal will pass through the more linear portion of the nonlinearity during threshold estimation. Changes in signal level at the input and output of the function will be similar, the standard deviation of the underlying distributions will be small, and the corresponding PF slope will be steep. In conditions that produce higher masked thresholds, the signal will pass through the more compressive region of the function. The signal level will have to change by a greater amount to produce the same change in signal level at the output of the nonlinearity, the underlying standard deviation will be larger, and the PF slopes will be shallower.

The model uses a two-line approximation to the function representing compression, with linear growth of response at the signal frequency for low-level on-frequency maskers and constant compression above some breakpoint. The amount of compression can be determined by fitting a two-line function to the data.

There is a consensus in literature that the difference in on- and off-frequency GOM or TMC functions can be used to estimate compression at signal frequencies of 1 kHz or higher, although Wojtczak and Oxenham (2007) recently have questioned the necessary assumption that the rate of recovery is the same for on- and off-frequency forward maskers. The literature on compression at lower frequencies, reviewed here in a later section, suggests that the comparison of on- and off-frequency masking may be problematic because the assumption of linearity in the off-frequency reference condition may not be valid. Estimates of compression based on PF slopes for forward-masked tones do not require comparison of results obtained in on- and off-frequency conditions.

B. Cochlear nonlinearity reflected in PF slopes

PFs showing percent correct (PC) as a function of the level of the variable stimulus take different forms and have different meanings in VS and VM paradigms. In the VS paradigm, PC increases as the signal level increases. In a VM paradigm, PC decreases as the masker level increases, but the data can be fitted using the same equations and procedures. Schairer *et al.* (2003b) demonstrated that PFs obtained in a VS paradigm in LNH, with either on- or off-frequency forward maskers, have shallower slopes under conditions that result in large amounts of forward masking and interpreted the change in slope as a measure of compression. The rationale for this argument is shown in Fig. 1 (from Schairer *et al.*, 2003b). In a VS paradigm, conditions that produce low signal levels at masked threshold will result in the signal passing through the more linear portion of the peripheral

nonlinearity as the signal level varies during the threshold estimation procedure. PF slopes would be steep in this case because a small range of signal levels would be necessary to establish threshold. In conditions that produce higher signal levels at masked threshold, the signal will pass through the more compressive region of the function and will have to change by a greater amount to produce the same change at the output of the nonlinearity. In this case, the PF slope would be shallower. The nonlinearity is depicted in Fig. 1 as a two-line approximation, as represented in the Plack and Oxenham (1998) model. It is assumed that the true underlying nonlinearity has a more gradual transition into the compressive region (see, e.g., Neely and Jesteadt, 2005). The decrease in PF slope as a function of signal level at threshold should be the same for the on-frequency and off-frequency masker conditions in a VS paradigm, because for any given PF, signal level is the only varying parameter. The signal is "on frequency" by definition and will be affected by the nonlinearity at its place regardless of the masker frequency.

In a VM paradigm, however, PF slopes for on- and off-frequency conditions should be different because in this case, the varied parameter is the masker level, and the masker response growth will be different at the place of the signal depending on the relation between the masker and the signal frequency. In on-frequency conditions, results should be similar to the VS case because the masker will be compressed just as the signal is compressed. In off-frequency masker conditions, however, the masker response growth will be linear at the place of the signal and PF slopes should be steep and parallel, regardless of the masker level at the threshold. Schairer *et al.* (2003b) provided evidence to support these predictions in LNH using a 4 kHz signal and on- and off-frequency maskers.

C. Use of PF slopes to investigate compression at low frequencies

Behavioral studies have demonstrated a significant amount of compression at signal frequencies of 1 kHz and above (see, e.g., Nelson *et al.*, 2001 and Oxenham and Plack, 1997), but earlier studies showed less compression at lower frequencies (see, e.g., Hicks and Bacon, 1999 and Plack and Oxenham, 2000). These studies relied on the assumption that response growth at the place of the signal is linear for off-frequency maskers. Several authors [Lopez-Poveda *et al.* (2003), Plack and Drga (2003), Plack *et al.* (2004), and Plack and O'Hanlon (2003)] have pointed out that if compression affects a wider range of frequencies relative to CF (Rhode and Cooper, 1996), then response growth of on-frequency and off-frequency maskers at the place of the signal would be similar and would confound methods that rely on the linearity of the off-frequency response. TMC and GOM methods that do not rely on the assumption of linear growth at the signal place for off-frequency maskers, or that use off-frequency growth at higher frequencies as a linear reference, have demonstrated low-frequency compression that is greater than in previous studies and is comparable to compression estimates for higher-frequency signals (see, e.g., Lopez-Poveda and Alves-Pinto, 2008; Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; and Williams and Bacon, 2005). A recent report (Stainsby and Moore, 2006) suggests, however, that decay of forward masking is not independent of signal frequency and that it may not be appropriate to use a high-frequency reference to estimate compression at low frequencies. Additivity of masking (Plack and O'Hanlon, 2003; Plack *et al.*, 2005) methods do not depend on an off-frequency masker condition or on uniform decay of forward masking across frequency and can be used to directly compare compression at low and high frequencies. In addition, Lopez-Poveda and Alves-Pinto (2008) described a method of using TMCs to estimate compression that does not require the assumption of uniform decay of forward masking across frequencies. In their new method, the comparison is made between TMC slopes obtained at two different probe levels within masker-frequency conditions. PF slopes are another measure of cochlear compression that requires no assumptions regarding the decay of forward masking or the degree of off-frequency compression and could provide an independent measure of the frequency range of compression at low frequencies.

D. Use of PF slopes to investigate compression in ears with hearing loss

There are few studies that address PFs for detection of tones in quiet in LHL. Marshall and Jesteadt (1986) obtained PFs for LHL and LNH for 0.5 and 4 kHz tones. PFs were estimated by straight-line least-squares fits weighted by number of trials for d' as a function of level. They reported no differences in PF slopes between the two groups. In contrast, Arehart *et al.* (1990) estimated PF slopes using linear regression and probit analysis for 0.5, 2, 4, and 8 kHz tones in quiet in groups of LHL and LNH. The slopes of the two groups overlapped, but the LHL had significantly steeper slopes in the 2 kHz condition and some LHL had abnormally high slopes across frequencies.

Thresholds for forward-masked tones under a number of different conditions have been reported for LHL due to OHC dysfunction. However, PFs for those conditions have not been reported. Following the logic described for LNH (Fig. 1), it is predicted that PF slopes for forward-masked tones in a VS condition should be steeper for LHL than for LNH. This is because LHL presumably lack the nonlinearity that is hypothesized to be responsible for the decrease in PF slope with masked threshold in the LNH. The function in Fig. 1 would be a straight line with a slope near 1.0, rather than a two-part function with a compressive region. Thus, any changes in the external signal across the range would be represented by the same amount of change internally, similar to conditions that would produce thresholds in the lower-level (linear) portion of the function for LNH. PFs should remain steep and parallel across the masker frequency and the threshold level.

II. EXPERIMENT 1: PSYCHOMETRIC FUNCTIONS FOR FORWARD-MASKED, LOW- AND HIGH-FREQUENCY TONES IN A VARIABLE-SIGNAL PARADIGM IN LISTENERS WITH NORMAL HEARING

The purpose of experiment 1 was to use PF slopes to test the hypothesis that compression is comparable at low- and

high-signal frequencies. Note that the purpose is to compare relative compression between frequencies and not to estimate specific parameter or compression values. This method has the benefit that it does not rely on a comparison of off- and on-frequency conditions. Because both types of maskers should have a similar effect on the PF slope, based on the argument in the Introduction and in Schairer *et al.* (2003b), only on-frequency conditions were included. Forward masking was measured in on-frequency VS conditions at both 0.25 and 4 kHz in a group of LNH. It was predicted that PF slopes would decrease similarly as a function of signal threshold in both the 0.25 and 4 kHz conditions, suggesting comparable cochlear compression at these frequencies.

A. Subjects

Five paid adults, two males and three females, ages 19–32 years (mean=25.2, standard deviation [SD]=6.3) served as subjects. Three were college students, one was the first author, and one was a research associate from another laboratory. Hearing for the three college students had been screened at 0.5, 1, 2, and 4 kHz within the past year using the same two-interval forced-choice (2IFC) adaptive procedure used in the experiment. Thresholds were at or better than 15 dB sound pressure level (SPL) for all test frequencies, bilaterally, for each subject. Hearing for the other two subjects had been tested, using clinical procedures, as part of a research protocol for another laboratory. Hearing thresholds were at or better than 15 dB hearing level (dBHL), bilaterally, at the same frequencies for both subjects.

B. Stimuli and apparatus

Signals and maskers were all pure tones. Signal frequencies were 0.25 or 4 kHz, signal duration was 10 ms (5 ms rise/fall), and signal delay was 10 ms (from offset of masker to onset of signal). A 10 ms delay was selected in order to provide a sufficient amount of masking (in comparison to decreased masking at longer delays) and to avoid abnormally steep PFs observed in some listeners in very short (e.g., 0 ms) signal delay conditions (as observed in Schairer *et al.*, 2003b). The on-frequency forward masker duration was 200 ms (2 ms rise/fall) and the maskers were presented at levels of 30, 50, 70, and 90 dB SPL in separate conditions. All stimuli were generated with ramps that were shaped using a half-cycle of a raised-cosine function. Thresholds for each signal in quiet also were obtained.

Stimuli were generated digitally at a sampling rate of 50 kHz using a Tucker-Davis Technologies (TDT) array processor (TDT AP2) and 16 bit digital-to-analog converters (TDT DD1). The forward masker was generated on one channel of the DD1, while the signal was generated on the other. The output of each channel was low-pass filtered at 20 kHz (TDT FT6) and attenuated (TDT PA4), then the outputs of the two channels were combined (TDT SM3) and presented monaurally to the left ear through a headphone buffer (TDT HB6), a remote passive attenuator in the sound treated room, and a Sennheiser HD 250 Linear II headphone. Parallel use of multiple attenuators, summers, and headphone buffers made it possible to simultaneously test up to four listeners. Subjects 078 and 175 were tested individually. Subjects 100, 102, and 108 were tested as a group.

C. Adaptive-procedure thresholds

Thresholds were obtained in a VS paradigm using a two-track 2IFC adaptive procedure with decision rules to estimate 71 PC (two-down, one-up) on one track and 87 PC (five-down, one-up) on the other track (Levitt, 1971), with a 4 dB step size. Two tracks with different decision rules were used to obtain a larger range of PCs for the PF fits. Threshold for each track was calculated as the mean of the reversal levels after the fourth reversal. Five 200-trial repetitions were obtained in each condition. The first repetition was excluded as practice. A total of 800 trials, or 400 trials per track, were available for further analysis. There were two exceptions. Four repetitions were obtained in the 0.25 kHz quiet-threshold condition for subject 102. After excluding the first repetition as practice, a total of 600 trials remained for this subject in this condition. For subject 175, one repetition in the 4 kHz, 90 dB masker condition yielded only one reversal after the fourth with which to calculate the threshold. This repetition was discarded and an extra repetition was obtained. Mean thresholds were calculated across repetitions for each track, and then the mean across the two tracks was calculated as the adaptive-procedure threshold for each subject in each condition. This threshold was an estimate of the level required for 79 PC.

D. Psychometric-function fits

All trials from both tracks were combined to fit PFs. The combined data were trimmed such that signal levels that were presented on <30 trials across both tracks and/or associated with <50 PCs were excluded. As the signal level increased, subsequent signal levels were deleted after the first occurrence of 100 PC. The purpose was to remove multiple levels with associated PCs of 100 that would skew the fits to be shallower than they probably were, and to avoid nonmonotonic functions. PFs were fitted for each condition using Dai's (1995) modification of the equation proposed by Egan *et al.* (1969), in which $d' = (I/a)^b$, where I is signal power, $10 \log(a)$ is the signal level required for $d' = 1$, and b is the slope of a line in $\log d'$ versus signal level. Dai's (1995) fitting procedure minimizes the deviation between expected and obtained proportion correct, expressed in units of χ^2. To provide a more familiar measure of goodness of fit in terms of variance accounted for, an r^2 was calculated for each PF. For a total of 50 PF fits, including the quiet-threshold conditions, the r^2 values ranged from 0.79 to 1.00 (mean=0.95). Note that the fitting procedure yields a measure of the threshold, $10 \log(a)$, that is computed very differently than the adaptive threshold described above. The correlation between the two threshold types across frequency, masker level, and subject (not including quiet thresholds, but including all masked thresholds regardless of associated slopes) was 0.998.

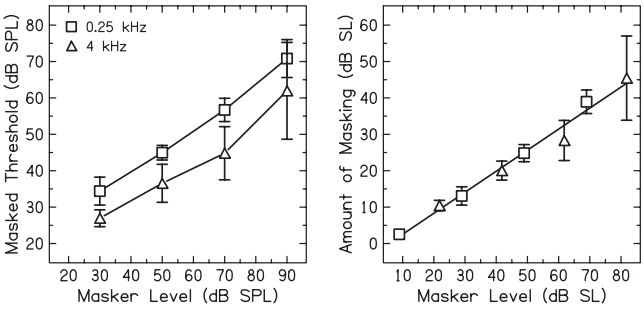


FIG. 2. Mean masked adaptive threshold in dB SPL across listeners with normal hearing (LNH) as a function of masker dB SPL (left panel) and mean amount of masking (i.e., signal dB SL) as a function of masker dB SL (right panel) for 0.25 kHz (squares) and 4 kHz (triangles) signal conditions in experiment 1. The mean masker thresholds in quiet for the LNH in experiment 2 were used to estimate masker dB SL in experiment 1. The error bars represent +/−1 standard deviation. The on-frequency forward masker duration was 200 ms, and maskers were presented at 30, 50, 70, and 90 dB SPL in separate conditions. Signal duration was 10 ms and signal delay was 10 ms, presented in a variable signal (VS) paradigm. The amount of masking increases as a function of masker level at a similar rate for both frequencies. The line in the right panel is fitted with all data points (across frequency) and has a slope of 0.58 dB/dB.

E. Results and discussion

1. Adaptive-procedure thresholds

Jesteadt *et al.* (1982) noted that forward masking appears more uniform across frequency conditions when thresholds are plotted in units of sensation level (SL). Figure 2 shows mean masked threshold across subjects as a function of masker dB SPL in the left panel and mean amount of masking as a function of dB SL in the right panel. Amount of masking was calculated for each subject in each condition by subtracting the signal threshold in quiet from the masked thresholds. Thresholds for the masker in quiet were not obtained for this group and some members of the group were not available to return to run those conditions. The mean masker thresholds in quiet for the LNH in experiment 2 were therefore used to estimate masker levels in dB SL in experiment 1.Threshold increases as a function of masker level at a similar rate for both frequencies, but appears to be consistently greater in the 0.25 kHz condition when expressed in terms of masker dB SPL (left panel). However, when corrected by masker threshold in quiet, the amount of masking is similar across signal-frequency conditions (right panel) and is well described by a single line with a slope of 0.58 dB/dB.

2. Psychometric-function fits

Figure 3 shows fits to the individual data points for the 30 dB (low threshold) and 90 dB (high threshold) masker conditions for both signal frequencies for each subject. There does not appear to be a difference in the goodness of fit to the data points as a function of frequency or masker level. Thus, the trend of shallower slopes as a function of signal level cannot be accounted for by poorer PF fits at higher levels. Table I provides a summary of the PF parameters and r^2 values for each subject in each condition. Figure 4 shows PF slope as a function of PF threshold in dB SL for each subject and for the geometric mean across subjects (geometric mean slope as a function of arithmetic mean PF threshold). One data point from the 4 kHz, 30 dB masker condition is missing from the panel for subject 175 because the slope is excessive (5.6). Despite the variability across subjects, on average PF slope decreases as a function of threshold at a similar rate for both frequencies. The pattern of slope change

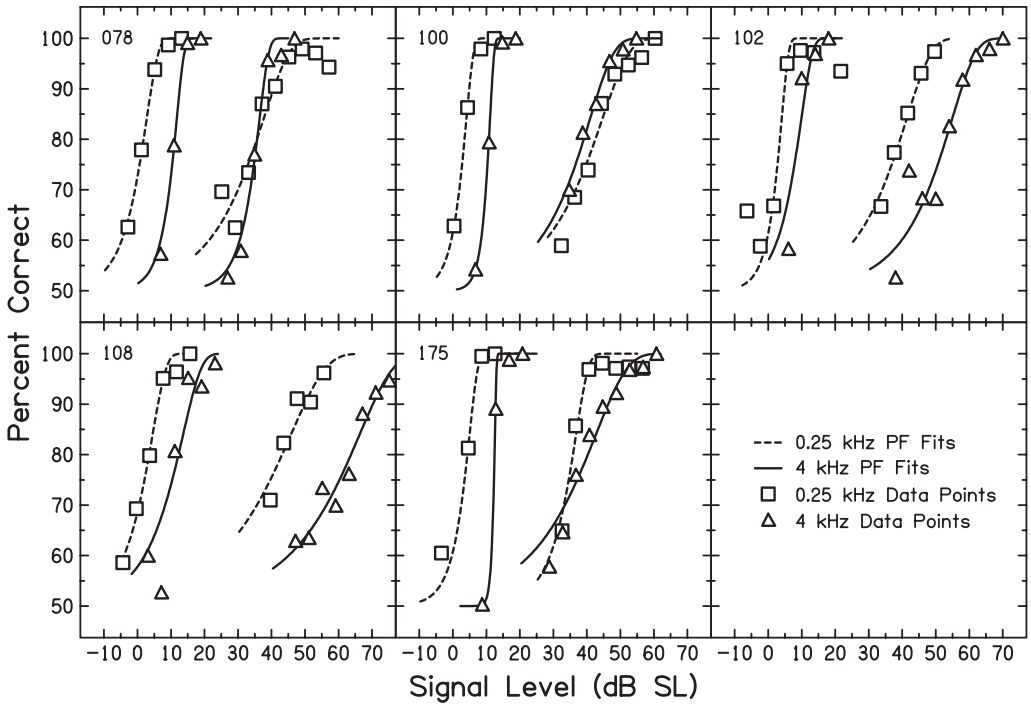


FIG. 3. Psychometric-function (PF) fits to the data points in the lowest (30 dB SPL) and highest (90 dB SPL) masker-level conditions for each subject in experiment 1. Data points are represented as in Fig. 2. The dashed and solid lines represent PFs in the 0.25 and 4 kHz signal conditions, respectively. PF threshold, slope, and goodness of fit (r^2) parameters for all subjects and conditions (including those not shown here) are listed in Table I.

TABLE I. Experiment 1 psychometric function (PF) parameters for listeners with normal hearing (LNH) in variable signal (VS), on-frequency forward-masking conditions [QT=quiet threshold; SD =standard deviation].

Signal frequency	Masker level	Parameter	078	100	102	108	175	Mean(SD)
0.25 kHz	QT	Threshold	30.84	27.64	38.37	32.36	31.39	32.12 (3.92)
		Slope	1.79	0.99	0.74	0.63	2.04	1.11
		r^2	0.99	0.98	0.88	0.92	0.95	
	30	Threshold	31.72	30.55	41.29	34.74	35.07	34.67 (4.17)
		Slope	0.80	1.31	1.34	0.69	1.13	1.02
		r^2	0.99	0.99	0.87	0.99	0.96	
	50	Threshold	43.14	45.33	48.51	44.34	45.09	45.28 (2.0)
		Slope	0.72	0.57	1.10	0.37	1.25	0.73
		r^2	0.98	0.98	0.97	0.94	1.0	
	70	Threshold	53.20	56.42	59.29	57.74	55.14	56.36 (2.34)
		Slope	0.48	0.47	0.50	0.28	0.73	0.47
		r^2	0.96	0.89	0.93	0.82	0.92	
	90	Threshold	64.40	68.00	75.77	72.74	66.24	69.43 (4.71)
		Slope	0.37	0.35	0.38	0.27	0.76	0.40
		r^2	0.93	0.96	0.98	0.89	0.98	
4 kHz	QT	Threshold	17.15	13.28	22.01	16.91	15.27	16.93 (3.24)
		Slope	1.05	1.93	1.16	2.42	1.36	1.58
		r^2	0.98	1.00	0.98	0.99	1.00	
	30	Threshold	27.68	23.71	30.10	27.67	27.57	27.35 (2.30)
		Slope	1.24	2.23	0.83	0.52	5.57	1.46
		r^2	0.99	0.99	0.85	0.84	1.00	
	50	Threshold	36.16	30.66	43.39	37.32	33.04	36.11 (4.84)
		Slope	1.11	0.73	0.78	0.30	0.91	0.71
		r^2	0.99	0.92	0.97	0.86	0.95	
	70	Threshold	40.94	38.86	51.49	50.56	39.08	44.18
		Slope	1.07	0.52	0.53	0.20	0.83	0.55
		r^2	0.99	0.96	0.93	0.79	0.95	
	90	Threshold	52.09	50.42	73.14	77.92	53.04	61.32 (13.11)
		Slope	1.00	0.39	0.40	0.28	0.30	0.42
		r^2	0.99	0.98	0.90	0.95	0.95	

with level is related to the form of the compressive nonlinearity. The function fitted to the mean data in Fig. 4 is the reciprocal of a quadratic compression function, as described by Neely and Jesteadt (2005).

In summary, thresholds in on-frequency VS conditions increased as a function of masker level and PF slopes decreased as a function of signal threshold similarly for both the 0.25 and 4 kHz signals. Results suggest comparable cochlear on-frequency compression at these two frequencies.

III. EXPERIMENT 2: PSYCHOMETRIC FUNCTIONS FOR FORWARD-MASKED, LOW- AND HIGH-FREQUENCY TONES IN A VARIABLE-MASKER PARADIGM IN LISTENERS WITH NORMAL HEARING

The purpose of experiment 2 was to test the hypothesis that the bandwidth of compression at 0.25 Hz is wider than at 4 kHz. VS conditions do not provide information about the range of frequencies that are compressed at each CF because on- and off-frequency conditions produce similar PF slopes. However, for any given PF in a VM condition, the varied parameter is the masker level, which will grow compressively at the place of the signal in the on-frequency condition, and linearly or less compressively at the place of the signal for the off-frequency condition. If the bandwidth of compression at 0.25 kHz is wider than at 4 kHz signal, then PF slopes might decrease as a function of threshold in both on-frequency and off-frequency conditions, because the off-frequency masker may grow compressively at the place of the signal, just as the on-frequency masker does. Thus, the prediction is that in the 4 kHz signal condition, PF slopes will decrease as a function of masker threshold for the on-frequency conditions and will remain steep for off-frequency conditions; in the 0.25 kHz signal condition, PF slopes will decrease as a function of threshold similarly in on- and off-frequency conditions.

A. Subjects

Four paid adults, one male and three females, ages 19–23 years (mean=20.5, SD=1.7) served as subjects. Three were college students; and one was the second author, who was a graduate research assistant in the laboratory. Subjects had hearing thresholds less than 20 dB SPL at 0.5, 1, 2, and 4 kHz, bilaterally.

B. Stimuli and apparatus

The stimuli were delivered through the same equipment as experiment 1. Signal frequencies were 0.25 or 4 kHz, signal duration was 10 ms (5 ms rise/fall), and signal delay was 5 or 10 ms. Forward masker duration was 200 ms (5 ms

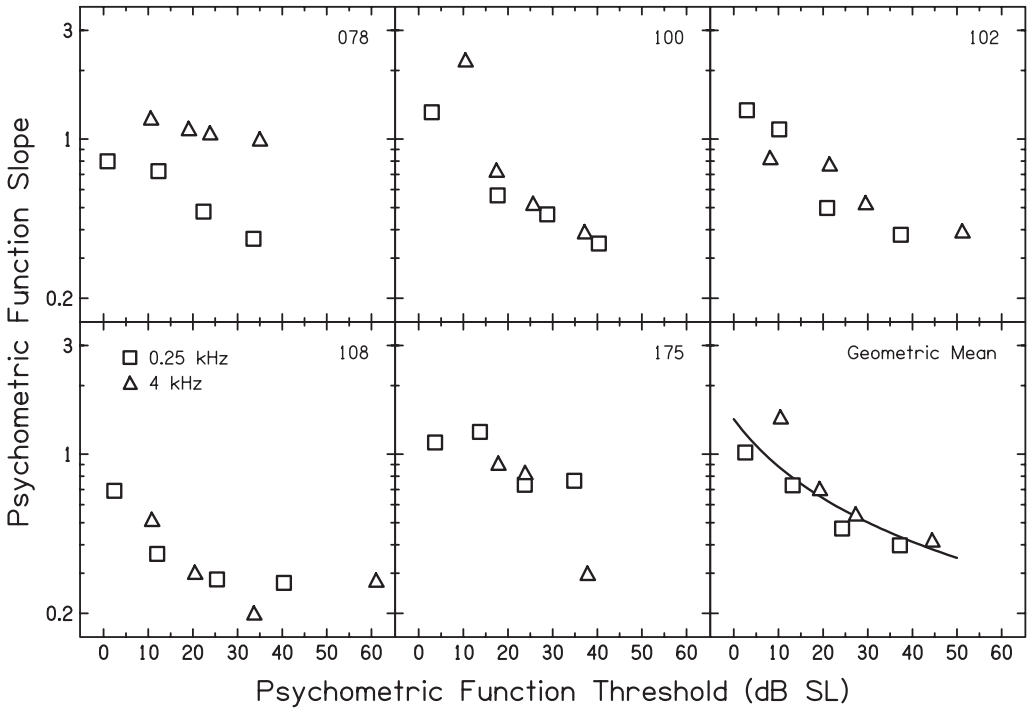


FIG. 4. Psychometric-function (PF) slope as a function of PF threshold in dB SL for each subject and the geometric mean across subjects for experiment 1. Frequencies are represented as in Fig. 2. On average, although there is variability across subjects, PF slope decreases as a function of masked threshold for both frequencies at approximately the same rate. The function fitted to the geometric mean data is the reciprocal of the quadratic compression function described by Neely and Jesteadt (2005).

rise/fall) and masker frequencies were 0.15 and 2.4 kHz in the off-frequency conditions. The off-frequency maskers were selected such that the ratios between the masker and the signal frequencies were identical for 0.25 and 4 kHz signals. Fixed signal levels were selected to produce masker levels at threshold that were in the moderate (i.e., compressive) range. Signal levels were restricted to a range in which masker levels did not consistently exceed 90 dB SPL during the adaptive procedure for any subject. Different delays were used in an attempt to produce masker levels at threshold that covered overlapping ranges for the on- and off-frequency conditions. For the 0.25 kHz signal condition, signal levels in the off-frequency conditions were fixed at 45, 50, and 55 dB SPL; signal levels in the on-frequency conditions were fixed at 50, 55, and 57 dB SPL. All signal delays were 10 ms except for the 50 dB on-frequency condition in which signal delay was 5 ms. Signal levels in the off-frequency, 4 kHz signal condition were fixed at 30, 40, and 50 dB SPL; in the on-frequency condition, signal levels were fixed at 30, 40, and 45 dB SPL. All signal delays were 5 ms, except for the 50 dB off-frequency condition in which signal delay was 10 ms. Thresholds also were obtained for each signal and masker in quiet.

C. Adaptive-procedure thresholds

Thresholds were obtained in a VM paradigm using the same two-track procedure used in experiment 1 with the exception that the step size was initially 4 dB until after the fourth reversal, and then step size was 2 dB. Six to eight 200-trial repetitions were obtained in each condition. The last six repetitions were included in the analyses. There were 1200 trials, or 600 trials per track, available for further analysis. Mean adaptive thresholds were calculated as described in experiment 1.

D. Psychometric-function fits

All trials from both tracks were combined to fit PFs. PF threshold and slope were estimated as in experiment 1. Because VM PFs are "backwards" from VS PFs, the software routine and the rules used to fit PFs could not handle the data as extracted from the original data files. To obtain slopes following the VS procedure as closely as possible, masker levels were transformed by subtracting each from 100 before the PFs were fitted. There was a total of 72 PF fits, including the quiet-threshold conditions. The correlation between the masked adaptive and PF thresholds (excluding quiet-threshold conditions) across signal frequency, signal level, masker frequency, and subject was 0.988.

E. Results and discussion

1. Adaptive-procedure thresholds

Figure 5 shows mean threshold as a function of signal dB SPL in the left panel and as a function of dB SL (with regard to mean signal threshold in quiet) in the right panel. Thresholds increase as a function of signal level, and the level of the masker at threshold is higher in the off-frequency masker condition than in the on-frequency masker condition for both signal frequencies. For the 4 kHz signal condition, the rate of growth, estimated using slopes of linear regression line fits, is shallower for the off-frequency masker (0.78 dB/dB) than for the on-frequency masker (1.81 dB/dB). This

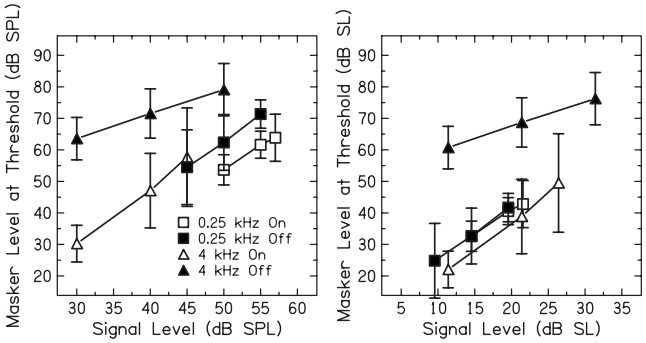


FIG. 5. Mean masker level at threshold across listeners with normal hearing (LNH) as a function of signal dB SPL (left panel) and dB SL (with regard to mean signal threshold in quiet, right panel) for 0.25 and 4 kHz signal (squares and triangles, respectively), on- and off-frequency masker conditions (open and filled, respectively) in experiment 2. The error bars represent +/−1 standard deviation. The forward masker duration was 200 ms and the masker levels were varied to estimate the level required to just mask the fixed signal. Signal duration was 10 ms and the signal delay was 5 or 10 ms. The rate of growth is shallower for the off-frequency than for the on-frequency case for the 4 kHz signal condition. Growth for the on- and off-frequency, 0.25 kHz signal conditions was similar to each other and to the on-frequency, 4 kHz signal condition.

difference in the slope of growth of maskability (Nelson *et al.*, 2001), which is another term for GOM functions obtained using a VM paradigm, can be used to estimate cochlear compression. If the off-frequency condition is assumed to be a linear reference, the slope ratio indicates on-frequency compression by a factor of 2.32. For the 0.25 kHz condition, however, the slopes of the off-frequency (1.69) and on-frequency (1.48) functions are similar. At first glance, this might suggest that there is little compression at 0.25 kHz, but both functions are similar to the on-frequency function at 4 kHz rather than to the off-frequency function. This is particularly clear in the right-hand panel. This suggests that the off-frequency function at 0.25 kHz cannot be used as a linear reference and that both functions reflect the effects of compression that is similar to the amount of compression observed at 4 kHz.

2. Psychometric-function fits

The interpretation of the thresholds obtained in on- and off-frequency conditions at 0.25 kHz is supported by the PFs. Figure 6 shows fits to the individual data points for the lowest and highest signal level conditions. The data appear to be more nonmonotonic in this experiment than in experiment 1, probably because the smaller step size (2 dB in this experiment and 4 dB in experiment 1) resulted in fewer trials per point. In general, the fits appear to be acceptable, with the exception of subject 231. It is unclear why this subject had a difficult time with this paradigm, considering all these subjects participated in VS conditions in experiment 3, where subject 231's fits appear to be more orderly (see Fig. 12). As in all regression analyses, poor fits result in shallower slopes. Table II provides a summary of the PF parameters and r^2 values for each subject in each condition.

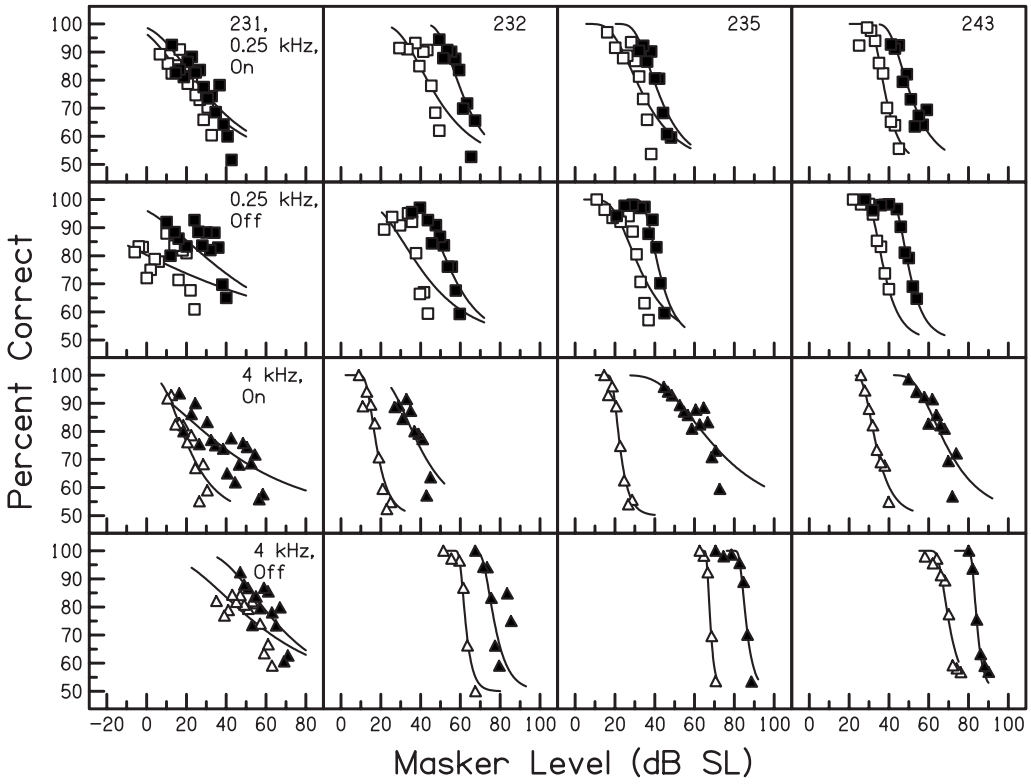


FIG. 6. Psychometric-function (PF) fits to the individual data points for the lowest (open) and highest (filled) signal level conditions, for on-frequency and off-frequency maskers, in 0.25 and 4 kHz variable masker (VM) conditions, for listeners with normal hearing (LNH) in experiment 2. The associated PF parameters (slope and threshold) and goodness of fit (r^2) are shown in Table II for each subject in each condition. Subject 231 stands out as having widely varying data and poor fits in comparison to the other subjects. This is not the case in the variable signal (VS) conditions of experiment 3, for which 231's data are more orderly (see Fig. 12).

TABLE II. Experiment 2 psychometric function (PF) parameters for listeners with normal hearing (LNH) in variable masker (VM), on- and off-frequency forward-masking conditions [QT=quiet threshold; SD =standard deviation].

Signal frequency	Masker frequency	Signal level	Parameter	231	232	235	243	Mean (SD)
0.25 kHz	None	QT	Threshold	38.455	22.228	36.208	34.228	32.78(7.24)
			Slope	0.433	0.132	0.283	1.124	0.37
			r^2	0.804	0.523	0.775	0.969	
None	0.15 kHz	QT	Threshold	33.969	18.314	33.143	26.020	27.86(7.30)
			Slope	0.433	0.446	0.166	1.094	0.43
			r^2	0.902	0.833	0.601	0.978	
None	0.25 kHz	QT	Threshold	25.334	10.617	23.834	18.979	19.69(6.63)
			Slope	0.916	0.908	0.286	0.290	0.51
			r^2	0.952	0.824	0.849	0.807	
0.25 kHz	0.25 kHz	50	Threshold	49.050	57.184	57.135	56.926	55.07(4.02)
			Slope	0.171	0.237	0.280	0.701	0.30
			r^2	0.756	0.623	0.724	0.970	
		55	Threshold	57.437	66.579	62.831	65.395	63.06(4.06)
			Slope	0.240	0.237	0.434	0.529	0.34
			r^2	0.595	0.634	0.839	0.940	
		57	Threshold	54.115	71.138	65.431	69.298	65.00(7.63)
			Slope	0.172	0.361	0.365	0.414	0.31
			r^2	0.711	0.778	0.838	0.867	
0.25 kHz	0.15 kHz	45	Threshold	46.564	57.730	65.279	63.241	58.20(8.39)
			Slope	0.064	0.200	0.311	0.647	0.23
			r^2	0.190	0.559	0.802	0.982	
		50	Threshold	58.510	65.224	70.962	70.927	66.41(5.91)
			Slope	0.112	0.213	0.183	0.406	0.21
			r^2	0.524	0.704	0.580	0.916	
		55	Threshold	30.383	72.707	74.951	75.929	73.30(2.80)
			Slope	0.111	0.314	0.597	0.672	0.34
			r^2	0.207	0.872	0.885	0.985	
4 kHz	None	QT	Threshold	17.031	18.076	20.662	16.233	18.00(1.93)
			Slope	1.366	0.389	0.879	1.053	0.84
			r^2	0.973	0.806	0.965	0.989	
None	2.4 kHz	QT	Threshold	7.046	6.594	−0.491	−3.996	2.29(5.43)
			Slope	1.098	0.743	0.515	0.747	0.75
			r^2	0.983	0.928	0.943	0.977	
None	4 kHz	QT	Threshold	3.622	11.118	9.444	4.202	7.10(3.75)
			Slope	0.963	0.306	0.386	1.154	0.60
			r^2	0.921	0.795	0.728	0.993	
4 kHz	4 kHz	30	Threshold	23.725	28.716	31.739	37.526	30.43(5.77)
			Slope	0.335	0.838	1.141	0.643	0.67
			r^2	0.839	0.928	0.972	0.965	
		40	Threshold	32.955	44.780	57.980	59.299	48.75(12.41)
			Slope	0.089	0.378	0.225	0.377	0.23
			r^2	0.620	0.883	0.850	0.954	
		45	Threshold	39.055	49.613	77.917	72.308	59.72(18.43)
			Slope	0.111	0.287	0.161	0.277	0.19
			r^2	0.644	0.706	0.722	0.768	
4 kHz	2.4 kHz	30	Threshold	58.890	68.980	67.294	65.810	65.24(4.43)
			Slope	0.117	1.451	2.364	0.817	0.76
			r^2	0.465	0.974	0.992	0.942	
		40	Threshold	66.719	74.881	77.347	74.687	73.41(4.62)
			Slope	0.165	1.187	2.357	0.648	0.74
			r^2	0.654	0.973	0.986	0.876	
		50	Threshold	70.056	83.310	85.199	80.024	79.65(6.74)
			Slope	0.163	0.743	1.309	1.630	0.71
			r^2	0.572	0.192	0.949	0.985	

Figure 7 shows PF slope as a function of PF threshold for each subject and for the geometric mean across subjects. If the PF slope in a VM paradigm reflects compression of the masker at the signal place, we would expect steep slopes in the off-frequency conditions regardless of the masker level and shallower slopes at high masker levels in the on-

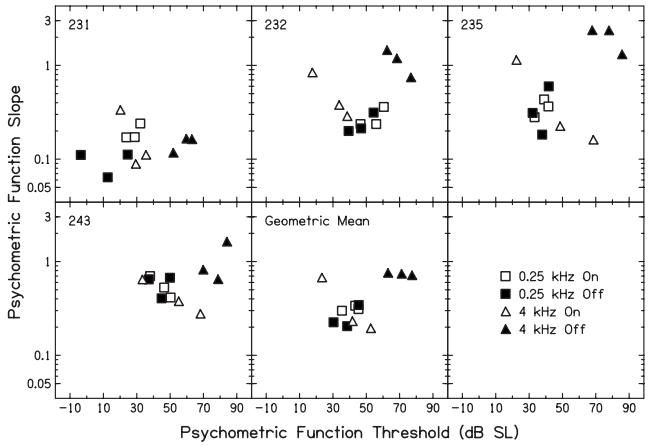


FIG. 7. Psychometric-function (PF) slope as a function of PF threshold for each subject and the geometric mean across subjects in experiment 2. Signal frequencies are represented as in Fig. 5. The trend for shallow slopes in all conditions for 231 is evident in this plot. On average the PF slopes in the off-frequency masker, 4 kHz signal conditions are steeper than the other conditions, including the off-frequency, 0.25 kHz signal conditions. On average, these results are consistent with comparable compression, but a wider bandwidth of compression at 0.25 in comparison to 4 kHz.

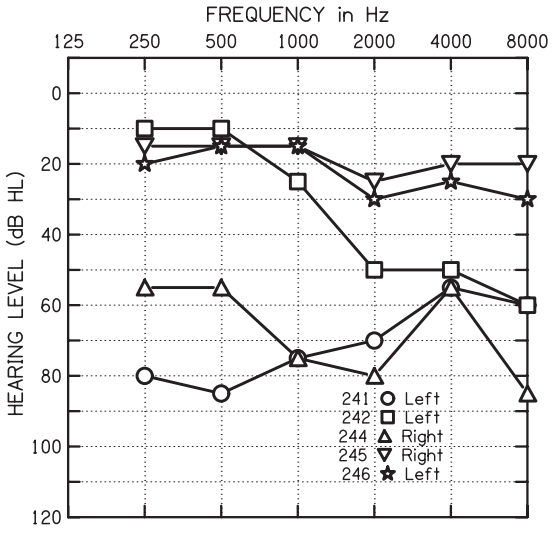


FIG. 8. Audiometric hearing thresholds for listeners with hearing impairment for experiment 3. The better ear is shown in this plot for each subject, and it is also the test ear in the experiment.

frequency conditions. With the exception of subject 231, all of the subjects had steeper PFs in the off-frequency condition at 4 kHz than in the on-frequency conditions, and slopes in the on-frequency condition became shallower as the masker level increased. The pattern is clear in the mean data, where the off-frequency PF slopes for the 4 kHz signal conditions remain constant at a value that is observed for the on-frequency conditions only at the lowest level. For the 0.25 kHz signal conditions, however, PF slopes are shallower for both the off-frequency and on-frequency masker conditions. The overlap of these slopes with those observed in on-frequency masker, 4 kHz signal conditions suggests compression in both the off-frequency and on-frequency conditions at the lower frequency. On average, these results are consistent with comparable compression with a wider bandwidth of compression at 0.25 in comparison to 4 kHz.

IV. EXPERIMENT 3: PSYCHOMETRIC FUNCTIONS FOR FORWARD-MASKED, LOW- AND HIGH-FREQUENCY TONES IN A VARIABLE-SIGNAL PARADIGM IN LISTENERS WITH NORMAL HEARING AND WITH HEARING LOSS

The purpose of experiment 3 was to test the hypothesis that PF slopes reflect reduced compression in LHL in comparison to LNH. As in experiment 1, the purpose is to compare relative compression between frequencies and groups of LNH and LHL, and not to estimate specific parameter or compression values. Off-frequency maskers were included in this experiment to provide a more complete set of data. Although the off-frequency masker, 4 kHz signal condition was used in Schairer *et al.* (2003b), the off-frequency masker, 0.25 kHz signal condition has yet to be reported in this context. In addition, results in LHL using this PF slope method have not been reported for any of the conditions, and thus, it was deemed appropriate to collect data in the full complement of conditions (low- and high-signal frequency conditions and on- and off-masker conditions) for comparison with the LNH. It was predicted that (1) in LNH, PF slopes will decrease as a function of threshold for both on- and off-frequency masker conditions, for both 0.25 and 4 kHz signal conditions, further supporting the hypothesis that there is comparable cochlear compression at these two frequencies, and (2) PF slopes in LHL will be steeper at frequencies with HL than for the LNH, and they will decrease less (or not at all) as a function of threshold.

A. Subjects

The group of LNH included the same subjects as in experiment 2. The group of LHL included four females and one male, ages 24–43 years (mean=31.2, SD=7.5). One other LHL was enrolled in the study but did not complete the data collection. Her data were not included. Audiometric hearing thresholds in the test ear for the group of LHL are shown in Fig. 8. In the experiment, left ears were tested in the group of LNH, and the better ear was tested in the group of LHL (as specified in Fig. 8). All five LHL were tested in the 0.25 and 4 kHz signal conditions, with the exception of subject 241, who was not tested in the 0.25 kHz signal conditions due to the degree of HL at that frequency.

B. Otoacoustic emission stimuli and apparatus

Distortion-product otoacoustic emission (DPOAE) input-output (I/O) functions at f_2=4 kHz were obtained in all test ears in order to demonstrate the presence of cochlear nonlinearity in the LNH and decreased or absent cochlear nonlinearity in the LHL. DPOAEs were not collected at f_2 =0.25 kHz because biological noise obscures the responses at low frequencies. Data were obtained as described in Schairer *et al.* (2003a) using a double-evoked technique (Keefe, 1998). DPOAEs are elicited by presenting two tones or primaries, and recording the emission at the distortion product frequency of $2f_1-f_2$. In the double-evoked technique, the SPL in the ear canal is recorded across three intervals: one primary presented alone ($p1$), the second pri-

mary presented alone ($p2$), and both primaries presented together ($p12$). The DPOAE is calculated as $(p1+p2)-p12$. In this manner, the linear distortion of the system is presumably canceled and the residual is the nonlinear emission.

DPOAEs were elicited with an f_2/f_1 of 1.21. L_2 levels for the DPOAE conditions were presented in descending 5 dB steps from 85 dB SPL down to 0 dB SPL. For L_2 levels of 65 dB SPL and above, $L_1=L_2$. At each L_2 below 65 dB SPL, $L_1=0.4L_2+39$ dB SPL, as proposed by Kummer *et al.* (1998). The current data were compared to the 25th and 75th percentile values from 15 left ears with NH from the study of Schairer *et al.* (2003a).

C. Forward masking stimuli and apparatus

The stimuli were delivered through the same equipment as experiments 1 and 2. Signal and masker frequencies were the same as for experiment 2, and a VS instead of a VM paradigm was used. Signal delays were 10 ms, as in experiment 1. Thresholds in quiet were obtained for each signal and masker (thresholds for LNH were presented in Table II as part of the experiment 2 results). In the masked conditions, masker levels were fixed at 50, 70, or 90 dB SPL for both on- and off-frequency conditions for the LNH. An additional set of conditions was added in order to obtain masked thresholds in the group of LNH that were similar to the highest masked thresholds of the listeners with the greatest HL. In this set of conditions, on- and off-frequency maskers were 90 dB SPL, for both signal frequencies, but the signal delay was shortened to 5 ms. Schairer *et al.* (2003b) did not find a significant independent effect of signal delay on PF slopes. Thus combining the 10 and 5 ms delay conditions for the current purposes was deemed appropriate.

For the LHL, in general, masker levels were selected individually based on thresholds in quiet for the maskers. The lowest masker level was selected such that it was estimated to produce at least 5–10 dB of masking (based on a comparison of masked thresholds after one to two practice repetitions with the average quiet threshold of the signal). The highest masker level was selected such that it would not require a signal level to exceed 90 dB SPL during the adaptive procedure. The exceptions were as follows. For listener 244 in two conditions, and listeners 241 and 242 in one condition each, the starting signal level was 94 dB SPL. This is because we attempted to use starting levels that were 14–20 dB above the estimated threshold for all conditions, and levels above 90 dB SPL were required to meet that target in these cases. The goal was to obtain data for three masker levels for each masker-frequency/signal-frequency combination. This was not possible for listeners 244 and 245 due to time and dynamic range constraints. In some cases, the same masker levels were used with the LHL that were used with the LNH. This occurred when a LHL had a normal or near-normal threshold at the signal frequency. Listener 246 was tested at the same masker levels for both signal frequencies as the group of LNH, and listener 242 was tested at the same masker levels as the group of LNH for the 0.25 kHz signal condition.

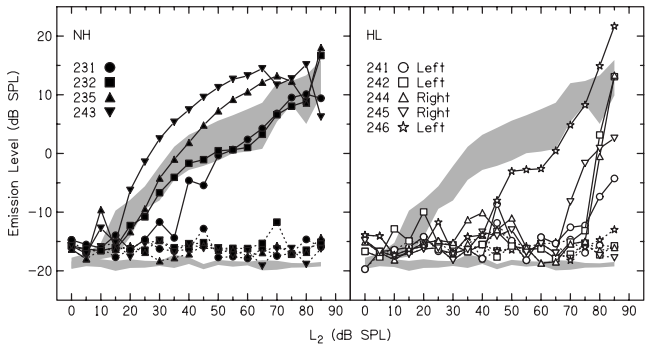


FIG. 9. Distortion-product otoacoustic emission (DPOAE) input-output (I/O) functions with $f_2=4$ kHz for ears with normal hearing (NH) (left panel) and ears with hearing loss (HL) (right panel) for experiment 3. The symbols represent individual subjects. The symbols connected by solid lines represent the DPOAE levels, and the symbols connected by dashed lines represent the noise levels. The shaded areas represent the 25th to 75th percentile of responses from left ears with NH from Schairer *et al.* (2003a). DPOAEs were not obtained for $f_2=0.25$ kHz (the other signal frequency in the behavioral tests) because biological noise at low frequencies obscures the responses. The results suggest normal cochlear nonlinearity in ears with NH and reduced or absent nonlinearity in ears with HL.

D. Adaptive-procedure thresholds

Thresholds were obtained in a VS paradigm using a two-track 2IFC adaptive procedure with decision rules to estimate 71 PC on one track and 87 PC on the other track (Levitt, 1971). The initial and final step sizes (after four reversals) were 4 and 2 dB, respectively. The threshold for each track was calculated as the mean of the reversal levels after the fourth reversal. Eight 200-trial repetitions were obtained in each condition. The first two repetitions were excluded as practice. A total of 1200 trials, or 600 trials per track, were available for further analysis. The exception was for subject 245 for whom only six repetitions of each condition were obtained due to time constraints. Mean thresholds were calculated across repetitions for each track, and then the mean across the two tracks was calculated as the adaptive-procedure threshold for each subject in each condition. Mean thresholds across LNH were calculated for comparison with individual LHL.

E. Psychometric-function fits

PFs were fitted as for experiments 1 and 2. All trials from both tracks were combined to fit the PFs. There was a total of 144 PFs to fit (64 for LNH, 80 for LHL, but two could not be fitted). This total includes the quiet thresholds for the LHL but excludes the quiet thresholds for the LNH because they were already presented in experiment 2. The correlation between the masked adaptive and PF thresholds (not including quiet thresholds) across signal frequency, masker level, masker frequency, and subject was 0.995.

F. Results

1. DPOAE I/O functions

Figure 9 shows the DPOAE I/O functions obtained in individual ears with NH (left panel) and with HL (right panel) compared to the 25th to 75th percentile responses (shaded areas) recorded in a group of left ears with NH

(Schairer *et al.*, 2003a). The responses in ears with NH in the current study exceed the 25th percentile of the normal range with the exception of listener 231 at L_2's below 50 dB SPL. The responses from ears with HL do not reach the 25th percentile of the normal range except at the highest L_2 levels. This suggests normal cochlear nonlinearity for the ears with NH and decreased or absent nonlinearity in ears with HL. Note that despite the fact that subjects 245 and 246 have near-normal audiometric thresholds at 4 kHz (Fig. 8), their DPOAE I/O functions suggest loss of OHC function and, presumably, loss of cochlear nonlinearity. Further, the thresholds in quiet for the 4 kHz signal obtained in the laboratory were higher for subjects 245 and 246 (39.5 and 40.6 dB SPL, respectively) than for the LNH (16.3–21.0 dB SPL) For these reasons, it was deemed appropriate to include them in the group of LHL for the 4 kHz signal conditions.

DPOAEs were not obtained in the f_2=0.25 kHz case because biological and environmental noise make it difficult to measure reliable robust responses in a reasonable amount of time. Thus, DPOAEs cannot be used to compare the degree of cochlear nonlinearity between the two subject groups at that signal frequency. Subjects 242, 245, and 246 (of the group of LHL) had audiometric thresholds within normal limits at 0.25 kHz (Fig. 8) despite having HL (or borderline HL) at 4 kHz. They also had thresholds for the 0.25 kHz signal in quiet (39.7, 33.3, and 35.9 dB SPL) that were comparable to the LNH (in the range 30.9–39.4 dB SPL). These three subjects in the group of LHL were therefore considered to have NH at 0.25 kHz. They could in a sense serve as their own controls (NH at 0.25 kHz and HL at 4 kHz). It was demonstrated in experiment 1 that in LNH, on-frequency masking produces similar PF slope results at 0.25 and 4 kHz (and presumably reflects cochlear nonlinearity). It follows that any difference in PF slope between the 0.25 and 4 kHz signal conditions in these three individuals with presumably normal function at 0.25 kHz and impaired function at 4 kHz is likely due to loss of cochlear nonlinearity.

2. Adaptive forward-masked thresholds

Figure 10 shows mean adaptive threshold across LNH as a function of masker dB SPL in the left panel and mean amount of masking as a function of masker dB SL in the right panel. All signal- and masker-frequency conditions are represented in each panel. Slopes of GOM (fit without the 5 ms duration condition) are similar for on-frequency (0.57) and off-frequency (0.66) conditions in the 0.25 kHz signal case and for the on-frequency (0.48) condition in the 4 kHz signal case. The slope is steeper in the off-frequency masker condition in the 4 kHz case (0.97) than for the other three conditions. This result is consistent with the data shown in Fig. 5 that were obtained using a VM paradigm.

Figure 11 shows the means across LNH (filled squares in all panels) along with the individual thresholds for each LHL (open symbols, as in Fig. 8), with different signal and masker frequencies represented in different rows. In the 4 kHz signal conditions, the masked thresholds for the LHL are generally higher than those for LNH (left panels), with the exception of subjects 245 and 246 in the highest masker-level condition. In the 0.25 kHz signal conditions, responses were not obtained in subject 241 due to degree of HL. Subjects 242, 245, and 246 had normal thresholds. Subject 244 was the only listener who had HL at 0.25 kHz but was still able to perform the task over a range of masker levels. Slopes of GOM for the LHL are similar in most cases to those for the LNH (see Table III for specific values).

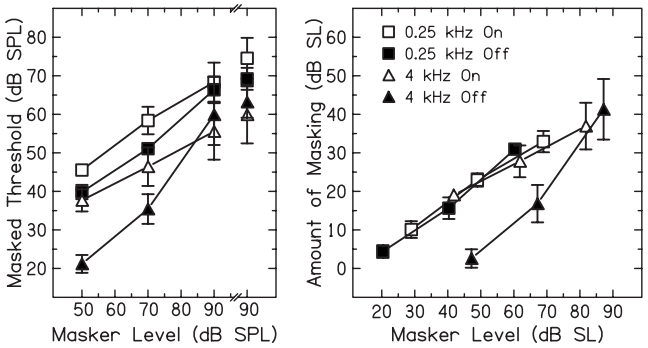


FIG. 10. Mean masked threshold across listeners with normal hearing (LNH) as a function of masker dB SPL (left panel) and mean amount of masking as a function of masker dB SL (with regard to mean masker thresholds in quiet; right panel) for 0.25 and 4 kHz signal conditions (squares and triangles, respectively) and on- and off-frequency masker conditions (open and filled, respectively) in experiment 3. The error bars represent +/−1 standard deviation. Masker duration was 200 ms with a 10 ms signal delay. Maskers were presented at 50, 70, and 90 dB SPL in different conditions. Another condition with a masker level of 90 dB SPL and a 5 ms signal delay (disconnected symbols in left panel) was also presented. Signal duration was 10 ms and signals were presented in a variable signal (VS) paradigm. The amount of masking increases as a function of masker level at a similar rate for all but the 4 kHz, off-frequency condition (see Table III for slopes).

3. Psychometric-function fits

The threshold, slope, and r^2 values for each condition for LNH and LHL are shown in Tables IV and V, respectively. Examples of PF fits to individual data points for the lowest and highest masker-level conditions for LNH and LHL are shown in Figs. 12 and 13, respectively. Figure 14 shows the PF slope as a function of the PF threshold for the individual LNH along with the mean results across subjects (geometric mean slope as a function of arithmetic mean PF threshold). Because the PFs in a VS paradigm provide a measure of compression at the signal frequency, no difference is expected between PF slopes in on-frequency and off-frequency conditions. If there were markedly less compression at 0.25 kHz, slopes would not be expected to decline with increasing threshold in those conditions. Although the slope estimates vary widely for individual subjects, in general, slopes decrease as a function of threshold and the decrease is similar for on-frequency and off-frequency maskers and for 0.25 and 4 kHz signal conditions. As in Fig. 4, the function fitted to the mean data in Fig. 14 is the reciprocal of a quadratic compression function, as described by Neely and Jesteadt (2005).

The mean data points and line fitted to the mean data from Fig. 14 are replotted in Fig. 15, along with the individual data for the LHL. Note that the scale is larger in Fig. 15 than in Fig. 14 to accommodate the larger slopes in the LHL. In the 4 kHz signal conditions, although there appears to be some decrease in PF slope as a function of threshold,

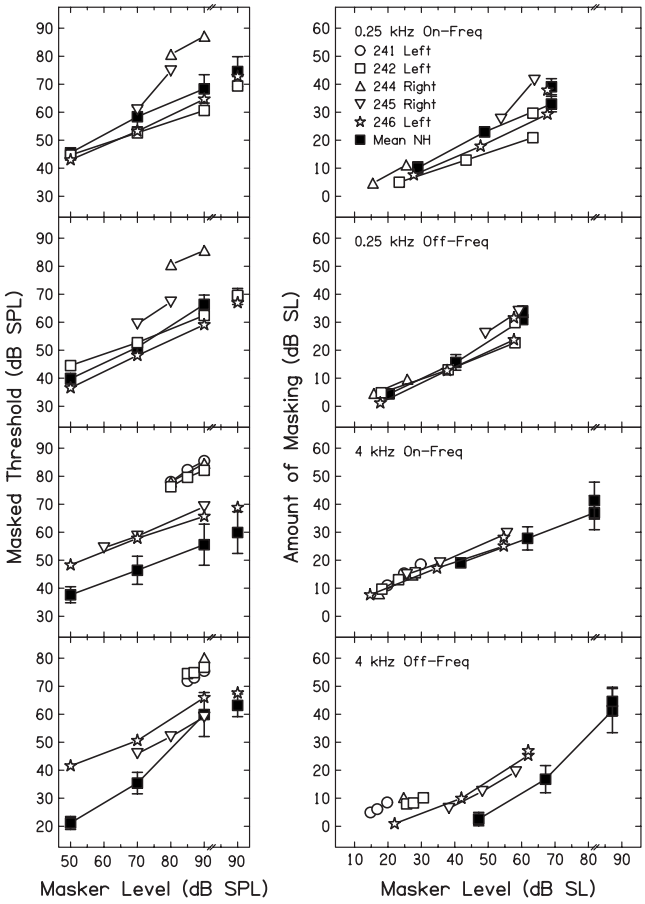


FIG. 11. Mean data across listeners with normal hearing (LNH; mean NH) from Fig. 10 are reproduced here (filled squares) along with individual data from listeners with hearing loss (LHL; symbols as in Fig. 8) from experiment 3. Signal and masker-frequency conditions are represented in rows, with adaptive threshold as a function of masker dB SPL in the left column and amount of masking as a function masker dB SL in the right column. Although it appears that more masking was produced in general in LNH (left panel), when masker levels are expressed in dB SL (with regard to masker thresholds in quiet), there is actually a comparable amount of masking across subjects except in the 4 kHz signal, off-frequency masker condition.

the slopes are steeper in all cases (except for one data point for subject 245 in the on-frequency masker case) than the fit to the data from the LNH. In the 0.25 kHz signal conditions, subject 244, who has HL at 0.25 kHz, has steeper slopes than the LNH except for one on-frequency condition. Subjects 242, 245, and 246 served as their own controls because they have NH at 0.25 kHz and impaired hearing at 4 kHz. Subject 246 has slightly steeper PF slopes in comparison to LNH in the 0.25 kHz signal conditions. In general, subjects 242 and 245 have PF slopes that are comparable or lower than LNH in the 0.25 kHz signal conditions. The very shallow slopes in these conditions may be due to poor goodness of PF fits rather than perceptual variability. For subject 242, the two conditions that have slopes that are out of line with the rest of the data have associated r^2 values of 0.597 and 0.443 (see Table V for the 50 and 70 dB off-frequency masker, 0.25 kHz signal conditions). The r^2 values for all other masked conditions are >0.60 for these two subjects and the other subjects. For subject 245, the r^2 for the 70 dB on-frequency masker condition was 0.453. As stated in experiment 2, poor PF fits can be associated with PF slopes that are shallower than expected.

In summary, the on-frequency results in LNH are in agreement with experiment 1. PF slopes decreased as a function of threshold for both on- and off-frequency masker conditions, for both 0.25 and 4 kHz signal conditions. In addition, preliminary results in LHL demonstrated that PF slopes were steeper at frequencies with HL. The PF slopes demonstrated differences in compression between the groups that were not obvious in the GOM functions, because slopes of GOM were similar between the LNH and the LHL (see Fig. 11 and Table III).

V. DISCUSSION

A. Summary of experiments

The goals of the current set of experiments were to use trends in PF slopes across conditions to compare compression and bandwidth of compression at low- and high-signal frequencies, and to compare degree of compression in LNH and LHL. For LNH, PF slopes decreased as a function of signal threshold for VS, 4 kHz, on-frequency conditions (experiments 1 and 3) and at a similar rate for off-frequency conditions (experiment 3), consistent with Schairer *et al.* (2003b). In addition, PF slopes decreased as a function of signal threshold for VM, 4 kHz, on-frequency but not off-frequency conditions (experiment 2), also consistent with Schairer *et al.* (2003b). These results suggest that PF slopes reflect the compressive nonlinearity at the place of the signal.

The current results extend this conclusion to the 0.25 kHz signal condition and suggest comparable compression in the 0.25 and 4 kHz conditions (experiments 1 and 3). The finding of comparable compression at the low- and high-signal frequencies is consistent with recent behavioral data (see, e.g., Lopez-Poveda and Alves-Pinto, 2008; Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; and Williams and Bacon, 2005) and DPOAE data in LNH (Gorga *et al.*, 2007). Gorga *et al.* (2007) reported that although the I/O functions

TABLE III. Slope of linear regression line fits to growth of amount of masking as a function of masker dB SL for experiment 3: Means across listeners with normal hearing (NH) and individual listeners with hearing impairment.

Signal frequency	Masker frequency	241	242	244	245	246	NH
0.25 kHz	On		0.40	0.65	1.39	0.54	0.57
0.25 kHz	Off		0.45	0.51	0.78	0.56	0.66
4 kHz	On	0.75	0.58	0.67	0.50	0.44	0.48
4 kHz	Off	0.72	0.45		0.65	0.61	0.97

TABLE IV. Experiment 3 psychometirc function (PF) parameters for listeners with normal hearing (LNH) in variable signal (VS), on- and off-frequency forward-masking conditions [SD=standard deviation].

Signal frequency	Masker frequency	Masker level	Parameter	231	232	235	243	Mean (SD)
0.25 kHz	0.25 kHz	50	Threshold	45.535	43.274	45.436	45.288	44.88(1.08)
			Slope	0.510	0.576	0.451	1.469	0.66
			r^2	0.717	0.877	0.730	0.980	
		70	Threshold	58.840	51.398	61.243	56.340	56.96(4.21)
			Slope	0.379	0.444	0.360	0.734	0.46
			r^2	0.806	0.775	0.940	0.913	
		90	Threshold	69.756	63.140	72.836	62.723	67.11(4.99)
			Slope	0.368	0.312	0.231	0.712	0.37
			r^2	0.785	0.860	0.862	0.958	
		90, 5 ms delay	Threshold	78.031	70.993	77.174	66.855	73.26(5.30)
			Slope	0.313	0.398	0.425	0.484	0.40
			r^2	0.725	0.851	0.937	0.956	
0.25 kHz	0.15 kHz	50	Threshold	41.198	37.269	37.856	38.891	38.80(1.73)
			Slope	0.851	0.476	0.171	1.112	0.53
			r^2	0.823	0.759	0.537	0.958	
		70	Threshold	50.665	49.148	50.133	49.493	49.86(0.68)
			Slope	0.458	0.333	0.466	0.753	0.48
			r^2	0.938	0.824	0.951	0.960	
		90	Threshold	65.966	62.810	66.138	62.388	64.33(2.00)
			Slope	0.189	0.380	0.419	0.407	0.33
			r^2	0.612	0.760	0.533	0.832	
		90, 5 ms delay	Threshold	72.137	67.238	69.105	65.809	68.57(2.73)
			Slope	0.567	0.544	0.388	0.707	0.54
			r^2	0.904	0.907	0.910	0.909	
4 kHz	4 kHz	50	Threshold	34.399	38.942	39.050	33.846	36.56(2.82)
			Slope	0.266	0.422	0.748	1.179	0.56
			r^2	0.735	0.916	0.927	0.982	
		70	Threshold	44.047	49.478	43.455	39.533	44.13(4.09)
			Slope	0.167	0.263	0.369	0.628	0.32
			r^2	0.650	0.767	0.749	0.936	
		90	Threshold	53.778	59.283	46.405	46.781	51.56(6.16)
			Slope	0.140	0.186	0.160	0.829	0.24
			r^2	0.676	0.826	0.593	0.918	
		90, 5 ms delay	Threshold	53.392	67.762	51.891	55.337	57.10(7.25)
			Slope	0.152	0.243	0.554	0.337	0.29
			r^2	0.663	0.719	0.905	0.946	
4 kHz	2.4 kHz	50	Threshold	22.003	17.794	21.254	18.162	19.81(2.14)
			Slope	0.534	0.717	1.384	1.154	0.88
			r^2	0.796	0.937	0.986	0.947	
		70	Threshold	39.370	31.387	32.362	34.533	34.41(3.56)
			Slope	0.504	1.196	1.268	0.288	0.69
			r^2	0.925	0.982	0.990	0.496	
		90	Threshold	67.912	59.204	51.551	55.297	58.49(7.02)
			Slope	0.198	0.165	0.911	0.446	0.34
			r^2	0.673	0.547	0.969	0.952	
		90, 5 ms delay	Threshold	64.035	63.494	56.626	63.266	61.86(3.50)
			Slope	0.405	0.236	0.525	0.318	0.36
			r^2	0.899	0.678	0.840	0.915	

differed across f_2 frequencies of 0.5 and 4 kHz overall, the maximum compression ratios were comparable (approximately 4:1). The fact that there is evidence from behavioral and nonbehavioral studies strengthens the argument that compression is comparable at low and high frequencies.

The fact that PF slopes in the 0.25 kHz signal, on- and off-frequency masker conditions decreased at a similar rate in the VM conditions (experiment 2) provides support for the conclusion of Lopez-Poveda *et al.* (2003) and Plack and Drga (2003) that the bandwidth of compression is wider at lower frequencies. PF slopes provide a uniform measure of compression that can be used in all conditions, allowing a direct assessment of the range of frequencies around CF that have compressive response growth at CF. Another benefit of the current approach is that it does not require assumptions about the rate of recovery from forward masking for either on-frequency or off-frequency maskers, or across signal frequency, thus avoiding issues raised by Stainsby and Moore

TABLE V. Experiment 3 psychometric function (PF) parameters for listeners with hearing loss (LHL) in variable signal (VS), on- and off-frequency forward-masking conditions [QT=quiet threshold].

Signal frequency	Masker frequency	Masker level	Parameter	241	242	244	245	246
0.25 kHz	None	QT	Threshold		39.260	74.781	32.871	35.308
			Slope		0.633	0.438	0.641	1.182
			r^2		0.953	0.762	0.969	0.948
None	0.15 kHz	QT	Threshold		31.222	64.298	20.276	29.507
			Slope		0.447	1.099	0.677	0.198
			r^2		0.980	0.983	0.894	0.634
None	0.25 kHz	QT	Threshold		25.516	64.491	15.666	19.392
			Slope		0.445	0.931	0.833	0.290
			r^2		0.861	0.989	0.980	0.841
0.25 kHz	0.25 kHz	50	Threshold		44.062			42.840
			Slope		0.733			1.127
			r^2		0.939			0.915
		70	Threshold		52.654		57.881	52.987
			Slope		1.314		0.145	1.199
			r^2		0.950		0.450	0.936
		80	Threshold			80.594	71.934	
			Slope			2.010	0.253	
			r^2			0.975	0.661	
		90	Threshold		59.595	85.297		64.055
			Slope		0.449	0.495		0.551
			r^2		0.814	0.717		0.883
		90, 5 ms delay	Threshold		68.462			72.055
			Slope		0.439			0.683
			r^2		0.928			0.958
0.25 kHz	0.15 kHz	50	Threshold		42.448			36.197
			Slope		0.227			0.836
			r^2		0.597			0.976
		70	Threshold		47.479		56.340	48.034
			Slope		0.159		0.220	0.850
			r^2		0.443		0.691	0.894
		80	Threshold			80.780	60.782	
			Slope			1.529	0.214	
			r^2			0.997	0.811	
		90	Threshold		62.058	85.728		58.338
			Slope		0.397	1.399		0.594
			r^2		0.917	0.972		0.947
		90, 5 ms delay	Threshold		68.603			66.600
			Slope		0.315			0.531
			r^2		0.885			0.911
4 kHz	None	QT	Threshold	65.300	66.782	70.216	39.651	40.030
			Slope	0.392	1.235	1.415	1.798	0.614
			r^2	0.468	0.985	0.976	0.980	0.738
None	2.4 kHz	QT	Threshold	70.295	59.353	65.495	31.104	27.903
			Slope	1.022	0.965	1.117	1.225	1.139
			r^2	0.993	0.882	0.979	0.884	0.989
None	4 kHz	QT	Threshold	60.418	61.971	62.756	34.148	35.203
			Slope	1.577	1.521	0.913	1.085	1.597
			r^2	0.997	0.988	0.956	0.938	0.953
4 kHz	4 kHz	50	Threshold					CNF
			Slope					
			r^2					
		60	Threshold				54.457	
			Slope				2.192	
			r^2				0.987	
		70	Threshold				58.997	57.986
			Slope				2.508	1.775
			r^2				0.977	0.992
		80	Threshold	78.829	76.619	78.017		
			Slope	3.723	3.553	2.394		

TABLE V. (*Continued.*)

Signal frequency	Masker frequency	Masker level	Parameter	241	242	244	245	246
			r^2	0.947	0.987	0.926		
		85	Threshold	82.369	79.662			
			Slope	2.341	1.902			
			r^2	0.988	0.967			
		90	Threshold	85.160	81.975	85.247	63.492	64.927
			Slope	1.415	1.131	4.857	0.216	0.830
			r^2	0.956	0.907	0.971	0.778	0.912
		90, 5 ms delay	Threshold					69.071
			Slope					1.246
			r^2					0.980
4 kHz	2.4 kHz	50	Threshold					41.536
			Slope					1.372
			r^2					0.958
		70	Threshold				45.845	51.306
			Slope				1.925	1.632
			r^2				0.991	0.887
		80	Threshold				52.281	
			Slope				2.179	
			r^2				0.988	
		85	Threshold	71.887	74.564			
			Slope	2.024	1.292			
			r^2	0.916	0.971			
		87	Threshold	73.532	75.100			
			Slope	7.500	1.157			
			r^2	0.996	0.965			
		90	Threshold	75.478	CNF	80.166	58.894	65.578
			Slope	1.725		2.115	0.943	0.982
			r^2	0.818		0.964	0.964	0.922
		90, 5 ms delay	Threshold					67.214
			Slope					0.984
			r^2					0.982

(2006) and Wojtczak and Oxenham (2007). Finally, preliminary results in LHL suggest that the decrease in compression at frequencies with HL is reflected in steeper PF slopes. Because the PF slope can be measured in on-frequency conditions independent of the recovery function, this measure is less likely to be confounded by differences in frequency or temporal analysis between LNH and LHL.

B. Methodological issues

Schairer *et al.* (2003b) noted a number of methodological issues related to the procedures for fitting PFs and the assumed form of the PF that also apply to the current study. In both studies, PFs were reconstructed from adaptive-procedure data and fitted using procedures described by Dai (1995). The current study used two interleaved adaptive tracks with different decision rules, as suggested by Dai (1995), in an effort to sample the range of the PF more evenly. The use of two tracks also makes it possible to estimate the slope of the PF from the two adaptive thresholds by computing the slope of the line connecting the 71- and 87-PC points on the PF. We compared the two measures using the data obtained in experiment 1. Slope estimates obtained by connecting the two points were shallower on average than those obtained by fitting reconstructed PFs, but showed a similar pattern as a function of level. Both measures yielded one extreme slope estimate out of a total of 40 estimates (5 subjects $\times$ 8 conditions). In both cases, the estimate obtained with the other measure fell well within the range of slopes observed across subjects and conditions. Additional work will be required to simulate the two estimation approaches and to obtain estimates of the stability of the two measures. A simplified method for estimating the PF slope would facilitate use of slopes as a measure of compression.

Based on simulations, Dai (1995) recommended use of a 4 dB final step size when reconstructing PFs rather than the typical 2 dB value, because the larger step size would concentrate more trials on a smaller number of levels, resulting in better estimates of the proportions correct at those levels. It is worth noting that a 4 dB step size was used in experiment 1, but a 2 dB step size was used in the later experiments. PF fits were generally poorer in those experiments, consistent with Dai's (1995) simulations.

The off-frequency maskers were selected such that the ratios between the masker and the signal frequencies were identical for 0.25 and 4 kHz signals. However, it is possible that the off-frequency maskers were not low enough. Lopez-Poveda and Alves-Pinto (2008) found evidence of compression using a 2.2 kHz off-frequency masker in their 4 kHz signal conditions and suggested that studies that used $\geq$2.2 kHz tones as off-frequency maskers may not have

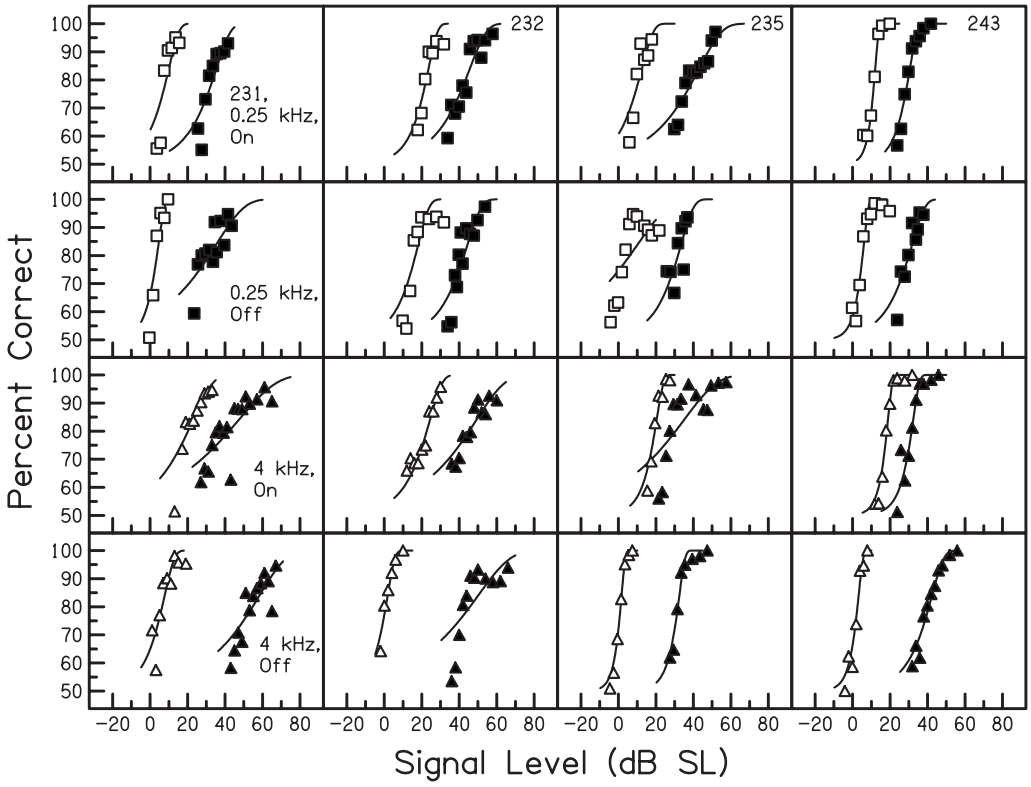


FIG. 12. Psychometric-function (PF) fits to the individual data points for the 50-(open) and 90-(filled) dB SPL on-frequency and off-frequency maskers in 0.25 and 4 kHz signal conditions for listeners with normal hearing (LNH) in experiment 3. The PF thresholds, slopes, and goodness of fit measures (r^2) for all conditions and subjects are listed in Table IV.

been purely linear reference conditions. Compression was not evident in their 1.6 kHz off-frequency masker conditions. The 4 kHz off-frequency masker conditions of the current experiment 2 are relevant to this issue. The theory is that in the VM conditions, PF slope should become shallow as a function of masker threshold for on-frequency masker conditions but remain steep in the off-frequency masker conditions. The on-frequency masker should be compressed be-

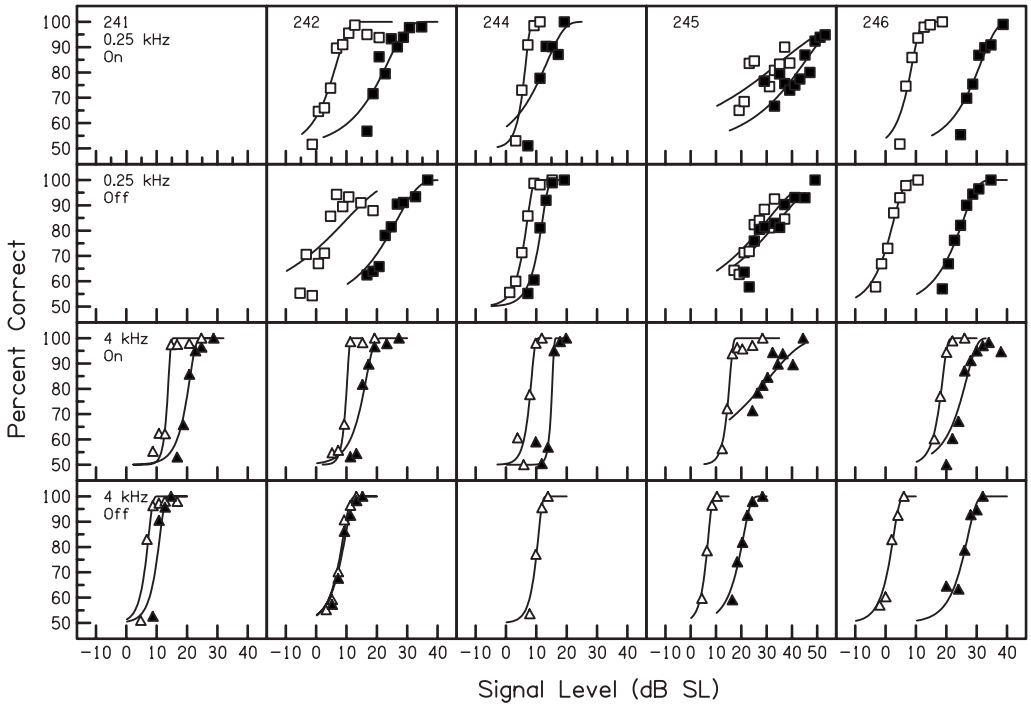


FIG. 13. Psychometric-function (PF) fits to individual data points for the lowest (open) and highest (filled) masker-level conditions (varied by subject, see Table V) for listeners with hearing loss (LHL) in experiment 3. Conditions are represented as in Fig. 12. The PF thresholds, slopes, and goodness of fit measures (r^2) for all conditions and subjects are listed in Table V.

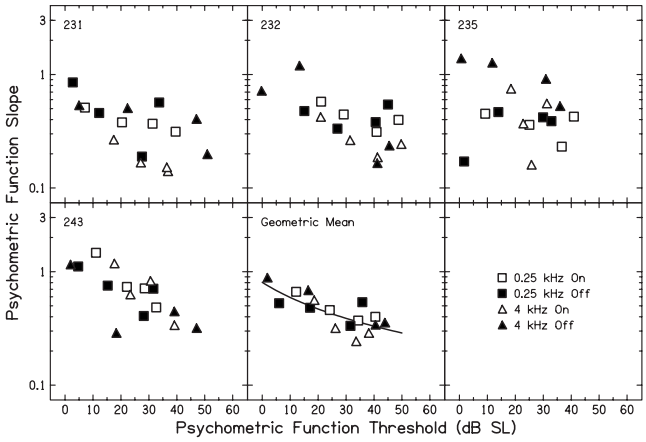


FIG. 14. Psychometric-function (PF) slope as a function of PF threshold for on-frequency (open symbols) and off-frequency (filled symbol) masker conditions for 0.25 (squares) and 4 kHz (triangles) signal conditions in listeners with normal hearing (LNH) in experiment 3. In general, slopes decrease as a function of threshold similarly for on- and off-frequency maskers, and for 0.25 and 4 kHz signal conditions. The function fitted to the geometric mean data is the reciprocal of the quadratic compression function described by Neely and Jesteadt (2005).

cause its traveling wave should peak at the place of signal but the off-frequency masker should not be compressed because its traveling wave peaks at a lower frequency, presumably out of the range of compression around the signal frequency. If compression had influenced the results in the off-frequency masker condition, the slopes should have become shallow as a function of masker level at threshold. In Fig. 7, there may be a slight decrease in slopes in subjects 232, 235,

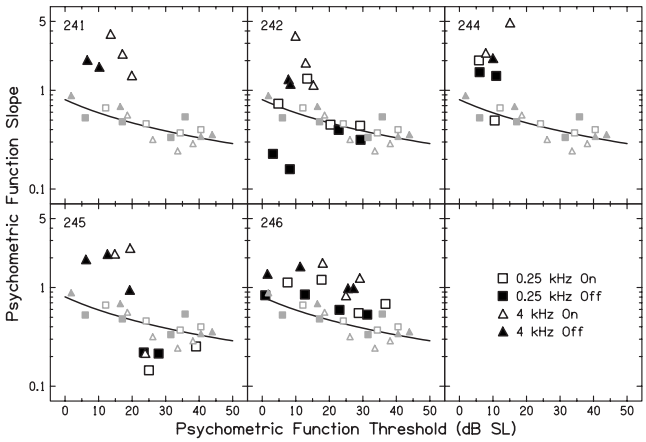


FIG. 15. Psychometric-function (PF) slope as a function of PF threshold for the individual listeners with hearing loss (LHL; dark symbols, as in Fig. 14) in experiment 3. The line fit to the geometric mean across listeners with normal hearing (LNH) is replotted from Fig. 14 along with the mean data points (small, gray symbols). Note that the axes are on a different scale than in Fig. 14 in order to accommodate the wider range of values in the LHL. In the 4 kHz signal conditions, although there appears to be some decrease in PF slope as a function of threshold, in general the slopes are steeper than the data from the LNH. In the 0.25 kHz signal conditions, subject 244, who has impaired hearing at 0.25 kHz, has steeper slopes than the LNH except for one on-frequency condition. Subjects 242, 245, and 246 have normal or near NH at 0.25 kHz (see Fig. 8). Subjects 242 and 245 have PF slopes that are comparable or lower than LNH in the 0.25 kHz conditions, although the shallowest slopes may be related to poorer PF fits in those conditions (see text for explanation and Table V for PF fit parameters). Subject 246 has slightly steeper PF slopes in comparison to LNH in the 0.25 kHz conditions.

and 243, but slopes in the 4 kHz off-frequency condition were still much steeper than in the remaining conditions. The data for subject 231 may agree with the results of Lopez-Poveda and Alves-Pinto (2008), although caution should be taken when making this comparison due to the relatively poorer r^2 values (i.e., PF fits) for this subject in several conditions. If the data are taken at face value regardless of PF fit, slopes for the 4 kHz signal, off-frequency masker condition were similar to the remaining conditions, suggesting comparable compression in the on- and off-frequency conditions and a wider range of compression around 4 kHz for this subject than for the other three subjects.

A 10 ms, 0.25 kHz signal was used in the current study in order to achieve the range of masker and signal thresholds that were necessary to test the hypotheses across experiments. In particular, a longer duration signal would have required excessive masker levels in experiment 2. Spectral splatter from the short-duration signal may have affected the results, but only for the off-frequency masker conditions for which there may have been unintended overlap in the spectra of the signal and the masker. This potential confounding factor would not affect the conclusions drawn from the PF slopes for experiments 1 and 3 for three reasons: (1) off-frequency maskers were not used in experiment 1, (2) the comparison of compression at 0.25 and 4 kHz requires only the on-frequency conditions, and (3) the comparison of compression between LHL and LNH requires only the on-frequency conditions. However, spectral splatter from the short-duration 0.25 kHz signal may confound conclusions drawn from the results of experiment 2. The off-frequency masker in the 0.25 kHz signal condition produced higher masker levels at threshold than the on-frequency masker (see left-hand panel of Fig. 5), which suggests enough separation in the off-frequency masker and signal spectra that a higher masker level was necessary to just mask the signal than in the on-frequency masker conditions. Overlapping spectra of the masker and signal may have influenced the trends in PF slopes. PF slopes for on- and off-frequency maskers were similar to each other and to the slopes for the 4 kHz signal, on-frequency condition. This may be due to a wider range of compression around the 0.25 kHz signal, to spectral splatter from the short-duration signal that overlapped the spectrum of the off-frequency masker, or to a combination of both.

C. Results for listeners with hearing loss

PF slopes in LHL were consistent with reduced compression at frequencies at which there was hearing impairment, and with compression that was comparable to LNH at frequencies at which hearing was normal. Having some frequencies of NH and some frequencies with impaired hearing allows subjects to serve as their own "controls." The need to measure PF slope over a range of threshold levels required LHL to have mild or moderate losses at the test frequencies. The use of an independent DPOAE measure of compression strengthened the interpretation of the data. It might be feasible to compare the PF slope and the DPOAE measures in a larger group of LHL tested only at high levels.

It should be noted that a slight decrease in PF slope as a

function of PF threshold was observed in LHL despite the fact that slopes remained steep in comparison to LNH. Given the mild and moderate HLs in this group, the decrease in PF slope may reflect varying degrees of residual compression. This is consistent with the VS results from Lopez-Poveda *et al.* (2005) in which nearly normal compression was observed in two ears with sensorineural HL, but reduced compression in a third ear with HL, and with VM results from Oxenham and Plack (1997) in which some residual compression was evident in two of their LHL.

In some cases, the steepness of the PF slope may be due to a reduced dynamic range, as a result of loss of amplification and compression mechanisms in the inner ear. A shallow PF slope implies a large range over which signal levels are detected, from just audible to 100% audible. Ears with HL have reduced ranges in comparison to ears with NH, and therefore PF slopes could be steeper simply due to the restricted range.

D. Cochlear nonlinearity as a function of level

In principle, PF slopes can be used to estimate the form of the function describing cochlear nonlinearity as a function of level. To accept one proposed nonlinearity function and reject the others, or to obtain reliable estimates of parameter values for individual subjects, would require a larger number of more stable PF slope estimates than were obtained in the present experiments. Geometric mean slopes across subjects in Figs. 4 and 14 were fitted with quadratic compression functions described by Neely and Jesteadt (2005). The parameters were estimated by iterative fits of the equation: $1/\text{slope}=a+bT+cT^2$, where T is the PF threshold, the signal level in dB SL required for $d'=1$, a specifies the compression at absolute threshold, and b describes the rate at which compression increases with level. The best fitting value of c was zero and it can be ignored for purposes of the current discussion. The functions account for the general form of the data and suggest that compression is comparable at the two signal frequencies, but the parameter values differ from those used by Neely and Jesteadt (2005) and differ from one set of data to the next. Neely and Jesteadt (2005) assumed $a=0.6$ and $b=0.1$. In Fig. 4, $a=0.7$ and $b=0.04$. In Fig. 14, $a=1.2$ and $b=0.04$. Differences among conditions in experiments 1 and 3 may have influenced the parameter estimates. The largest difference was the inclusion of off-frequency masker conditions in experiment 3. Fitting a quadratic compression function to the slopes obtained in only the on-frequency conditions yielded parameter estimates of $a=0.9$ and $b=0.06$. We have no reason to believe, however, that the relation between the PF slope and the signal level should vary for on- and off-frequency maskers.

Schairer *et al.* (2003b) noted that the PF slope appeared to be more highly correlated with the masker level than with the signal level. The high correlation observed between the masker level and the signal level made it difficult to separate the two effects, but partial correlation of the PF slope was much lower with the signal threshold (−0.06) than with the masker level (−0.71). In the current experiments 1 and 3, masker level and PF threshold were significantly correlated with each other and with PF slope. Because the range of signal delays was limited, it was not possible to separate the masker and signal level effects on the PF slope in the data reported here. The assumption that the signal level is the critical factor governing changes in the PF slope leads to the most straightforward interpretation of the data, but the relative contributions of the signal and the masker level have yet to be thoroughly explored.

VI. CONCLUSIONS

The trends in the PF slopes across conditions suggest comparable compression at 0.25 and 4 kHz, and potentially a wider bandwidth of compression in relative frequency at 0.25 kHz. This is consistent with other recent behavioral studies that have revised earlier estimates of less compression at lower frequencies. The preliminary results in LHL demonstrate that PF slopes are abnormally steep at frequencies with HL, but are similar to those for LNH at frequencies with NH. Overall, the results are consistent with the notion that PF slopes reflect degree of cochlear nonlinearity, and can be used as an additional measure of compression across frequency. More data are required in VM conditions in LNH to investigate bandwidth of compression and in VS conditions in LHL, particularly in the lower frequency signal condition.

ACKNOWLEDGMENTS

We thank Tom Creutz for creating the multitrack function of our data collection program, and Hongyang Tan for developing our data analysis program. We also thank Brian Moore and two anonymous reviewers for helpful comments on previous versions of the manuscript. Research was supported by NIH Grant Nos. R01 DC006648 and R03 DC006342. Subject recruitment was supported by NIH Grant No. P30 DC04662.

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.

Chatterjee, M. (**1999**). "Temporal mechanisms underlying recovery from forward masking in multielectrode-implant listeners," J. Acoust. Soc. Am. **105**, 1853–1863.

Chatterjee, M., Galvin, J. J. III, Fu, Q., and Shannon, R. V. (**2006**). "Effects of stimulation mode, level and location on forward-masked excitation patterns in cochlear implant patients," J. Assoc. Res. Otolaryngol. **7**, 15–25.

Dai, H. (**1995**). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," J. Acoust. Soc. Am. **98**, 3135–3139.

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). "Masking-level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopu, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2007**). "Low-frequency and high-frequency cochlear nonlinearity in humans," J. Acoust. Soc. Am. **122**, 1671–1680.

Hicks, M. L., and Bacon, S. P. (**1999**). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," J. Acoust. Soc. Am. **105**, 326–338.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth

behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lopez-Poveda, E. A., and Alves-Pinto, A. (**2008**). "A variant temporal-masking-curve method for inferring peripheral auditory compression," J. Acoust. Soc. Am. **123**, 1544–1554.
Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.
Lopez-Poveda, E. A., Plack, C. J., Meddis, R., and Blanco, J. L. (**2005**). "Cochlear compression in listeners with moderate sensorineural hearing loss," Hear. Res. **205**, 172–183.
Luscher, E., and Zwislocki, J. (**1949**). "Adaptation of the ear to sound stimuli," J. Acoust. Soc. Am. **21**, 135–139.
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.
Meddis, R., and O'Mard, O. P. (**2005**). "A computer model of the auditory-nerve response to forward-masking stimuli," J. Acoust. Soc. Am. **117**, 3787–3798.
Neely, S. T., and Jesteadt, W. (**2005**). "Quadratic-compression model of auditory discrimination and detection," Acust. Acta Acust. **91**, 1–12.
Nelson, D. A., and Freyman, R. L. (**1984**). "Broadened forward-masked tuning curves from intense masking tones: delay-time and probe-level manipulations," J. Acoust. Soc. Am. **75**, 1570–1577.
Nelson, D. A., and Freyman, R. L. (**1987**). "Temporal resolution in sensorineural hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 709–720.
Nelson, D. A., and Pavlov, R. (**1989**). "Auditory time constants for off-frequency forward masking in normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **32**, 298–306.
Nelson, D. A., and Schroder, A. C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2221–2233.
Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.
Oxenham, A. J. (**2001**). "Forward masking: Adaptation or integration?," J. Acoust. Soc. Am. **109**, 732–741.
Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.
Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.
Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.
Plack, C. J., and O'Hanlon, C. G. (**2003**). "Forward masking additivity and auditory compression at low and high frequencies," J. Assoc. Res. Otolaryngol. **4**, 405–415.
Plack, C. J., O'Hanlon, C. G., and Drga, V. (**2005**). "Additivity of masking and auditory compression," in *Auditory Signal Processing: Physiology, Psychophysics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York).
Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.
Plack, C. J., and Oxenham, A. J. (**2000**). "Basilar-membrane nonlinearity estimated by pulsation threshold," J. Acoust. Soc. Am. **107**, 501–507.
Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.
Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**a). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944–966.
Schairer, K. S., Nizami, L., Reimer, J. F., and Jesteadt, W. (**2003**b). "Effects of peripheral nonlinearity on psychometric functions for forward-masked tones," J. Acoust. Soc. Am. **113**, 1560–1573.
Shannon, R. V. (**1990**). "Forward masking in patients with cochlear implants," J. Acoust. Soc. Am. **88**, 741–744.
Stainsby, T. H., and Moore, B. C. J. (**2006**). "Temporal masking curves for hearing-impaired listeners," Hear. Res. **218**, 98–111.
Williams, E. J., and Bacon, S. P. (**2005**). "Compression estimates using behavioral and otoacoustic emission measures," Hear. Res. **201**, 44–54.
Wojtczak, M., and Oxenham, A. J. (**2007**). "Verification of the assumption of frequency-independent recovery from forward masking," J. Acoust. Soc. Am. **121**, 3133.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–219.

Critical ratios of beluga whales (*Delphinapterus leucas*) and masked signal duration

Christine Erbe[a)]
JASCO Research Ltd., 55 Fiddlewood Crescent, Bellbowrie, Qld 4070, Australia

(Received 28 February 2008; revised 7 July 2008; accepted 21 July 2008)

This article examines the masking of a complex beluga vocalization by natural and anthropogenic noise. The call consisted of six 150 ms pulses exhibiting spectral peaks between 800 Hz and 8 kHz. Comparing the spectra and spectrograms of the call and noises at detection threshold showed that the animal did not hear the entire call at threshold. It only heard parts of the call in frequency and time. From the masked hearing thresholds in broadband continuous noises, critical ratios were computed. Fletcher critical bands were narrower than either $\frac{1}{5}$ or $\frac{1}{11}$ of an octave at the low frequencies of the call (<2 kHz), depending on which frequency the animal cued on. From the masked hearing thresholds in intermittent noises, the audible signal duration at detection threshold was computed. The intermittent noises differed in gap length, gap number, and masking, but the total audible signal duration at threshold was the same: 660 ms. This observation supports a multiple-looks model. The two amplitude modulated noises exhibited weaker masking than the unmodulated noises hinting at a comodulation masking release.
 [DOI: 10.1121/1.2970094]

PACS number(s): 43.66.Dc, 43.66.Gf, 43.80.Lb [WWA] Pages: 2216–2223

I. INTRODUCTION

In general terms, we speak of masking when a noise interferes with or obscures a signal. Auditory masking is a matter of everyday experience, and likely affects all life forms with an acoustic sense. The American National Standards Institute defines *masking* as both a process and a quantity (Sec. 12.29):[1]

> (a) The process by which the threshold of hearing for one sound is raised by the presence of another (masking) sound.
>
> (b) The amount by which the threshold of hearing for one sound is raised by the presence of another (masking) sound, expressed in decibels.

Numerically, if SL_0 is the signal level at detection threshold in the absence of noise, and SL_n is the signal level at detection threshold in the presence of noise, then the amount of masking M is simply:

$$M = SL_n - SL_0(\text{dB}). \tag{1}$$

A. Frequency selectivity in masking

In 1940, Fletcher[2] published the results of a number of auditory signal detection experiments and postulated that the human auditory system can be modeled as a series of overlapping bandpass filters. In one of the experiments described, humans detected pure tones in white noise of constant spectral density level but of varying bandwidth. With the tone located at the center frequency of the masking band, Fletcher found that masking (in the sense of definition b) increased with increasing bandwidth up to—what he called—a critical bandwidth. For noise wider than the critical bandwidth, masking remained constant. From this band-widening experiment, Fletcher postulated that at detection threshold and in the case of white noise, the critical bandwidth was numerically equal to the ratio of the intensity of the tone to the average intensity per cycle of the noise.

The American National Standards Institute defines the *critical band* (Sec. 6.26)[3] as

> that frequency band of sound, being part of a continuous-spectrum noise covering a wide band, that contains sound power equal to that of a pure tone centered in the critical band and just audible in the presence of the wideband noise;

and the *critical ratio* (*Fletcher critical band*) (Sec. 6.27)[3] as the

> difference [dB] between the sound pressure level of a pure tone just audible in the presence of a continuous noise of constant spectral density and the sound pressure spectrum level for that noise.

Numerically, if I_t denotes the intensity of the tone and PSD_n the power spectral density (intensity per Hertz) of wideband noise at the levels where the tone is just audible through the noise, then the critical ratio CR becomes

$$CR = 10 \log_{10} \frac{I_t}{PSD_n}. \tag{2}$$

Remembering Fletcher's concept of auditory filters, only a certain part (critical band) of this wideband noise is effective at masking the tone. Under the *equal-power assumption* that the intensity of the tone I_t equals the total intensity of the noise in the corresponding auditory filter, of width Δf, when the tone is just audible, so-called *Fletcher critical bands* can be computed. The equal-power assumption is

$$I_t = PSD_n \, \Delta f. \tag{3}$$

[a)]Electronic mail: christine@jasco.com

 0001-4966/2008/124(4)/2216/8/$23.00

Substituting into Eq. (2) yields the critical ratio in decibels:

$$\mathrm{CR} = 10 \log_{10} \Delta f. \quad (4)$$

Solving for Δf, the Fletcher critical band is

$$\Delta f = 10^{\mathrm{CR}/10}. \quad (5)$$

Fletcher critical bands are estimates of auditory filter widths from experiments that measured critical ratios. Critical bands without Fletcher's name in front are usually measured with direct techniques, such as band-widening experiments. In humans,[4] bottlenose dolphins,[5] and three species of pinnipeds listening in air,[6] critical bands were wider than Fletcher critical bands by a factor 2.5, 5.6, and 3.2—14.2, respectively. The technique used might affect the results; critical bands measured by the two-tone-masking method showed no consistent relationship to critical ratios and were occasionally narrower than Fletcher critical bands in a harbor seal.[7]

The difference between the critical bandwidth and the Fletcher critical bandwidth can be explained in various ways. In the general *power-spectrum model of masking*,[8] the equal-power assumption is discarded. The signal detection threshold is instead assumed to correspond to a critical signal-to-noise power ratio K within the auditory filter:

$$I_t/(\mathrm{PSD}_n \Delta f) = K. \quad (6)$$

Under the equal-power assumption, $K=1$, see Eq. (3). In the power-spectrum model of masking, K is about 0.4 for humans,[4] indicating that tones can be detected when their power is less than that of the noise in the corresponding critical band.

Computing critical bandwidths with Eqs. (5) and (6) assumes that auditory filters are rectangular. Outside of the band Δf, noise power does not add to the masking of a tone in the center; inside of Δf, the noise (if distributed evenly over all frequencies) masks independent of frequency. To allow for nonrectangular auditory filter shapes, a weighting function $W(f)$ is introduced in the power-spectrum model of masking:[9]

$$I_t = K \int \mathrm{PSD}_n(f) W(f) df. \quad (7)$$

Rather than being rectangular, the various weighting functions commonly used slope at the low-frequency and the high-frequency boundaries of the critical band. The actual width of the auditory filter is then defined as either the half-power width or the *equivalent rectangular bandwidth* (ERB).[10] The latter is computed by equating the area underneath $W(f)$ with the area underneath a rectangle of the same height as $W(f)$.

B. Signal duration

Signal duration can affect signal detection. Below some critical signal duration, the detection threshold increases with decreasing signal length. For longer signals, the threshold remains constant. In humans and some terrestrial mammals, the critical duration of signals at communication frequencies is about 150–500 ms.[10–14] Similar values were measured from bottlenose dolphins at communication and echolocation frequencies.[15] One beluga whale showed threshold increases for single tones at 60 kHz that were shorter than 100 ms;[16] thresholds were lower for tones in rapid succession, in line with temporal integration theory (see the following). In a harbor seal, underwater thresholds increased for tones shorter than about 100 ms.[17] A California sea lion and a harbor seal in air had critical tone durations of 300 ms.[18]

These observations are usually explained by the *temporal integration* (also called temporal summation) model[19,20]. The auditory system is assumed to behave like a power integrator resulting in a 3 dB decrease in detection threshold for every doubling of signal duration below a few hundred milliseconds.[21] Experiments on temporal resolution, modulation and gap detection, however, produce much shorter temporal integration times of the order of 1–5 ms.[10] This discrepancy has become known as the integration-resolution paradox[22]. One solution to this paradox is provided by the *multiple-looks model*[23]. In this model, the auditory filters compute outputs over short time windows (approximately 3 ms). These "looks" are stored in short-term memory (with a decay constant of a few hundred milliseconds) and are used intelligently for signal detection and discrimination. For signals much longer than 3 ms or for short signals in repetition, more looks are available to the listener, resulting in improved performance and lower thresholds.

The above-quoted temporal integration studies with marine mammals measured signal duration at detection threshold in the absence of noise; or in the presence of continuous broadband noise. In the presence of noncontinuous but intermittent noise, the masked signal detection threshold depends on the duration of the gaps in the noise and the spectral energy distribution.

C. Masking in beluga whales

Few studies have addressed masking in beluga whales. Johnson *et al.*[24] trained one beluga whale to detect pure tones in broadband noise. Fletcher critical bands were computed by integrating the noise spectral density over frequency, until the noise power equaled the tone power at detection threshold. Fletcher critical bands [Eq. (4)] were about 18 dB for tones below 1 kHz, and increased in bandwidth for tones of higher frequency. Finneran *et al.*[25] trained one beluga whale to detect pure tones in notched noise of varying notch width. Assuming a rounded exponential function for $W(f)$, they computed ERBs as an estimate of auditory filter widths. At 20 and 30 kHz, the ERBs were on average 9.1% and 15.3% of the center frequency, respectively. ERBs were larger at higher frequencies and at higher noise levels.

In a previous study, the author trained one beluga whale to detect complex beluga vocalizations in different types of natural and anthropogenic noise[26,27]. The goal was to estimate the extent of the zone of masking around icebreakers in the wild. Since its publication, the author has repeatedly been asked whether the results of this study could be used to infer more generic characteristics such as critical ratios or critical

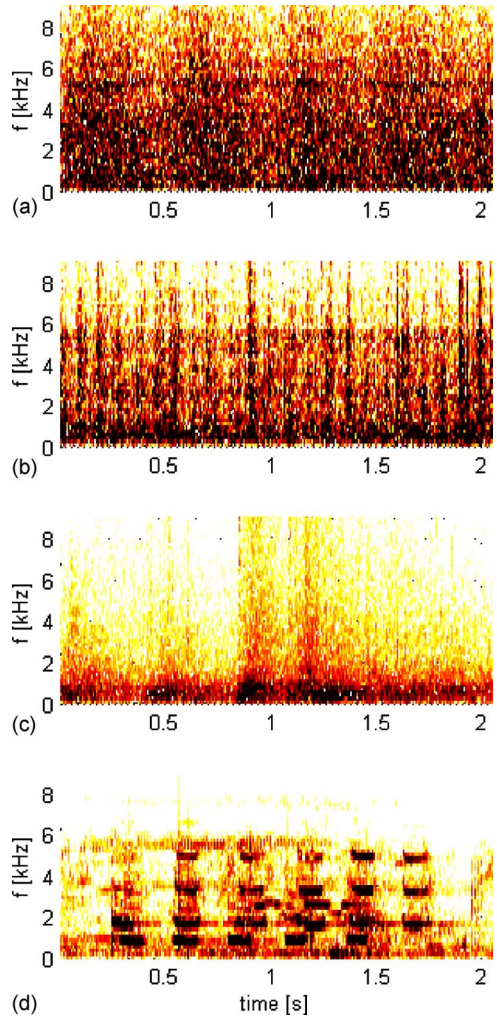


FIG. 1. (Color online) Spectrograms of (a) bubbler system noise and (b) propeller cavitation noise, both as emitted by an icebreaker, (c) naturally occurring, thermal ice-cracking noise, and (d) a beluga vocalization.

bands of the beluga auditory system. This article addresses this question.

II. METHODS AND RESULTS

A. Behavioral experiments

From 1995 to 1997 one beluga whale (Aurora) was trained and used for masking experiments at the Vancouver Aquarium.[26,27] This animal detected calls of its species in four types of noise: bubbler system noise and propeller cavitation noise (both as emitted by an icebreaker), naturally occurring, thermal ice-cracking noise, and Gaussian white noise (Fig. 1). Icebreaker and ice-cracking noises had been recorded in the Arctic, Gaussian white noise was created digitally. White noise, because of its flat frequency spectrum, would appear all black in Fig. 1 and is therefore not shown. The noise recordings were played continuously in the pool for an entire session, by repeating the 2 s clips shown in Fig. 1. The broadband noise level was about 160 dB re 1 μPa at 1 m for all four noises. The beluga vocalization was randomly inserted into the noise, with its level varying in 3 dB steps according to a titration paradigm. The beluga was stationed against a bar 1 m in front of the sound projector and

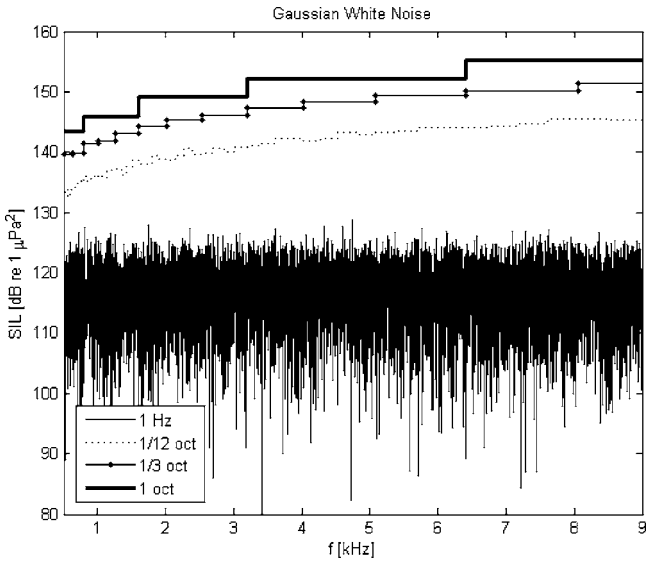


FIG. 2. The bottom line shows the power spectral density of Gaussian white noise. This is equal to the sound intensity level (SIL) in a series of adjacent bandpass filters of constant 1 Hz width. The next line above is the SIL integrated into adjacent bandpass filters that were $\frac{1}{12}$ of an octave wide. On a linear frequency plot, these bands get wider with higher frequency, and the output of the filter increases accordingly. One-third octave bands are wider than $\frac{1}{12}$ octave bands, therefore powers at the output are higher. Each octave band is as wide as three $\frac{1}{3}$ octave bands and twelve $\frac{1}{12}$ octave bands, therefore the power at the octave filters is highest.

indicated signal detection by a go/no-go response. The call detection threshold in the absence of noise was also measured in order to compute the threshold shift in the presence of the above-noted four noises. Masking was 36.5, 33.0, 26.6, and 32.0 dB for bubbler noise, propeller cavitation noise, thermal ice-cracking noise, and white noise, respectively. Details of the behavioral experiment were published earlier.[26,27]

B. Critical ratios

Gaussian white noise has a constant power spectral density (Fig. 2) and is thus an ideal noise for the computation of critical ratios. The call consisted of a fundamental tone at 800 Hz with harmonics and a weaker nonharmonic sidetone. The power spectral densities of the call and the noise *at detection threshold* were integrated over frequency in adjacent bandpass filters. The powers of the call and the noise at the outputs of the filters were compared. Under the equal-power assumption [Eq. (3)], when the power of the call equaled the power of the noise, the bandpass filter approximated the width of the beluga auditory filter, or the Fletcher critical band. The widths of the bandpass filters were varied as $1/n$, from 1 octave wide, to $\frac{1}{2}$ octave, $\frac{1}{3}$ octave, $\frac{1}{4}$ octave, etc. Bandpass filters were centered at the main call frequencies. For Gaussian white noise, the narrower the filters, the lower the power output at each filter (Fig. 2). For pure tones and the tones in the beluga call, the power at the output of the filter does not decrease with decreasing filter width, unless the filter becomes narrower than the tone bandwidth. Therefore, at some filter width, the powers of the call and the noise become equal. For the following analysis, call levels were computed only over the times of the call "phonemes."

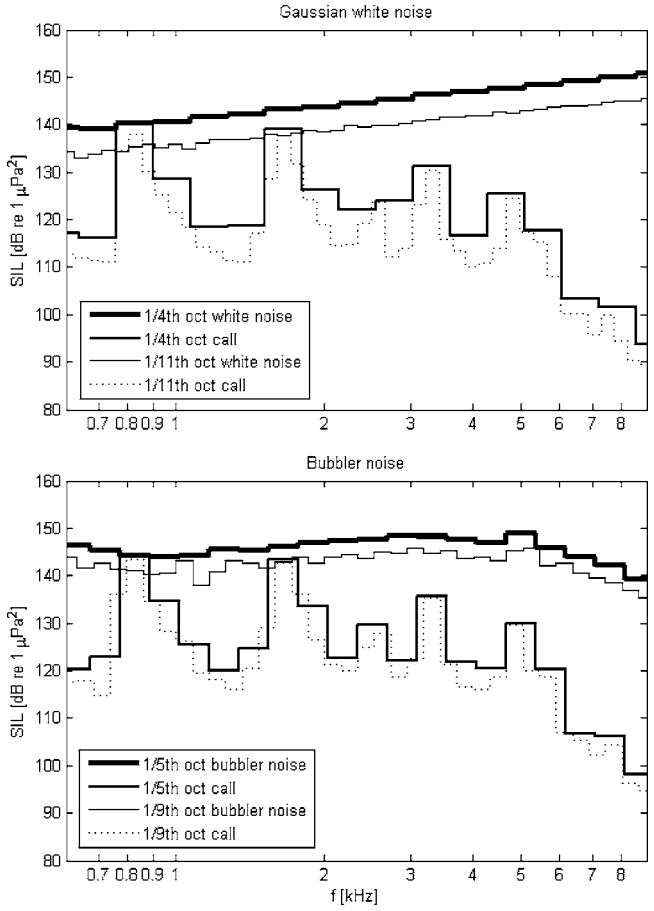


FIG. 3. Top: Band levels of white noise and the beluga call at detection threshold. Bottom: Band levels of bubbler noise and the beluga call at detection threshold. Bands are centered at the call tones. At the wider bandwidths, only the 800 Hz tone of the call was audible; at the narrower bandwidths, the 1600 Hz tone became audible, too (based on the equal-power assumption of masking). Higher call components were never audible, no matter how wide the auditory filter.

From Fig. 1, the call consisted of six bursts (phonemes). The six bursts were extracted and quiet times in between were discarded for the computation of power spectral densities as well as band levels.

Figure 3 shows $\frac{1}{4}$ octave levels of white noise and the beluga call at detection threshold. At the output of the $\frac{1}{4}$ octave band centered at the 800 Hz tone of the call, the call power equalled the noise power. For wider bands, the noise power exceeded the call power; for narrower bands, the call power exceeded the noise power. If the beluga auditory filter at 800 Hz is $\frac{1}{4}$ of an octave wide or narrower, then the animal heard the 800 Hz tone of the call at the behaviorally measured call detection threshold. If the beluga auditory filter at the call frequencies is $\frac{1}{11}$ of an octave wide or narrower, then the animal heard both the 800 Hz tone and the 1600 Hz harmonic at the detection threshold. No matter what the bandwidth, the higher tones of the call did not match the noise power at detection threshold. Under the equal-power assumption of masking, the higher tones of the call were thus not audible at detection threshold. The animal cued on the low-frequency part of the call.

Bubbler system noise is continuous and near white at low frequencies. It was therefore also used for the estimation

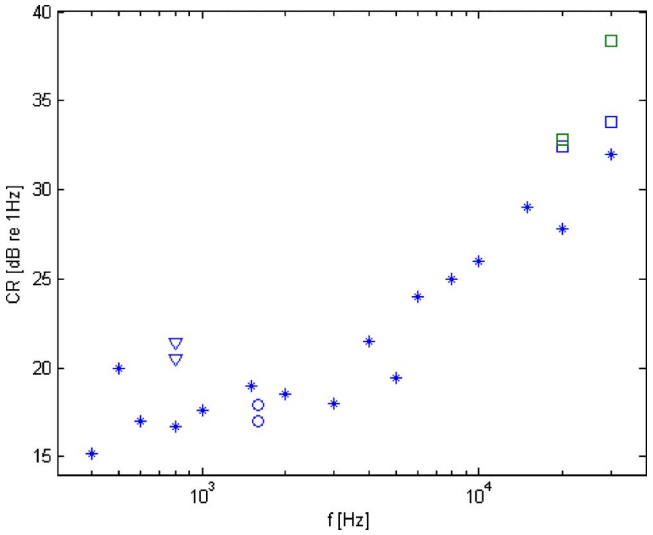


FIG. 4. (Color online) Maximum Fletcher critical bands at 800 Hz (▽) and 1600 Hz (○) for Gaussian white and bubbler noise from the beluga of this study. Fletcher critical bands of Johnson et al. (Ref. 24) from one beluga (∗). ERBs of Finneran *et al.* (Ref. 25) from one beluga (□).

of Fletcher critical bands. Figure 3 shows $\frac{1}{5}$ octave band levels of bubbler noise and the call at detection threshold. At this bandwidth, the power of the 800 Hz tone equaled that of the noise. For $\frac{1}{9}$ octave bands, the 800 Hz tone surpassed the noise in power, and the 1600 Hz harmonic just equaled the noise power. None of the higher tones of the call were able to match the noise power, no matter what the bandwidth.

In summary for white noise and bubbler noise, if the animal cued on the lowest call component at 800 Hz, then its Fletcher critical bands were $\frac{1}{5}$ of an octave wide or narrower. If the animal cued on the lowest two components (800 and 1600 Hz), then its Fletcher critical bands were $\frac{1}{11}$ of an octave wide or narrower. Figure 4 shows the maximum Fletcher critical bands at 800 Hz (CR=21 and 20 dB) and 1600 Hz (CR=17 and 18 dB) for both noises. The only other published Fletcher critical band measurements in belugas at these low frequencies are shown as stars.[24] The only published measurements of beluga ERBs[25] are shown as squares. Critical ratios increase with frequency. The data in Fig. 4 is pooled from three single beluga whales. Critical ratios of Aurora computed under the assumption that she heard both the 800 and the 1600 Hz component of the call (circles) fit the beluga data of Johnson *et al.* better than the critical ratios computed under the assumption that Aurora cued on only the 800 Hz component (triangles). It seems therefore more likely that Aurora cued on the lowest two frequencies of the beluga call rather than on only the very lowest frequency. The Q value (quality factor) of the auditory filter centered at f_0 =1600 Hz was $Q=f_0/\Delta f=13$ based on masking by bubbler noise, and 16 based on white noise.

C. Additional masking by ambient noise

When measuring detection thresholds and computing Fletcher critical bands, it is important to examine whether ambient noise adds to the masking by the test noises. Ambient noise at the Vancouver aquarium where the behavioral experiments were carried out was very variable. The test

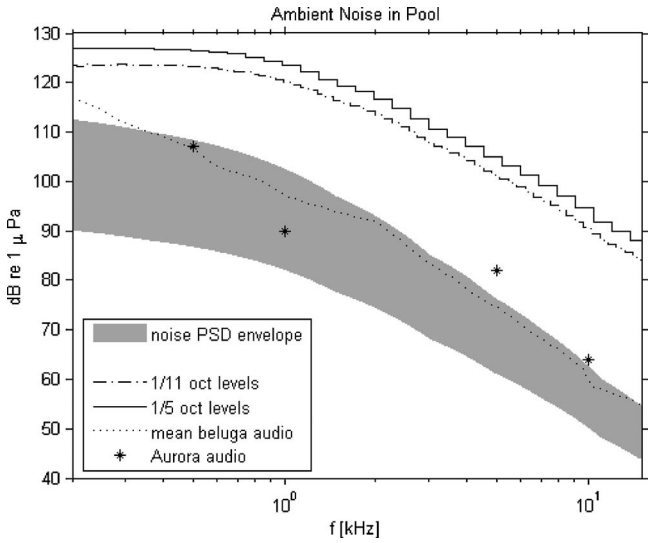


FIG. 5. Ambient noise in the pool: Power spectral densities [dB re 1 μPa^2/Hz] fell into the gray shaded area; also shown are $\frac{1}{5}$ and $\frac{1}{11}$ octave band levels [dB re 1 μPa] of the upper envelope. Further plotted are the mean of all published beluga audiograms and pure tone detection thresholds of the beluga Aurora at four frequencies.

pool was close enough to the machinery room for low-frequency noise to be easily detectable inside the water. Overflow and filtration equipment switched on and off automatically. The pool bordered public parkland, where pathway construction took place during 2 months. Last but not least, a quarter of the data were collected during the rainy season. Ambient noise measurements were done regularly prior to the masking experiments. Figure 5 shows the smoothed envelopes of the lowest and highest ambient noise PSDs.

The beluga audiogram (dotted line) is the mean of four published audiograms.[26,28–30] Pure tone detection thresholds (shown as stars) were also measured for the beluga Aurora used for the masking study. During these audiogram sessions, the equipment around the test pool was switched off; ambient noise levels corresponded to the lower envelope. Power spectral densities of ambient noise have to be at least a critical ratio below the audiogram for pure tone detection thresholds not to be masked. The above-determined maximum critical ratios were 21 dB if the animal cued on 800 Hz and 18 dB if the animal cued on 1600 Hz. At 1 and 10 kHz, ambient noise was not a critical ratio below the audiogram. At 500 Hz and 5 kHz, ambient noise was just about a critical ratio below the audiogram. The pure tone detection thresholds of the beluga whale Aurora, even though they fit in well with the other published audiograms, might therefore have been masked, yielding a masked audiogram. Fletcher critical band levels of the higher PSD envelope are plotted in Fig. 5 as well, showing that ambient noise in the pool at its loudest level was audible to the beluga over the full bandwidth. However, spectrum density levels of ambient noise were well below those of the test noises. Therefore, while ambient noise could have masked the audiogram measurements, it did not add to the masking of the test noises.

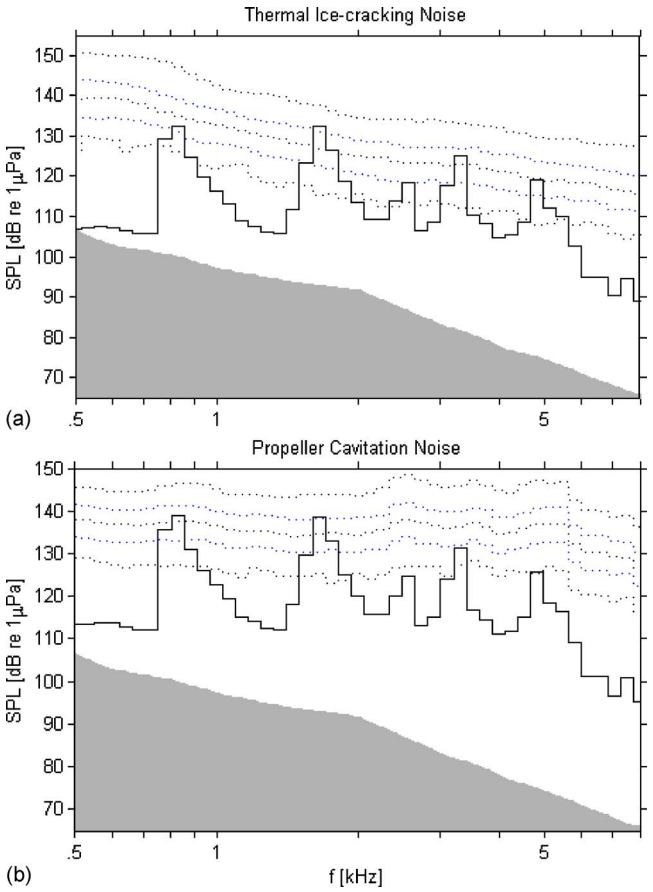


FIG. 6. (Color online) Percentiles of thermal ice-cracking noise and propeller cavitation noise (in $\frac{1}{11}$ octave band levels) as played in the pool. From top to bottom, the dotted lines correspond to the 95th, 75th, 50th, 25th, and 8th percentiles. One-eleventh octave band levels of the call at detection threshold are shown as a solid line. The gray shaded area is the area underneath the mean of the published beluga audiograms.

D. Masked signal duration

Revisiting Fig. 1, while white noise and bubbler system noise were temporally continuous, thermal ice-cracking noise and propeller cavitation noise were intermittent. It was argued earlier that the masking potential of these noises was less than that of continuous noise, because an animal might detect pieces of the call through the quieter gaps in the noise.[26,27] How much of the call did the animal hear at detection threshold?

Figure 6(a) shows percentiles of the thermal ice-cracking noise (in $\frac{1}{11}$ octave band levels) as it was played in the pool. From top to bottom, the dashed lines correspond to the 95th, 75th, 50th, 25th, and 8th percentiles. The noise consisted of 2 s samples which were played continuously in repetition. Ninety five percent of the time, the noise band levels were below the 95th percentile curve, etc. The 2 s call was inserted at random times; its level variation followed a titration paradigm. One-eleventh octave band levels of the call at detection threshold are plotted as a solid line in Fig. 6. Assuming $\frac{1}{11}$ octave wide critical bands at 1600 Hz, 75% of the time the noise was quiet enough for the 1600 Hz peak of the call to be audible (under the equal-power assumption). Given that the call phonemes only occupied about 44% (880 ms) of the 2 s call, the animal would have heard the

1600 Hz component for on average 33% (=75% × 44%) of the time, or for 660 ms. Assuming $\frac{1}{11}$ octave wide critical bands over the full bandwidth of the call, 25% of the time the noise was quiet enough for all of the spectral peaks (with the exception of the minor 7.5 kHz component) to be audible. The animal would have heard the full bandwidth of the phonemes for 220 ms.

Band levels of propeller cavitation noise were below the 1600 Hz peak of the call 75% of the time [Fig. 6(b)]. As in the above-mentioned case of thermal ice-cracking noise, the 1600 Hz call component would have been audible for 660 ms. Assuming $\frac{1}{11}$ octave wide bands over the full bandwidth of the call, the 800 Hz peak would have been audible for slightly less of the time, with the high-frequency components hardly ever audible.

III. DISCUSSION

While the data used in this article were obtained from a captive whale, the audiogram measurement showed that this animal had "normal" hearing (compared to other captive animals); and there was no reason to believe that its signal detection capabilities would differ from animals in the wild. While an enclosed pool has its own frequency response, potentially amplifying and damping certain frequencies at certain locations, the effect would be the same for signal and noise, therefore, the signal-to-noise ratio in the beluga's critical bands does not change. A full analysis of reverberation in the pool was done at the time of the behavioral experiments.[26]

A. Critical ratios

Call detection thresholds in four types of noises were measured behaviorally with a trained beluga whale (Aurora). Two types of noises were continuous and broadband: artificially created, Gaussian white noise and bubbler system noise recorded from an icebreaker in the field. Band levels of varying width were plotted for the call and noises at detection threshold. Only for the two lowest frequencies of the call was there a bandwidth for which the noise level equaled the call level. In other words, under the equal-power assumption of masking, only the two lowest frequencies could have been audible to the whale at detection threshold. If the animal cued on the 800 Hz component of the call, then its Fletcher critical bands were $\frac{1}{5}$ of an octave wide or narrower. If the animal cued on both the 800 and the 1600 Hz components, then its Fletcher critical bands were $\frac{1}{11}$ of an octave wide or narrower. Critical ratios at these frequencies have been published for one other beluga whale [24]. One-eleventh octave bands fit the published data better than $\frac{1}{5}$ octave bands; therefore, Aurora likely cued on 1600 Hz as well as on 800 Hz.

B. Masked signal duration

Two types of noises were intermittent and broadband: propeller cavitation noise recorded from an icebreaker and naturally occurring, thermal ice-cracking noise. Using $\frac{1}{11}$ octave band filters, intensity level percentiles of the noises were plotted and compared to the call level at threshold. For both thermal ice-cracking noise and propeller cavitation noise, the most intense frequency of the call (1600 Hz) reached the 75th percentile of the noise. In other words, for 75% of the time, the noise was quiet enough for the animal to detect the 1600 Hz component of the call—under the equal-power assumption. Considering the durations of the phonemes in the call, the animal would have heard the 1600 Hz component for a total length of 660 ms at detection threshold—in both propeller cavitation and ice-cracking noise. The animal might have cued on different frequencies when presented with different masking noises. However, the 1600 Hz component was audible for the largest percentage of time in both situations. This analysis is based on the outcomes of the critical ratio analysis, i.e., assuming Fletcher critical bands were $\frac{1}{11}$ of an octave wide. For wider critical bands, the call levels in Fig. 6 would remain roughly the same, while the noise levels would increase. The call frequencies would hence reach lower percentiles, decreasing the audible signal duration at detection threshold.

Taking a closer look at the gap structure of the two intermittent noises (Fig. 1), natural ice-cracking noise consisted of sharp pulses with up to 700 ms of quiet gaps in between. Propeller cavitation noise consisted of 12 pulses/*s*, leaving quiet gaps for less than 60 ms. Each phoneme in the beluga call had a duration of about 150 ms. The animal would have heard a number of phonemes uninterruptedly through the relatively long gaps in ice-cracking noise. But it would not have heard uninterrupted phonemes through the short gaps in propeller noise. In propeller noise, Aurora would have detected a series of short chunks of phonemes. As the signal was inserted randomly into the continuous noise, the time when the signal happened in the noise was a stochastic variable. On average, at detection threshold, the animal would have heard 660 ms of the 1600 Hz component. This is the combined length of the chunks of signal surpassing the noise, summed over the 2 s duration of the call. The other components of the call would have been audible for shorter durations.

While these durations are of the order of the critical signal lengths measured with terrestrial and marine mammals,[10–18] no mammal tested to date has indicated temporal integration times of up to 2 s. For the detection of signals through gaps in noise, it does not make sense for the subject to integrate energy over long time periods, because if the integration time extends beyond the gap, then the noise energy would quickly outgrow the integrated signal energy, hence increasing the masking. A temporal integration model seems implausible.

While the behaviorally measured, absolute masking [Eq. (1)] differed for ice-cracking (26.6 dB) and propeller cavitation noise (33.0 dB), the call detection threshold in these noises was such that the total duration of audible call components was the same (660 ms), despite the very different gap pattern of these two noises. A theory that can explain this observation is the theory of multiple looks. If each look lasts only a few millisecond[23] then it does not matter how the signal is chopped up, as long as the gaps in the noise (minimum 60 ms here) are not shorter than each look. Propeller

noise chopped the signal into more and shorter chunks than ice-cracking noise, but Aurora had the same number of looks at the signal in both noises: all looks combining to 660 ms.

C. Adaptations for masking

It is interesting to note that the slope of the call peaks roughly follows the slope of the audiogram as well as the slope of the naturally occurring, thermal ice-cracking noise [Fig. 6(a)]. While the two anthropogenic ship noises and the artificial white noise masked the high-frequency components of the call more than the low-frequency components, "forcing" the animal to cue on the low-frequency components during the detection task, Arctic ambient noise enabled the animal to hear the full bandwidth of the call at detection threshold. It is debatable whether this is coincidental or some adaptation of hearing and/or communication signals to natural ambient noise.

A few studies have shown an adaptation of marine mammal calls to prevalent background noise. Right whales in areas with substantial low-frequency shipping noise shifted their call frequencies up over the lifespan of individual animals, probably to compensate for masking.[31] A comparison of whistles from three separate populations of Indo-Pacific bottlenose dolphins showed that animals in habitats with less noise produced whistles over a broader range of frequencies and with greater modulations than animals in noisy habitats.[32] Bottlenose dolphins in Florida increased their whistle repetition rate upon watercraft approach, which was suggested as a compensation for masking.[33] Beluga whales in the St. Lawrence increased the production of falling tonal calls and pulsed calls, increased the repetition of calls, and shifted call frequencies up in the presence of ship noise.[34] They also increased the call level (Lombard response).[35] Killer whales in Puget Sound increased their call durations in the presence of boats.[36]

In the case of pinnipeds, calls are subject to masking by conspecifics, particularly during the breeding season. Harp seals have been shown to use frequency and temporal separation in conjunction with a wide vocal repertoire to reduce masking.[37] Polar seals further use predominantly downswept calls,[38] and downsweeps have lower detection thresholds than upsweeps.[39] Directional hearing capabilities help reduce masking when the signal and the noise arrive from different directions.[40–42] Another antimasking strategy is the use of within-call repetition.[43] The beluga call used for the current masking study consisted of six identical (in the sense of having the same time and frequency characteristics) phonemes. In line with multiple-looks theory, within-call repetition provides a larger number of looks, resulting in a lower detection threshold. The theory has been shown to hold in the case of signals with characteristics of communication sounds. In humans, the discrimination of syllables (each lasting 500 ms) improved when the stimuli were repeated.[44,45]

D. Comodulation masking release

The power-spectrum model of masking assumes that the listener focuses on the critical band around the signal frequency and that the detection threshold is related to a certain signal-to-noise ratio at the output of this filter. Across-filter comparisons, however, can aid signal detection under certain conditions in humans.[10] Given a narrow-band or pure-tone signal and a broadband noise, when the noise is amplitude modulated, and if this modulation is coherent or correlated across different frequency bands, then the masking is reduced compared to the case of unmodulated or incoherently modulated noise. This is termed the comodulation release of masking[46] (CMR).

The current study used four broadband noises. Gaussian white noise and bubbler noise are not amplitude modulated; propeller noise and ice-cracking noise, however, exhibit an amplitude modulation—coherently across frequency. The modulation rate for propeller noise was about 12 Hz and was constant in time. The modulation rate for ice-cracking noise was lower but random. The absolute masking as defined in Eq. (1) was lowest for ice-cracking noise (26.6 dB), followed by Gaussian noise (32.0 dB), then propeller noise (33.0 dB), then bubbler noise (36.5 dB). The weakest masker was a comodulated noise, the strongest masker was an unmodulated noise. The masking of propeller and Gaussian noise was similar. However, the four maskers were normalized by their total power over 2 s and 22 kHz. Gaussian noise had more energy at high frequencies outside of the call frequencies than the other three noises, thus reducing its on-band masking effect. While the data support a CMR in marine mammals, the amount of this release cannot be derived from the data. The noises used in this study differed in a multitude of variables (frequency, time, power spectral density, modulation, etc.); and the signal had a complex frequency and time structure. An animal in a complex acoustic environment might use a combination of differing strategies to reduce masking. To test for particular strategies or models, signals and maskers need to be chosen such as to reduce the number of variables.

[1] *American National Standard: Acoustical Terminology ANSI S1.1* (American National Standards Institute, Inc., and Acoustical Society of America, New York, 1994).
[2] H. Fletcher, "Auditory patterns," Rev. Mod. Phys. **12**, 47–65 (1940).
[3] *American National Standard: Bioacoustical Terminology ANSI S3.20* (American National Standards Institute, Inc., and Acoustical Society of America, New York, 1995).
[4] B. Scharf, "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York, 1970), Vol. **1**, pp. 159–202.
[5] W. W. L. Au and P. W. B. Moore, "Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **88**, 1635–1638 (1990).
[6] B. L. Southall, R. J. Schusterman, and D. Kastak, "Auditory masking in three pinnipeds: Aerial critical ratios and direct critical bandwidth measurements," J. Acoust. Soc. Am. **114**, 1660–1666 (2003).
[7] S. D. Turnbull and J. M. Terhune, "White noise and pure tone masking of pure tone thresholds of a harbour seal listening in air and underwater," Can. J. Zool. **68**, 2090–2097 (1990).
[8] R. D. Patterson and B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, New York, 1986), pp. 123–127.
[9] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138 (1990).
[10] B. C. J. Moore, *An Introduction to the Psychology of Hearing* (Academic, San Diego, 1997).
[11] J. W. Hughes, "The threshold of audition for short periods of stimulation," Proc. R. Soc. London, Ser. B **133**, 486–490 (1946).
[12] W. R. Garner and G. A. Miller, "The masked threshold of pure tones as a

function of duration,"J. Exp. Psychol. **37**, 293–303 (1947).
[13]T. D. Clack, "Effect of signal duration on the auditory sensitivity of humans and monkeys (*Macaca mulatta*)," J. Acoust. Soc. Am. **40**, 1140–1146 (1966).
[14]R. R. Fay, *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL, 1988).
[15]C. S. Johnson, "Relation between absolute threshold and duration-of-tone pulses in the bottlenosed porpoise," J. Acoust. Soc. Am. **43**, 757–763 (1968).
[16]C. S. Johnson, "Hearing thresholds for periodic 60-kHz tone pulses in the beluga whale," J. Acoust. Soc. Am. **89**, 2996–3001 (1991).
[17]J. M. Terhune, "Detection thresholds of a harbour seal to repeated underwater high-frequency, short-duration sinusoidal pulses," Can. J. Zool. **66**, 1578–1582 (1988).
[18]M. M. Holt, B. L. Southall, D. Kastak, R. J. Schusterman, and C. Reichmuth-Kastak, "Temporal integration in a California sea lion and a harbor seal: Estimates of aerial auditory sensitivity as a function of signal duration," J. Acoust. Soc. Am. **116**, 2531 (2004) (Abstract).
[19]R. Plomp and M. A. Bouman, "Relation between hearing and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758 (1959).
[20]R. Plomp, "Hearing thresholds for periodic tone pulses," J. Acoust. Soc. Am. **33**, 1561–1567 (1961).
[21]G. M. Gerken, V. K. H. Bhat, and M. H. Hutchison-Clutter, "Auditory temporal integration and the power-function model," J. Acoust. Soc. Am. **88**, 767–778 (1990).
[22]E. de Boer, "Auditory time constants: A paradox?," in *Time Resolution in Auditory Systems*, edited by A. Michelson (Springer, Berlin, 1975), pp. 141–158.
[23]N. F. Viemeister and G. H. Wakefield, "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865 (1991).
[24]C. S. Johnson, M. W. McManus, and D. Skaar, "Masked tonal hearing thresholds in the beluga whale," J. Acoust. Soc. Am. **85**, 2651–2654 (1989).
[25]J. J. Finneran, C. E. Schlundt, D. A. Carder, and S. H. Ridgway, "Auditory filter shapes for the bottlenose dolphin (*Tursiops truncates*) and the white whale (*Delphinapterus leucas*) derived with notched noise," J. Acoust. Soc. Am. **112**, 322–328 (2002).
[26]C. Erbe and D. M. Farmer, "Masked hearing thresholds of a beluga whale (Delphinapterus leucas) in icebreaker noise," Deep-Sea Res., Part II **45**, 1373–1388 (1998).
[27]C. Erbe, "Detection of whale calls in noise: Performance comparison between a beluga whale, human listeners, and a neural network," J. Acoust. Soc. Am. **108**, 297–303 (2000).
[28]F. T. Awbrey, J. A. Thomas, and R. A. Kastelein, "Low-frequency underwater hearing sensitivity in belugas Delphinapterus leucas," J. Acoust. Soc. Am. **84**, 2273–2275 (1988).
[29]C. S. Johnson, M. W. McManus, and D. Skaar, "Masked tonal hearing thresholds in the beluga whale," J. Acoust. Soc. Am. **85**, 2651–2654 (1989).
[30]M. J. WhiteJr., J. Norris, D. Ljungblad, K. Baron, and G. diSciara, "Auditory thresholds of two beluga whales *Delphinapterus leucas*," Report No. 78-109, Hubbs/SeaWorld Research Institute for Naval Ocean System Center, San Diego, 1978, pp. 35.
[31]S. E. Parks, C. W. Clark, and P. L. Tyack, "Short- and long-term changes in right whale calling behavior: The potential effects of noise on acoustic communication," J. Acoust. Soc. Am. **122**, 3725–3731 (2007).
[32]T. Morisaka, M. Shinohara, F. Nakahara, and T. Akamatsu, "Effects of ambient noise on the whistles of Indo-Pacific bottlenose dolphin populations," J. Mammal. **86**, 541–546 (2005).
[33]K. C. Buckstaff, "Effects of watercraft noise on the acoustic behavior of bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida," Marine Mammal Sci. **20**, 709–725 (2004).
[34]V. Lesage, C. Barrette, M. C. S. Kingsley, and B. Sjare, "The effect of vessel noise on the vocal behavior of belugas in the St. Lawrence River Estuary, Canada," Marine Mammal Sci. **15**, 65–84 (1999).
[35]P. M. Scheifele, S. Andrew, R. A. Cooper, and M. Darre, "Indication of a Lombard vocal response in the St. Lawrence River beluga," J. Acoust. Soc. Am. **117**, 1486–1492 (2005).
[36]A. D. Foote, R. W. Osborne, and A. R. Hoelzel, "Whale-call response to masking boat noise," Nature (London) **428**, 910 (2004).
[37]A. Serrano and J. M. Terhune, "Antimasking aspects of harp seal (*Pagophilus groenlandicus*) underwater vocalizations," J. Acoust. Soc. Am. **112**, 3083–3090 (2002).
[38]J. Terhune, "Antimasking strategies of underwater vocalizations and hearing abilities of polar seals," J. Acoust. Soc. Am. **116**, 2555 (2004) (Abstract).
[39]S. Turnbull and J. M. Terhune, "Descending frequency swept tones have lower thresholds than ascending frequency swept tones for a harbor seal (*Phoca vitulina*) and human listeners," J. Acoust. Soc. Am. **96**, 2631–2636 (1994).
[40]M. M. Holt and R. J. Schusterman, "Spatial release from masking of aerial tones in pinnipeds," J. Acoust. Soc. Am. **121**, 1219–1225 (2007).
[41]S. D. Turnbull, "Changes in masked thresholds of a harbor seal *Phoca vitulina* associated with angular separation of signal and noise sources," Can. J. Zool. **72**, 1863–1866 (1994).
[42]D. E. Bain and M. E. Dahlheim, "Effects of masking noise on detection thresholds of killer whales," in *Marine Mammals and the Exxon Valdez*, edited by T. R. Loughlin (Academic, San Diego, 1994), pp. 243–256.
[43]A. Serrano and J. M. Terhune, "Within-call repetition may be an antimasking strategy in underwater calls of harp seals (*Pagophilus groenlandicus*)," Can. J. Zool. **79**, 1410–1413 (2001).
[44]R. Frush Holt and A. Earley Carney, "Multiple looks in speech sound discrimination in adults," J. Speech Lang. Hear. Res. **48**, 922–943 (2005).
[45]R. Frush Holt and A. Earley Carney, "Developmental effects of multiple looks in speech sound discrimination," J. Speech, Language, J. Speech Lang. Hear. Res. **50**, 1404–1424 (2007).
[46]J. W. Hall, M. P. Haggard, and M. A. Fernandes, "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56 (1984).

Disentangling the effects of spatial cues on selection and formation of auditory objects[a)]

Antje Ihlefeld and Barbara Shinn-Cunningham[b)]
Auditory Neuroscience Laboratory, Boston University Hearing Research Center, 677 Beacon Street, Boston, Massachusetts 02215

(Received 19 December 2006; revised 8 July 2008; accepted 16 July 2008)

When competing sources come from different directions, a desired target is easier to hear than when the sources are co-located. How much of this improvement is the result of spatial attention rather than improved perceptual segregation of the competing sources is not well understood. Here, listeners' attention was directed to spatial or nonspatial cues when they listened for a target masked by a competing message. A preceding cue signaled the target timbre, location, or both timbre and location. Spatial separation improved performance when the cue indicated the target location, or both the location and timbre, but not when the cue only indicated the target timbre. However, response errors were influenced by spatial configuration in all conditions. Both attention and streaming contributed to spatial effects when listeners actively attended to location. In contrast, when attention was directed to a nonspatial cue, spatial separation primarily appeared to improve the streaming of auditory objects across time. Thus, when attention is focused on location, spatial separation appears to improve both object selection and object formation; when attention is directed to nonspatial cues, separation affects object formation. These results highlight the need to distinguish between these separate mechanisms when considering how observers cope with complex auditory scenes. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2973185]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Qp, 43.66.Ba [RLF] Pages: 2224–2235

I. INTRODUCTION

Attention is critical in enabling us to select important information from the overwhelming flow of events that continuously reaches our senses (Desimone and Duncan, 1995; Pashler, 1998). Because attention is often object based (O'Craven *et al.*, 1999; Scholl, 2001; Busse *et al.*, 2005; Cusack *et al.*, 2000; Shinn-Cunningham, 2008), the way we organize sensory inputs into perceptual objects is likely to affect how attention can modulate perception in all modalities (Darwin *et al.*, 2003; Wolfe *et al.*, 2003; Shomstein and Yantis, 2004; Serences *et al.*, 2005). Here, we examined how attention to spatial and/or nonspatial auditory cues can affect performance in a speech identification task with two concurrent talkers.

Spatial cues are critical both for forming visual objects and selecting objects from a complex visual scene (e.g., see Egly *et al.*, 1994; Knudsen 2007). The important role of spatial cues on visual object selection and object formation is not surprising when one considers how visual sensory inputs are encoded. The retina is topographically organized and directly encodes two-dimensional visual spatial information in parallel channels comprised of distinct neural populations. Visual objects are primarily determined by local spatial structure (e.g., edges define the boundaries of connected regions); higher-order spatial and nonspatial features determine which regions belong to one object.

In contrast, sound from all source directions adds acoustically before entering the ear, where the cochlea processes inputs in distinct frequency channels. Unlike visual spatial information, all auditory spatial information is computed from the sound mixture reaching the ears. Moreover, whereas even a static two-dimensional visual scene is often rich in information, auditory information is conveyed by changes in sound over time.

Perhaps as a result, spectrotemporal sound structure rather than spatial information dominates how a sound mixture is segmented over short time scales (i.e., how auditory objects are formed at the syllable level; Kubovy, 1981; Culling and Summerfield, 1995a; Darwin, 1997; Darwin, 2008; Shinn-Cunningham, 2008; Ihlefeld and Shinn-Cunningham, 2008). Higher-order features (e.g., timbre, pitch, and perceived location) are thought to determine how these local segments are organized across longer time scales to form auditory "streams" (Bregman, 1990; Darwin, 1997; Deutsch, 1999; Darwin and Hukin, 2000; Shinn-Cunningham, 2008). In general, masking can be reduced if target and masker are dissimilar in one or more attributes, including fundamental frequency, timbre (e.g., vocal tract length and intonation), overall signal intensity, and perceived spatial location (Culling *et al.*, 1994; Culling and Summerfield, 1995b; Freyman *et al.*, 2005; Rakerd *et al.*, 2006).

When listening for a target voice in a mixture of other voices, short-term segmentation of concurrent speech is often not the main factor limiting performance, presumably because of the rich spectrotemporal structure of speech. In particular, at least for closed-set speech identification in the

[a)]Portions of this work were presented at the 2006 Mid-Winter meeting of the Association for Research in Otolaryngology.

[b)]Also at Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 022105; electronic mail: shinn@cns.bu.edu

 0001-4966/2008/124(4)/2224/12/$23.00

Coordinate Response Measure (CRM) paradigm (Bolia *et al.*, 2000), listeners rarely report words that are not present in the sound mixture (what we will henceforth label as *drop* errors [e.g., see Kidd *et al.* (2005a)]. Instead, listeners usually err by either reporting a mixture of words from different sources (*mix* errors) or reporting all of the words from the wrong source (*masker* errors).

Many studies demonstrate that spatial separation of target and masker reduces response errors (e.g., Kidd *et al.*, 2005a; Best *et al.*, 2006; Ihlefeld and Shinn-Cunningham, 2008; Brungart, 2001; Brungart *et al.*, 2001; Brungart and Simpson, 2004). When the dominant form of interference is energetic masking, such that the neural representation of the masker energy at the auditory periphery occludes that of the target energy (e.g., see Durlach *et al.*, 2003; Kidd *et al.*, 2008), spatial differences between target and masker can improve target intelligibility, even when the spatial attributes of target and/or masker are ambiguous or inconsistent across frequency (Edmonds and Culling, 2005a; Edmonds and Culling, 2005b; Culling *et al.*, 2006). However, studies showing this effect all employed maskers that differed from the target in some nondirectional attributes (so that the target was easy to select from the mixture). This suggests that when the dominant interference is energetic masking, attention to a particular direction is not necessary to exploit spatial cues. Other studies show that when target and masker are perceptually similar or when the listener is uncertain about target features (i.e., when "informational masking" is the dominant form of perceptual interference, see Durlach *et al.*, 2003; Kidd *et al.*, 2008), directing spatial attention can improve target identification and/or detection (Arbogast and Kidd, 2000; Freyman *et al.*, 2005; Rakerd *et al.*, 2006; Kidd *et al.*, 2005a).

While there are many studies showing benefits of spatially separating target and masker signals, no past study has disentangled whether these improvements come solely from the listener directing attention to the target location or whether spatial continuity can contribute to performance through automatic improvements in streaming (linking together words from one source over time) even when attention is not spatially directed (Shinn-Cunningham, 2008). Specifically, spatial separation may increase the likelihood that the keywords from a sound source are linked together properly across time. Such automatic streaming should reduce mix errors, even when listeners attend to a nonspatial attribute of the target voice or when they happen to attend to another voice in the acoustic mixture. Similarly, even without spatial attention, spatial separation may make it easier to selectively attend to the keywords that have a desired timbre, which could reduce both mix and masker errors.

Here, we examined whether the rates at which different kinds of response errors occurred were similar when listeners directed their attention to spatial and timbral cues, and how overall performance and errors varied with spatial separation of the target and masker. We manipulated what features subjects attended to when listening for a target message played simultaneously with a concurrent masker. In order to deemphasize the role of nonspatial higher-order acoustic cues (such as fundamental frequency or vocal tract length), we used sine-wave vocoded speech.[1] Both target and masker stimuli were derived from utterances of the same talker. The keywords in these stimuli were nearly synchronous and possessed no strong pitch.

Results suggest that the increasing spatial separation of the competing talkers improves the ability to *select* the desired source only when a listener is attending to space. In contrast, spatial separation may improve streaming (the linking of sound from one source across time) both when attention is spatially directed and when attention is directed to a nonspatial feature.

II. METHODS

A. Subjects

Nine normal-hearing fluent speakers of American English (ages 20–32) were paid to participate. All subjects gave written informed consent (as approved by the Boston University Charles River Campus Institutional Review Board) before participating in the study.

B. Stimuli

Raw speech stimuli were derived from the CRM corpus (see Bolia *et al.*, 2000), which consists of sentences of the form "Ready ⟨call sign⟩ go to ⟨color⟩ ⟨number⟩ now." Target and masker [⟨color⟩ ⟨number⟩] phrases were extracted from the original utterances by time windowing. ⟨Color⟩ was one of the set (white, red, blue, and green). ⟨Number⟩ was one of the digits between 1 and 8, excluding the two-syllable digit seven. Five (arbitrarily selected) instances of the word "ready" were also extracted to serve as cue words.[2] The cue word was processed in the same way as the target and masker phrases.

In each trial, two different [⟨color⟩⟨number⟩] phrases were used as sources. The numbers and colors in the competing utterances were randomly chosen but constrained to differ from each other in each trial. In each trial, the designated target message was preceded by the cue word "ready," chosen randomly from the five instances. The cue word, which was approximately 300 ms long, was concatenated with the target phrase without any inserted delay. In order to minimize differences between concurrent messages, the same talker was used for both phrases (talker 0 was chosen because it is the talker with the smallest variance in speaking rate in the CRM corpus).

In order to reduce peripheral interference between the competing messages and to better isolate object- and attention-related effects on performance, the raw target, masker, and cue words were processed to produce intelligible, spectrally sparse signals that used nonoverlapping frequency bands (Arbogast and Kidd, 2000; Shinn-Cunningham *et al.*, 2005a). Each raw target phrase, masker phrase, and cue word was filtered by ten one-third-octave wide bandpass filters (fourth-order Butterworth filter) with center frequencies spaced linearly on a logarithmic scale between 250 and 3342 Hz.[3] The Hilbert envelope of each band was used to amplitude-modulate a sinusoidal carrier whose frequency matched the corresponding band's center frequency.

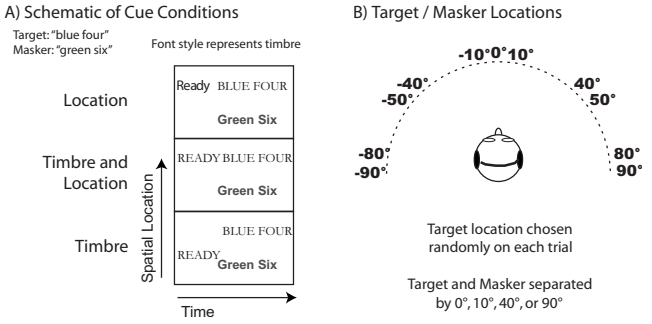


FIG. 1. Illustration of the experimental conditions. (A) Diagram of the cue conditions. The target message ("blue four") and the cue ("ready") match either in location (represented by vertical displacement), timbre (represented by font style), or timbre and location. The masker message ("green six") never matches the timbre of the cue or target. (B) The diagram of the spatial locations tested. The target was equally likely to come from each of the possible locations. The separation between the target and masker varied from trial to trial and was either 0°, 40°, or 90°.

On each trial, five of the resulting amplitude-modulated sinusoids were summed to produce the spectrally sparse target signals; the remaining five bands were summed to create the masker. For each subject, there were three possible sets of timbre from which target and masker signals were constructed. Each subject-specific set of timbres consisted of two combinations of five frequency bands each (randomly set for each subject at the start of the experiment). To ensure that spectral content was comparable across target and masker signals and across the three sets of timbres, each frequency band combination consisted of three from the lower six frequency bands (250–1056 Hz) and two from the upper four bands (1408–3342 kHz). On each trial, one of the three sets of timbres was selected, and one of its frequency band combinations was used to create the target; the other frequency band combination from the same set of timbres was used to generate the masker.

On a given trial, the frequency bands for the cue word "ready" either matched the target (in the timbre and timbre-and-location conditions) or were randomly chosen with the constraint that they did not match either target or masker [in the location condition; see Fig. 1(a)]. As stated above, there were three possible sets of timbres for each subject, a number chosen (somewhat arbitrarily) so that the number of possible timbres matched the number of possible spatial separations between target and masker.

C. Spatial synthesis

The broadband root-mean-square (rms) energy of the spectrally sparse target and masker signals was equalized. Then the equalized signals were processed to produce spatial cues by filtering them with pseudo-anechoic head-related transfer functions (HRTFs) measured on a Knowles Electronics Manikin for Acoustic Research (KEMAR) [for details, see Shinn-Cunningham *et al.*, (2005a, 2005b)]. HRTFs were measured in the horizontal plane containing the ears for sources at a distance of 1 m and at various azimuthal locations (±90°, ±80°, ±50°, ±40°, ±10°, and 0°).[4] The resulting binaural signals contained all of the appropriate spatial auditory cues for a source from the simulated location.

The target location was chosen randomly from trial to trial and was equally likely to be from any of the 11 locations. On each trial, the angular separation between target and masker was randomly chosen [either 0°, 10°, 40°, or 90°; however, in the location condition, the separation was never 0°; see Fig. 1(b)]. The cue location was either the same as the target (in the location and timbre-and-location conditions) or chosen randomly to differ from both target and masker locations [in the timbre condition; see Fig. 1(a)]. Following spatial synthesis, the cue word (in quiet) was concatenated with the sum of the target and masker signals.

Although target and masker had equal broadband rms energy prior to spatial processing with HRTFs, the spatial processing introduced level differences in the presentation of the sources at each ear. To remove any possible artifacts caused by variations in overall loudness with spatial configuration, on each trial, the overall level of the spatially processed target and masker pair was randomly roved over a range of 10 dB (average level set to 65 dB SPL).

D. Procedures

Stimuli were digital/analog (D/A) converted, amplified using Tucker-Davis System 3 hardware, and presented over Sennheiser HD 580 headphones to subjects seated in a sound-attenuated chamber. Following each trial, subjects indicated the perceived target keywords using a graphical user interface (GUI), after which the GUI indicated the correct response.

Prior to the experiment, subjects were screened to ensure that they could identify the color and number of the spectrally sparse processed speech in quiet. All subjects achieved 90% correct or better over the course of 50 trials. At the beginning of each session, subjects completed two short (100-trial) tests to sensitize them to the stimulus timbres and locations, respectively. In these tests, a cue/target phrase was presented in quiet. In the first test, cue and target had the same timbre on half of the trials and different timbres on the other half of the trials. Similarly, in the second test, cue and target had the same location on half of the trials and different locations on the other half of the trials. For both tests, the locations and timbres of cue and target were randomly selected on each trial and differed from each other, encouraging subjects to focus on the feature of interest. Subjects were asked whether the cue and target had the same or different timbre or location (first and second tests, respectively).[5]

In each block, subjects were instructed to report the color and number associated with the cue word timbre, location, or timbre and location, ignoring the message of the masker. In the timbre condition, listeners were instructed as follows: "Listen to what 'ready' sounds like and report the color and number that sound similar to 'ready.'" In the location condition, subjects were instructed as follows: "Listen to where 'ready' is coming from and report the color and number that come from the same location as 'ready.'" In the timbre-and-location condition, subjects were instructed as follows: "Listen to what 'ready' sounds like and where 'ready' is coming from and report the color and number that sound similar to 'ready' and that come from the same loca-

tion as 'ready.'" After each trial, correct-answer feedback was provided. A trial was scored as correct, and subjects were given feedback that they were correct if and only if they reported both target keywords. After each 5 min block, subjects were given the opportunity to take a break.

Each subject completed four sessions on four different days. Each session consisted of the sensitization tests followed by 12 experimental blocks of 50 trials each. Within each 5 min block, the condition (timbre, location, and timbre-and-location) was fixed. Each session contained four blocks of each of the three cue conditions. Four consecutive blocks always had the same cue condition and the same instructions. The four blocks for a given cue condition were presented one after the other within a session, while the ordering of the conditions was separately randomized for each subject and session. Within each session, the selection of the source locations was balanced such that each source location was presented the same number of times. The first session was for training purposes only, and the results of that session were discarded. During the three experimental sessions, subjects performed 600 trials in each cue condition: 200 repetitions for each of three spatial separations in the location condition and 150 repetitions for each of four spatial separations in the timbre and timbre-and-location conditions.

III. RESULTS

A. Percentage responses

Let C_xN_x represent the response color and number, where x denotes when the subject reported a target (T) or a masker (M) keyword. Then responses can be categorized into one of five distinct types: correct responses (C_TN_T), masker errors (C_MN_M), C_TN_M mix errors, C_MN_T mix errors, and drop errors (in which one or both of the reported keywords were not present in either target or masker). The rates at which each of these different response types occurred were calculated separately for each subject, condition, and spatial separation. These rates were then analyzed in primary repeated-measures two-way analysis of variance (ANOVAs) with independent factors of cue condition and spatial separation (excluding the 0° separation, which was not performed in the location condition). When significant interactions between condition and spatial separation were found, separate secondary ANOVAs were run for each of the three cue conditions.

Given the four possible colors and seven possible numbers, the probability of making either a correct response (C_TN_T), masker error (C_MN_M), C_TN_M mix error, or C_MN_T mix error by chance is 1/28 or 4%, while the probability of making a drop error by chance is 84%. However, listeners rarely made drop errors; they nearly always responded with some mixture of keywords from the target and masker, as shown below.

Figure 2 shows the across-subject mean percent correct as a function of spatial separation. Error bars show the 95% confidence intervals around the mean (1.96 times the standard error of the mean across subjects). When listeners were cued about where to listen, overall performance improved with increasing spatial separation (solid and dashed lines in Fig. 2). In contrast, in the timbre condition, performance was essentially independent of spatial separation (dotted lines in Fig. 2). Listeners tended to be better at reporting the target when there were two redundant features (timbre and location) than when there was a single feature to attend to (in Fig. 2, the dashed line is consistently above the solid line, and it is roughly equal to or above the dotted line).

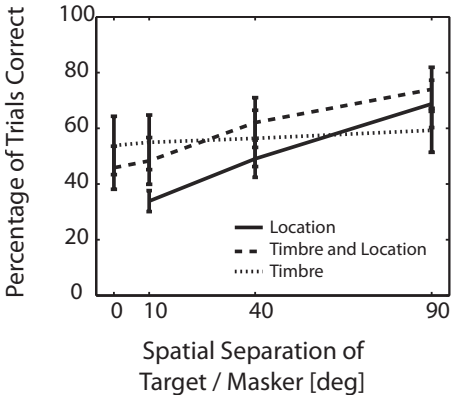


FIG. 2. Percent correct performance as a function of target and masker spatial separation. Performance improves with spatial separation of the target and masker for conditions in which the subjects know the target location ("attend location," solid line; "attend timbre and location," dashed line) but not when they are attending to a nonspatial feature ("attend timbre," dotted line). The across-subject mean in percent correct performance is shown as a function of spatial separation between the target and masker for conditions differing in which target features are known: location, timbre and location, and timbre. Error bars show the 95% confidence intervals around the mean.

The primary ANOVA of percent correct responses found a significant interaction between cue condition and spatial separation [$F(4,32)=25.374$, $p<0.001$], as well as significant effects of both main factors [$F(2,16)=7.599$ (cue condition), 127.683 (spatial separation); $p=0.005$ and $p<0.001$, respectively]. The follow-up ANOVAs found that for the timbre condition, the main effect of spatial separation on percent-correct performance was not significant [$F(2,16)=1.777$, $p=0.201$]. However, for both the location condition and the timbre-and-location condition, spatial separation significantly affected percent-correct performance [$F(2,16)=116.646$ (location) and 77.801 (timbre-and-location); $p<0.001$ for both tests]. The fact that overall performance improved with increasing spatial separation in the location and timbre-and-location conditions confirms that listeners can select a target based on location. In contrast, the fact that performance was independent of spatial separation in the timbre condition supports the idea that spatial separation helps overall performance only when listeners can direct attention to the target location. The spatial gain analysis in Sec. III C (below) considers the effect of spatial separation in more detail.

Figure 3 plots the across-subject average of the different error rates (with error bars showing the 95% confidence intervals around the mean) as a function of the spatial separation between target and masker [(a) masker errors, (b) drop errors, (c) C_TN_M mix errors, and (d) C_MN_T mix errors].

Masker errors (where listeners reported the wrong masker message) decreased with spatial separation in the location and timbre-and-location conditions [solid and dashed

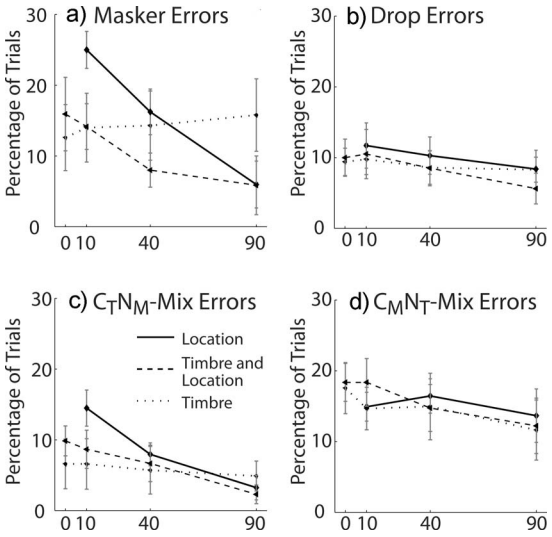


FIG. 3. Masker, C_TN_M mix, C_MN_T mix, and drop errors. Spatial separation reduces masker errors (reporting the masker) when listeners know the target location [solid and dashed lines in (a)] but not when they are instructed to attend the target timbre [dotted lines in (a)]. Spatial separation reduces the mix errors [improves the perceptual segregation of the target and masker over time; see (c) and (d)] and drop errors [decreases the likelihood of reporting words not present in the target or masker; see (b)] in all conditions. In all panels, the across-subject mean is shown as a function of spatial separation between the target and masker for each condition. Error bars show the 95% confidence intervals around the mean.

lines in Fig. 3(a)] but were independent of spatial configuration in the timbre condition [dotted line in Fig. 3(a)]. Overall, masker errors were the most common type of error, particularly in the timbre condition and/or at small spatial separations. Drop errors were relatively rare and decreased with increasing spatial separation for all three conditions [Fig. 3(b)]. C_TN_M mix errors were also relatively uncommon, showing that if a listener heard the first keyword correctly, they were unlikely to switch from target to masker and report the second keyword from the masker [Fig. 3(c)]. In the rare cases in which these errors occurred, their error rates tended to decrease with increasing spatial separation in the location and the timbre-and-location conditions but not in the timbre condition. It was more common to report a wrong initial keyword (report the masker color) and then to switch to the target number [C_MN_T, Fig. 3(d)]: the C_MN_T mix errors tended to decrease with increasing spatial separation in all three conditions [Figs. 3(c) and 3(d)].

The above summaries were generally supported by statistical analyses. For the masker errors [Fig. 3(a)], the primary ANOVA found a significant interaction between cue condition and spatial separation [$F(4,32)=20.193$; $p<0.001$] as well as significant main effects of both spatial separation [$F(2,16)=32.919$, $p<0.001$] and cue condition [$F(2,16)=5.559$, $p=0.04$, with Greenhouse–Geisser correction]. In the secondary ANOVA of the timbre condition, spatial separation did not significantly affect masker errors [$F(2,16)=0.938$, $p=0.377$, with Greenhouse–Geisser correction]. In both the location condition and the timbre-and-location condition, the main effect of spatial separation was significant [$F(2,16)=46.207$ (location) and 12.538 (timbre-and-location); $p<0.001$ and $p=0.001$, respectively].

The primary ANOVA of mix errors [Figs. 3(c) and 3(d)] found significant interaction terms [$F(4,32)=14.744$ (interaction, C_TN_M) and 2.953 (interaction, C_MN_T); $p<0.001$ and $p=0.035$, respectively] and a significant effect of spatial separation for both types of mix error [$F(2,16)=43.331$ (spatial separation, C_TN_M) and 6.43 (spatial separation C_MN_T); $p<0.001$ and $p=0.009$, respectively]. The main effect of cue condition was significant for C_TN_M mix errors [$F(2,16)=4.219$, $p=0.034$] but was not significant for C_MN_T mix errors [$F(2,16)=1.172$, $p=0.335$]. A secondary ANOVA analysis found that spatial separation had a significant effect on both types of mix error in the timbre-and-location condition [$F(2,16)=7.781$ (C_MN_T) and 28.464 (C_TN_M); $p=0.004$ and $p<0.001$, respectively], on C_TN_M mix errors in the location condition [$F(2,16)=49.438$, $p<0.001$], and on C_MN_T mix errors in the timbre condition [$F(2,16)=3.872$, $p=0.043$]. Spatial separation did not have a significant effect on either the C_MN_T mix errors in the location condition [$F(2,16)=1.882$, $p=0.184$ (location, C_MN_T)] or on the C_TN_M mix error in the timbre condition [$F(2,16)=1.39$, $p=0.278$].

A repeated-measure two-way ANOVA on the drop errors [Fig. 3(b)] found a significant effect of spatial separation [$F(2,16)=70.473$, $p=0.001$] and of cue condition [$F(2,16)=4.703$, $p=0.025$] but no significant interaction [$F(4,32)=1.492$, $p=0.228$].

B. Interim discussion

Results suggest a contribution of spatial cues to auditory object formation. In particular, in all conditions, the likelihood that listeners reported a mixture of target and masker words (as if the two messages were not perceptually distinct) tended to decrease with increasing spatial separation of target and masker. This suggests that perceptual separation of the target and masker improves with increasing spatial separation. However, there are a number of alternative explanations that could explain why spatial separation reduces the number of mix errors in all three conditions.

In the conditions in which the listener knows the target location, spatially directed attention can reduce mix errors. In particular, if listeners independently select each keyword based on its location, the probability of selecting both color and number correctly will increase with increasing spatial separation. As a result, both mix and masker errors will decrease (in the limit, if each keyword is selected properly based on its location with probability one, no mix errors will occur). It is difficult to judge from the pattern of response errors alone how much spatially directed attention contributes to the decline of mix errors with increasing spatial separation versus how much of this effect is due to automatic streaming induced by spatial separation, possibly even in the absence of spatially directed attention.

Similarly, in the timbre condition, there are a few possible explanations for the decrease in mix errors with increasing spatial separation that do not assume that spatial separation improves streaming of the target and masker. First, subjects could have been biased in their responses and attended to one side of space (for instance, by picking the

words that are better represented in the right ear). However, we separately analyzed the responses for leftward and rightward targets and did not find any consistent spatial bias for any of the subjects. Second, listeners could have simply responded by reporting the color and number keywords that were closest in space to the cue word "ready." However, there were no biases of this sort in the responses of any subjects. Third, listeners could arbitrarily pick either the target or the masker color and report the number from the same location. While the first two possibilities are not supported by the data, we cannot conclusively rule out this third possibility. However, listeners were specifically instructed to attend to the target timbre, not to the location. Moreover, listeners were able to select the proper keywords based on their timbre, as proven by the high percentage of trials in which they reported both keywords correctly in the timbre condition. Thus, on trials in which listeners were actively trying to attend to timbre, either (1) spatial separation helped listeners to properly stream the target and masker or (2) in the trials where attention to timbre failed, listeners attended to location, instead.

Even if listeners were sure that the color they reported was from the masker stream, there was little external motivation for them to switch to reporting the target color. Whether they made a switch (resulting in a mix error) or not (giving a masker error), they would receive a "wrong" score. However, listeners did make mix errors. Moreover, the patterns of these mix errors varied systematically with cue condition and spatial separation. For all three conditions, $C_M N_T$ mix errors were far more likely than $C_T N_M$ mix errors (or $C_T N_X$ drop errors). This suggests that even without being explicitly rewarded for reporting one of the two target keywords, listeners adopted a response strategy in which they reported as many target keywords as possible. In particular, the asymmetry in mix errors suggests that listeners often realized when they reported the wrong color and switched their attention to the target stream and reported the proper number. In the location and timbre-and-location conditions, masker errors were less likely than $C_M N_T$ mix errors at large spatial separations, and the likelihood of masker errors decreased with increasing spatial separation. This result suggests that spatial cues allowed the listener to detect their initial error and switch attention between streams when they knew the target location. In contrast, masker errors were roughly equally as likely as $C_M N_T$ errors for all spatial separations in the timbre condition, suggesting that when listeners attended to timbre, the spatial separation between target and masker was relatively unlikely to help them correct any initial error if they incorrectly reported the masker color.

In all three conditions, increasing spatial separation reduced drop errors [reporting words not from the target or the masker, Fig. 3(b)], suggesting that spatial separation also improves target audibility and/or short-term segmentation of auditory objects at the level of syllables, even when attention is not spatially directed.

C. Spatial gains

The influence of spatial separation on performance varied from subject to subject; however, an analysis of individual results shows that spatial separation influences performance in a consistent way across the population of subjects. For each subject in each condition and response category, we computed spatial gains. For percent correct performance, the spatial gain was computed as the increase in the probability of responding correctly when target and masker separation increased from 10° to 90°. Spatial gains in the error conditions were computed as the decrease in the percentage of each type of error as target and masker separation increased from 10° to 90°. From the resulting distribution of spatial gains, we computed the across-subject 95% confidence intervals of the mean spatial gain to determine whether there was a consistent effect of spatial separation across the subject population (see solid horizontal and vertical bars near the *x*- and *y*-axes in Fig. 4).

Figure 4 directly compares the spatial gains for different combinations of cue conditions for each individual subject (shown as individual points in each plot). Each row shows results for a different aspect of performance (overall percent

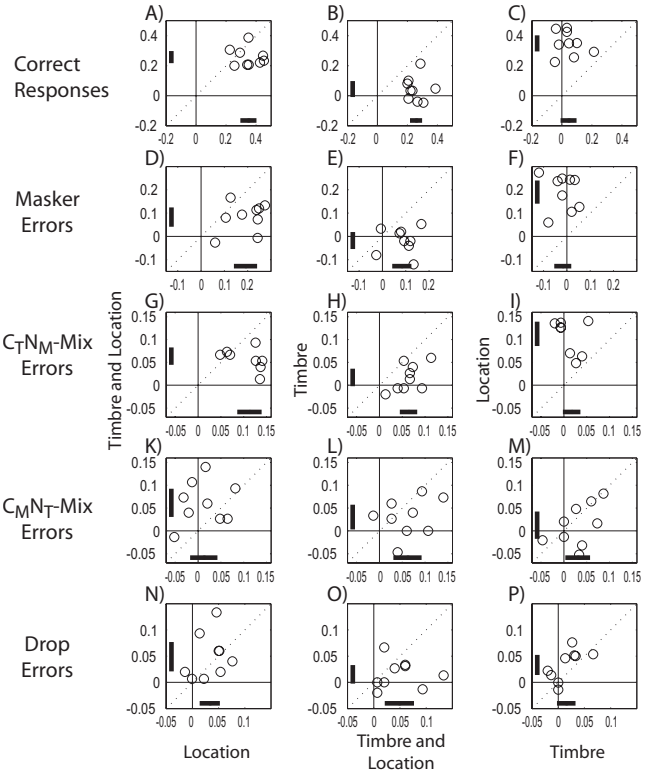


FIG. 4. Individual subject spatial gains (see text for definition) in correct performance, and masker, $C_T N_M$ mix, $C_M N_T$ mix, and drop errors contrasted across task conditions. Individual subjects all show large improvements in overall performance with spatial separation of the target and masker when they know the target location but not when they are attending to the target timbre. Overall, spatial separation of target and masker reduces response errors of all kinds (produces positive spatial gains); however, when listeners attend to the timbre, there is no significant reduction in masker errors. Each row compares the reduction in errors with spatial separation for one type of response error (correct performance, masker, $C_T N_M$ mix, $C_M N_T$ mix, and drop errors in top, second, third, fourth, and bottom rows, respectively). Within each panel, each point compares results for one of the nine individual subjects across two different task conditions. The horizontal and vertical bars within each panel show the 95% confidence intervals for the group mean of the spatial gain in the corresponding dimension. Left column: timbre-and-location vs location. Center column: timbre vs timbre-and-location. Right column: location vs timbre.

correct, masker errors, mix errors, and drop errors, respectively, from top to bottom in the figure). In each panel within a row, the spatial gains in two cue conditions are plotted against each other to allow an assessment of the relative size and direction of the spatial effects in the different cue conditions.

The absolute magnitude of the spatial gains differed from subject to subject (within each panel in Fig. 4, the 95% confidence intervals are on the order of 10%; see horizontal and vertical bars near abscissas and ordinates). However, there were consistent patterns in the relative size of the spatial gains in the different conditions across subjects (within each panel in Fig. 4, circles tend to cluster within one octant).

The spatial gain in overall performance is significantly greater than zero for all subjects and roughly equal in the location and timbre-and-location conditions [data fall near the diagonal and above and to the right of the origin in Fig. 4(a)], but there is no significant spatial gain in percent correct for the timbre condition [see Figs. 4(b) and 4(c), where spatial gains in the timbre condition cluster near zero].

The size of the spatial gains in the errors depends strongly on which target attribute the listener is instructed to attend to. The reduction in masker errors with spatial separation is consistently larger for location than for timbre-and-location conditions [data generally fall below the diagonal in Fig. 4(d)], while in the timbre condition there is no consistent spatial gain [spatial gains in the timbre condition in Fig. 4(e), vertical axis, and Fig. 4(f), horizontal axis, cluster near zero]. Spatial gains for mix errors tend to be positive in all conditions [most of the data points in Figs. 4(g)–4(m) are positive]. However, the size of this gain depends on the cue condition. The spatial gain is smaller in the timbre-and-location condition than in the location condition [data fall below the diagonal in Fig. 4(g)] and smaller for $C_T N_M$ mix errors in the timbre condition than in the timbre-and-location condition [data fall below the diagonal in Fig. 4(h)]. The spatial gains for $C_M N_T$ errors tend to be largest for the timbre-and-location condition [data tend to fall above the diagonal in Fig. 4(k) and below the diagonal in Fig. 4(l)] and slightly larger for the timbre condition than for the location condition [data tend to fall below the diagonal in Fig. 4(m)], although these trends are less consistent across subjects than the trends for the $C_T N_M$ mix errors. Finally, the spatial gains for drop errors are generally greater than zero for all three conditions [in Figs. 4(n)–4(p), data points tend to be positive] and are comparable in the different conditions [data in Figs. 4(n)–4(p) tend to fall around the diagonal].

IV. CONDITIONAL RESPONSE PROBABILITIES

In order to better understand how much of these spatial effects could be accounted for simply through spatially directed attention versus what improvements may come from automatic improvements in the perceptual segregation of the messages with spatial separation even when attention is not explicitly spatially directed, we analyzed conditional response probabilities (see the diagram in Fig. 5). This probabilistic analysis determined whether the color and number

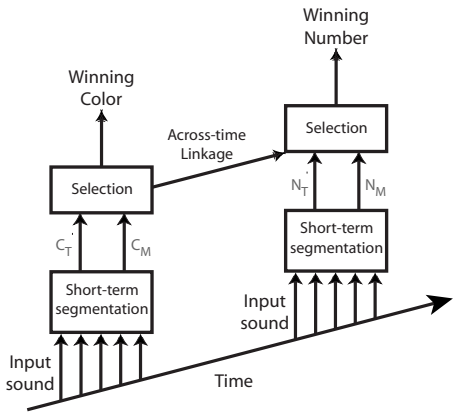


FIG. 5. Illustration of the model underlying the probabilistic analysis. Spatial cues may affect speech identification both through spatially directed attention and through automatic across-time linkage.

responses were independent of one another or whether the probability of responding with the correct number depended on whether listeners reported the target color (see also Cusack *et al.*, 2004). Specifically, we were interested in whether increasing the spatial separation between target and masker not only increases the likelihood of selecting the correct color and number when listeners know where the target is located, but also increases the probability of properly linking the keywords in the target and in the masker (increasing the perceptual segregation of the target and masker streams).

Target selection and across-time linkage of keywords are likely to occur at the same time, influencing each other. If each utterance is perceived correctly as one stream, the listener may only need to make one decision and report either both of the target words or both of the masker words. Instead, some mix errors occur, suggesting that, at least on some trials, the listener (1) makes two independent selections (selecting a color and then, separately, selecting a number), (2) decides to switch to the other stream upon hearing the color (i.e., decides, possibly incorrectly, that she was listening to the masker stream and therefore switches to the other stream), or (3) does not stream the target and masker properly and therefore makes a single decision, but the 'stream' she chooses to report is a mixture of target and masker.

If a listener independently selects color and number based on top-down attention and if there is no influence of across-time streaming (1, above), the initial choice of what color to report will be statistically independent of the second choice (the number reported). In other words, the probability of making a correct choice for the number will be the same for a given condition and spatial separation, independent of the color reported.[6] In contrast, if there is some automatic streaming of color and number across time, the probability of answering with the correct number should depend on whether or not the listener selected the correct color, regardless of what strategy the listener adopts (i.e., reporting the correct stream, switching attention to the other stream after reporting the color, or attending to an improperly formed stream). However, this analysis is not definitive. Dependencies between performance on color and number can arise for other reasons. Listeners' attention may lapse on some trials, so that the probability of missing the color and number both

increase together in those trials. Some trials may be inherently easier than others, even though they are analyzed together (for instance, a spatial separation of 10° may provide a stronger perceptual cue when target and masker are on the front rather than on the side). If so, the probability of being correct on the color and number will also be dependent. Nonetheless, it is worth examining whether such dependencies exist.

These ideas can be formulated through a simple probabilistic analysis. In general, whether or not color and number selections are independent,

$$P(N_T|C_T) = P(C_T N_T)/P(C_T), \quad (1)$$

where $P(N_T|C_T)$ is the conditional probability of reporting the target number in those trials where the target color was reported, $P(C_T N_T)$ is the probability of responding correctly, and $P(C_T)$ and $P(N_T)$ are the marginal probabilities of reporting the target color and number, respectively. Analogously, it is generally true that

$$P(N_T|C_M) = P(C_M N_T)/P(C_M). \quad (2)$$

If the selection of the number keyword is independent of the selection of the color keyword, then

$$P(C_T N_T) = P(C_T)P(N_T). \quad (3)$$

This also implies that when color and number choices are independent,

$$P(N_T|C_T) = P(N_T) = P(N_T|C_M). \quad (4)$$

Thus, if $P(N_T|C_T)$ is greater than $P(N_T|C_M)$ for a given stimulus configuration, this reflects a bias to report both target keywords over reporting a mix of target and masker keywords. The difference between $P(N_T|C_T)$ and $P(N_T|C_M)$ is expected to increase with increasing strength of perceptual across-time continuity, or streaming, between target color and target number.

For each subject, condition, and spatial separation, $P(N_T|C_T)$ and $P(N_T|C_M)$ were estimated from the observed percentages of responses. Figure 6 plots the means of the individual subject estimates of $P(N_T|C_T)$ and $P(N_T|C_M)$ as a function of the spatial separation between target and masker for the various conditions (dotted and dashed lines, respectively).

For all cue conditions, on trials when listeners properly reported the target color, listeners were also very likely to report the number from the target stream [solid lines are consistently above 50% in Figs. 6(a)–6(c)]. In contrast, listeners are much less likely to report the target number if they reported the masker color [dashed-dotted lines are between 36% and 70% in Figs. 6(a)–6(c)]. The mean difference between $P(N_T|C_T)$ and $P(N_T|C_M)$, averaged across the three spatial separations from 10° to 90° and averaged across listeners, is 40.1%, 29.7%, and 25.6% in the timbre, location, and timbre-and-location conditions, respectively. Thus, Eq. (4) is violated at each spatial separation and in each cue condition, proving that the color and number reports are not independent. Instead, the likelihood of reporting the target number depends on whether listeners reported the target or the masker color, with listeners more likely to get the target

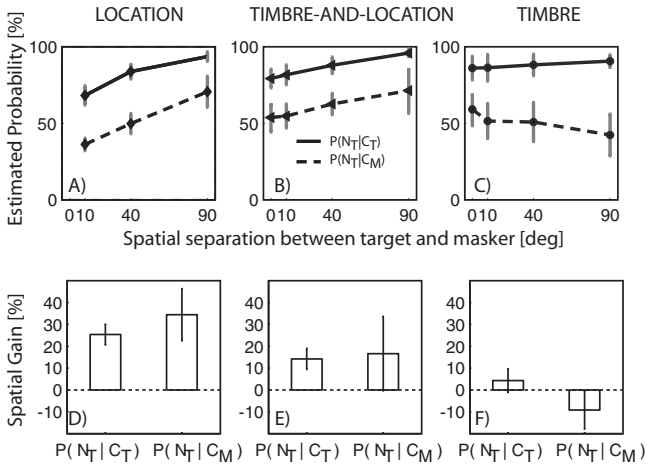


FIG. 6. Results of the probabilistic analysis from Eqs. (1)–(4). [(A)–(C)] Across-subject mean of the conditional probabilities, as a function of spatial separation (error bars show the standard error of the mean). [(C)–(E)] Across-subject mean of the percent spatial gain in conditional probabilities, respectively, with spatial separation for the three conditions (error bars show 95% confidence intervals).

number correct when they correctly report the target color. This suggests that listeners are likely to stay with the same stream for both color and number, either because they heard color and number from one source as a single perceptual unit or because the stimulus conditions were favorable for reporting both target keywords, regardless of whether they were streamed or not.

When listeners are told to attend to the target location, $P(N_T|C_T)$ and $P(N_T|C_M)$ both increase as the spatial separation between target and masker increases. In contrast, when listeners attend to the target timbre, the likelihood of reporting the target number is essentially independent of spatial separation or even tends to decrease with increasing spatial separation [$P(N_T|C_M)$, dashed line in Fig. 6(c)]. For each subject, we computed the difference in $P(N_T|C_T)$ when sources were 90° apart minus $P(N_T|C_T)$ when sources were 10° apart, and the corresponding effect of spatial separation on $P(N_T|C_M)$. The across-subject averages of these differences are shown in Figs. 6(d)–6(f) (error bars show the 95% confidence intervals).

In the location condition, spatial separation significantly increases the probability that listeners select the proper target number, both when they reported the target color and when they reported the masker color [both gains are positive in Fig. 6(d)]. In the timbre-and-location condition, spatial separation also improves the likelihood of reporting the proper color, but the effects are smaller, perhaps in part because of ceiling effects [in Fig. 6(e), gains are also positive]. In contrast, spatial separation has little effect on the probability of reporting the proper number when listeners attend to timbre [in Fig. 6(f), gains are not significantly different from zero], as if spatial separation improves streaming and makes it less likely that listeners switch streams.

Together, these results suggest that when attending to the target location (both in the location and the timbre-and-location conditions) and when target and masker are close together, listeners report the target and masker color and number in proper pairs and have difficulty detecting when

they made a wrong color selection. As a result, at small spatial separations, they fail to switch to the target stream when they initially listen to the masker. In contrast, in the timbre condition, when target and masker are close together, listeners are better able to select between the target and masker, presumably using nonspatial features, and are also more likely to switch streams when there is an initial error. This trend changes when target and masker are well separated in space: when listeners know the location of the target utterance (in the location and the timbre-and-location conditions), they are more likely to realize that they have reported the masker color and switch over to the target message than when they are attending to the target timbre. While this analysis cannot provide direct proof that the keywords of each source are more likely to be perceived in well formed streams when the competing sources are spatially separated, it is interesting to note that the conditional probabilities show patterns that differ across cue conditions. Future studies are needed to study more directly the automatic influence of spatial separation on object formation in the absence of spatially directed attention.

V. DISCUSSION

Most past studies of the effects of spatial cues on auditory perception show that listeners are more likely to report the target message correctly when competing sources arise from different directions compared to when they are at the target location (Bronkhorst, 2000). Multiple factors contribute to this effect, as considered below.

If a salient nonspatial acoustic attribute (e.g., timbre) differentiates the target from the masker (Brungart *et al.*, 2001; Darwin *et al.*, 2003) and the main form of interference is "energetic masking," spatial separation reduces peripheral interference between the target and masker and improves the effective target-to-masker energy ratio (Zurek, 1993; Shinn-Cunningham, 2005). This spatial effect does not require spatially directed attention (Edmonds and Culling, 2005a; Edmonds and Culling, 2005b; Culling *et al.*, 2006; Allen *et al.*, 2008) and appears to operate primarily by making it possible to detect near-threshold portions of the target (Shinn-Cunningham, 2005; Zurek, 1993).

Many studies point out that spatial cues carry little weight in enabling the segregation of sound locally in time at the level of syllables (Kubovy, 1981; Culling and Summerfield, 1995a; Darwin, 1997; Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007). While spatial cues provide little aid in segregating sources at a local time scale, spatial separation of competing sources improves the ability to selectively attend to a target when target location is the main cue differentiating the target from the other sources in the mixture (Freyman *et al.*, 2001; Gallun *et al.*, 2005; Shinn-Cunningham *et al.*, 2005a; Ihlefeld and Shinn-Cunningham, 2008). However, no past study has directly addressed whether spatial cues only allow a listener to select the proper object from the sound mixture or whether spatial separation automatically causes the competing messages to group properly across time and form more distinct auditory streams. Indeed, in discussing the effects of spatial cues on listening in complex environments, studies typically either acknowledge only one of these possibilities or use language that confounds the two.

The current results indicate that reports of color and number keywords are statistically dependent in all conditions. Moreover, when listeners were told to attend to the target location, the likelihood of selecting the proper keywords increased with increasing spatial separation for both the color and the number reports. Performance in the location and timbre-and-location conditions was biased toward reporting color and number keywords of one stream [as shown by differences in the conditional probabilities plotted in Figs. 6(a) and 6(b)]. An increase in the efficacy of spatially directed attention with increasing spatial separation between target and masker could account for the pattern of responses in these conditions. In contrast, when listeners were told to select the target based on a nonspatial feature, the likelihood of switching between streams decreases with increasing spatial separation, causing both an increase in masker errors and a decrease in mix errors. This suggests that the perceptual separation between the target stream and the masker stream increases with spatial separation (and perhaps that streaming itself becomes stronger as target and masker are displaced from each other in space). Of course, if timbre cues had been harder or easier to discern, the observed spatial effects might have differed. However, the current results show that when listeners were able to use timbre cues but were not perfect at it, spatial differences between target and masker influenced responses even when attention was not spatially directed. Overall, the effects of spatial separation were much smaller in the timbre condition than in the other two cue conditions. To the extent that these differences in the conditional probabilities are indicative of streaming, we conclude that the dominant contribution of spatial cues to understanding sources in a complex scene (with little energetic masking) comes through spatially directed attention, not through improvements in auditory streaming. As noted above, other possibilities for the dependency between color and number reports could also contribute to this effect. Further experiments are necessary to definitively address this question.

Recent physiological evidence suggests that spatial attention can modulate midbrain sensory responses (Winkowski and Knudsen, 2006). This result hints that sensory representations are altered by spatially directed attention in a manner that will tend to enhance the representation of a source at a desired location. This physiological mechanism could account for the observed psychophysical improvements in performance when listeners attended to a target at a known location.

The current results confirm that spatial attention can be directed toward the known location of a target, increasing the likelihood that the desired target source is selected and brought to the attentional foreground. Specifically, we see that spatially directed attention causes overall performance to improve with increasing spatial separation between target and masker when listeners attend to location or both timbre and location (Fig. 2). The importance of this selective spatial attention is greatest when space is the only acoustic feature

that a listener can use to identify the target (e.g., the spatial gain for masker errors is greater in the location and timbre-and-location conditions than in the timbre condition). When other features also aid in source selection, spatial information becomes wholly or partially redundant and thus less influential on performance.

It is worth noting that although having multiple cues identifying the target (timbre and location) reduces the influence of spatial configuration on performance, both of the redundant features contribute to the ability to report the target message (overall percent correct is generally higher for the timbre-and-location condition than for the corresponding single-feature conditions; see Fig. 2). At first glance, this result may seem at odds with visual theories that suggest that conjunctions of features do not provide large performance benefits (Treisman and Gelade, 1980; Wolfe and Bennett, 1997). However, most visual studies measure the time it takes to search for and detect a target in a complex visual scene. In contrast, in the current study, the auditory messages must be attended to and processed over time, so that the main factors affecting performance are the degree to which the target and masker are perceptually separated and how well a listener can maintain attention on the target, not how rapidly the target can be detected.

Among vision researchers, it has long been recognized that objects vie for attention in a complex scene and that top-down selection works in concert with bottom-up stimulus salience to determine which object will be processed and perceived (Desimone and Duncan, 1995; O'Craven *et al.*, 1999; Scholl, 2001; Serences and Yantis, 2006). A similar ability to selectively attend to a desired auditory stream normally enables communication in complex settings, where there are multiple talkers vying for attention (Shinn-Cunningham, 2008). This may help explain why listeners using hearing aids or wearing cochlear implants often find communication relatively easy in one-on-one settings but frequently experience communication difficulties or even communication breakdown in social settings such as at a restaurant (Gatehouse and Noble, 2004; Noble and Gatehouse, 2006; Harkins and Tucker, 2007). In a one-on-one conversation, there is no need to segregate the target source from a confusing sound mixture in order to process and understand it. However, in a noisy setting, sources must be segregated so that selective attention can be directed to whichever object is to be processed. In general, the acoustic cues critical for object formation (such as fine spectrotemporal resolution, robust timing information in the neural response in the auditory nerve, etc.) are degraded or absent in the signals many impaired listeners receive. These listeners may not be able to segregate and stream the sound mixture properly in complex settings and therefore may not be able to selectively attend to a desired sound source. Supporting this view, a high percentage of hearing-aid users is dissatisfied with their aids (Kochkin, 2005; but see also Edwards, 2007). Such descriptions are consistent with an inability to selectively attend to a desired source. This realization underscores the importance of further studies into the roles that various acoustic features (including location) play in both forming auditory objects and directing auditory attention.

VI. CONCLUSIONS

When trying to understand sources in a complex scene (with little energetic masking), spatial differences between target and masker improve the ability to select the target source from the mixture only when spatial location defines which object is the target. However, spatial separation affects listeners' responses as if spatial continuity helps to form streams even when attention is not spatially directed. The dominant contribution of spatial cues to listening selectively in a sound mixture comes through spatially directed attention, not through improvements in auditory streaming. Future work is needed to further delineate the different ways in which spatial cues affect object formation and object selection when listening in a sound mixture.

ACKNOWLEDGMENTS

Grants from the Office of Naval Research and the Air Force Office of Scientific Research to B.S.-C. supported this work. Jyrki Ahveninen, Virginia Best, Robert Carlyon, Steven Colburn, Frederick Gallun, Gerald Kidd, Nicole Marrone, Christine Mason, Richard Freyman, and three anonymous reviewers gave helpful feedback on earlier versions of this manuscript.

[1]Note that computational models of binaural processing can predict spatial release from masking for tasks dominated by energetic masking (e.g., Zurek, 1993). However, less is known about the role of spatial cues when informational masking limits performance (Kidd *et al.*, 2008). Here, we designed our stimuli to emphasize the role of informational masking and to de-emphasize the role of energetic masking (see also Arbogast *et al.*, 2002; Kidd *et al.*, 2005b; Gallun *et al.*, 2005; Brungart *et al.*, 2005; Shinn-Cunningham *et al.*, 2005a; Ihlefeld and Shinn-Cunningham, 2008).

[2]Each color-number pair and each cue word were carefully time windowed from recordings of talker 0 in the original corpus, using a routine programed in MATLAB 6.5 (both time-domain and short-term Fourier transform representations were used to monitor the quality of the resulting signal). Each color is preceded by the vowel /ə/ from "go to." This change from harmonic to inharmonic structure made it fairly easy to classify the beginning of the utterance. Similarly, each color in the CRM corpus is followed by "now," a syllable whose energy builds up slowly over time. Thus, even if a small bit of /n/ was left attached to the number, the intelligibility of the number was not adversely affected.

[3]The resulting spacing of the filters was roughly 42% of an octave. The center frequencies were 250, 333, 445, 593, 791, 1056, 1408, 1878, and 2505 Hz.

[4]In general, HRTFs differ across individual listeners. However, in the horizontal plane, the spatially dominant interaural difference cues in HRTFs tend to be grossly similar across listeners, allowing virtual spatial acoustic simulations that evoke sources at different lateral angles relative to the listener without using individualized HRTFs (Colburn and Kulkarni, 2005; Middlebrooks, 1999; Middlebrooks *et al.*, 2000). In the current study, all listeners perceived the sounds as having distinct lateral positions. However, absolute target and masker locations, not just angular separation, affect target intelligibility, e.g., due to differences in acoustic head shadow effects (Zurek, 1993; Shinn-Cunningham *et al.*, 2005a; Ihlefeld and Shinn-Cunningham, 2008). The perceptual difference between a target at 0° and a masker at 10° is not equivalent to that of a target at 80° and a masker at 90°. Because we collapsed performance across many different spatial source configurations with different better-ear effects, it is difficult to estimate how much performance should have improved with spatial separation due solely to the acoustic better-ear effects. However, statistically, the spatial stimulus features were the same across the three timbre, location, and timbre-and-location conditions. Therefore, performance in all three conditions was affected in similar ways by these better-ear acoustic advantages. Although it is desirable to measure performance as a function of both spatial separation between target and masker and their absolute locations, this would require considerably more data to be collected than was

done in the current study. To the extent that we find robust effects of spatial separation on performance, this averaging across different absolute locations is a source of uncontrolled variability. Thus, any conclusions we draw about the influence of spatial separation on performance are relatively conservative.

[5]The location cues used in the current study are much more clearly defined than the timbre cues. While the saliency of the timbre cues is difficult to establish, listeners could easily distinguish between the six different timbres during the sensitization tasks at the beginning of each session. Moreover, listeners achieved better-than-chance scores when attending only to timbre (see Sec. III).

[6]In general, it is difficult to directly compare number and color errors because the number of tokens, perceptual similarity of the tokens, and other attributes differ in the two sets. However, given how infrequent drop errors are, we believe that listeners are usually making binary decisions between responding with the target or the masker color and number. As a result, the decisions about which color and which number to report are probably more equal than one might expect based solely on the stimulus properties.

Allen, K., Carlile, S., and Alais, D. (**2008**). "Contributions of talker characteristics and spatial location to auditory streaming," J. Acoust. Soc. Am. **123**, 1562–1570.

Arbogast, T. L., and Kidd, G., Jr. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Best, V., Gallun, F. J., Ihlefeld, A., and Shinn-Cunningham, B. G. (**2006**). "The influence of spatial separation on divided listening," J. Acoust. Soc. Am. **120**, 1506–1516.

Bolia, R. S., Nelson, W. T., and Ericson, M. A. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acustica **86**, 117–128.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2004**). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. **115**, 301–310.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd, G., Jr. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Busse, L., Roberts, K. C., Christ, R. E., Weissman, D. H., and Woldorff, M. G. (**2005**). "The spread of attention across modalities and space in a multisensory object," Proc. Natl. Acad. Sci. U.S.A. **102**, 18751–18756.

Colburn, H. S., and Kulkarni, A. (**2005**). "Models of sound localization," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York).

Culling, J. F., Edmonds, B. A., and Hodder, K. I. (**2006**). "Speech perception from monaural and binaural information," J. Acoust. Soc. Am. **119**, 559–565.

Culling, J. F., and Summerfield, Q. (**1995a**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Culling, J. F., and Summerfield, Q. (**1995b**). "The role of frequency modulation in the perceptual segregation of concurrent vowels," J. Acoust. Soc. Am. **98**, 837–846.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1994**). "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," Speech Commun. **14**, 71–95.

Cusack, R., Carlyon, R. P., and Robertson, I. H. (**2000**). "Neglect between but not within auditory objects," J. Cogn Neurosci. **12**, 1056–1065.

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. (**1997**). "Auditory grouping," Trends Cogn. Sci. **1**, 327–333.

Darwin, C. (**2008**). "Spatial Hearing and Perceiving Sources," in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, edited by W. A. Yost (Springer, New York), Vol. **29**.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.

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. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970–977.

Desimone, R., and Duncan, J. (**1995**). "Neural mechanisms of selective visual attention," Annu. Rev. Neurosci. **18**, 193–222.

Deutsch, D. (**1999**). "Grouping mechanisms in music," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, San Diego).

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Edmonds, B. A., and Culling, J. F. (**2005a**). "The role of head-related time and level cues in the unmasking of speech in noise and competing speech," Acta Acust. **91**, 546–553.

Edmonds, B. A., and Culling, J. F. (**2005b**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.

Edwards, B. (**2007**). "The future of hearing aid technology," Trends Amplif. **11**, 31–45.

Egly, R., Driver, J., and Rafal, R. D. (**1994**). "Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects," J. Exp. Psychol. Gen. **123**, 161–177.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R., Helfer, K., and Balakrishnan, U. (**2005**). "Spatial and spectral factors in release from informational masking in speech recognition," Acta Acust. **91**, 537–545.

Gallun, F. J., Mason, C. R., and Kidd, G., Jr. (**2005**). "Binaural release from informational masking in a speech identification task," J. Acoust. Soc. Am. **118**, 1614–1625.

Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial, and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.

Harkins, J., and Tucker, P. (**2007**). "An internet survey of individuals with hearing loss regarding assistive listening devices," Trends Amplif. **11**, 91–100.

Ihlefeld, A., and Shinn-Cunningham, B. G. (**2008**). "Spatial release from energetic and informational masking in selective listening," J. Acoust. Soc. Am. **123**, 4369–4379.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. (**2005a**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005b**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.

Kidd, G., Jr., Mason, C. R., Richards, V., Gallun, F. J., and Durlach, N. I. (**2008**). "Informational masking," in *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*, edited by W. A. Yost (Springer, New York), Vol. **29**.

Knudsen, E. (**2007**). "Fundamental components of attention," Annu. Rev. Neurosci. **30**, 57–78.

Kochkin, S. (**2005**). "Customer satisfaction with hearing instruments in the digital age," Hear. J. **58**, 30–39.

Kubovy, M. (**1981**). "Concurrent-pitch segregation and the theory of indispensable attributes," in *Perceptual Organization*, edited by M. Kubovy and J. R. Pomerantz (Lawrence Erlbaum, Associates, NJ), pp. 55–98.

Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C., Macpherson, E. A., and Onsan, Z. A. (**2000**). "Psychophysical customization of directional transfer functions for virtual sound localization," J. Acoust. Soc. Am. **108**, 3088–3091.

Noble, W., and Gatehouse, S. (**2006**). "Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ)," Int. J. Audiol. **45**, 172–181.
O'Craven, K. M., Downing, P. E., and Kanwisher, N. (**1999**). "fMRI evidence for objects as the units of attentional selection," Nature (London) **401**, 584–587.
Pashler, H. (**1998**). *Attention (Studies in Cognition)* (Psychology, Hove, UK).
Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.
Scholl, B. J. (**2001**). "Objects and attention: The state of the art," Cognition **80**, 1–46.
Serences, J. T., Liu, T., and Yantis, S. (**2005**). "Parietal mechanisms of switching and maintaining attention to locations, objects, and features," in *Neurobiology of Attention*, edited by L. Itti, G. Rees, and J. Tsotsos (Academic, New York), pp. 35–41.
Serences, J., and Yantis, S. (**2006**). "Selective visual attention and perceptual coherence," Trends Cogn. Sci. **10**, 38–45.
Shinn-Cunningham, B. G. (**2005**). "Influences of spatial cues on grouping and understanding sound," in Forum Acusticum 2005, Budapest, Hungary, p. CD.
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. (**2005a**). "Bottom-up and top-down influences on spatial unmasking," Acta 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 responses," J. Acoust. Soc. Am. **117**, 3100–3115.
Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (**2007**). "Auditory nonallocation of a sound element lost in perceptual competition," Proc. Natl. Acad. Sci. U.S.A. **104**, 12223–12227.
Shomstein, S., and Yantis, S. (**2004**). "Control of attention shifts between vision and audition in human cortex," J. Neurosci. **24**, 10702–10706.
Treisman, A. M., and Gelade, G. (**1980**). "A feature-integration theory of attention," Cognit Psychol. **12**, 97–136.
Winkowski, D. E., and Knudsen, E. I. (**2006**). "Top-down gain control of the auditory space map by gaze control circuitry in the barn owl," Nature (London) **439**, 336–339.
Wolfe, J. M., and Bennett, S. C. (**1997**). "Preattentive object files: Shapeless bundles of basic features," Vision Res. **37**, 25–43.
Wolfe, J. M., Butcher, S. J., Lee, C., and Hyle, M. (**2003**). "Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons," J. Exp. Psychol. Hum. Percept. Perform. **29**, 483–502.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hochberg (College-Hill, Boston, MA).

Influences of modulation and spatial separation on detection of a masked broadband target[a)]

Norbert Kopčo[b)] and Barbara G. Shinn-Cunningham[c)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 22 June 2007; revised 24 June 2008; accepted 9 July 2008)

Experiments explored the influence of amplitude modulation and spatial separation on detectability of a broadband noise target masked by an independent broadband noise. Thresholds were measured for all combinations of six spatial configurations of target and masker and five modulation conditions. Masker level was either fixed (Experiment 1) or roved between intervals within a trial to reduce the utility of overall intensity as a cue (Experiment 2). After accounting for acoustic changes, thresholds depended on whether a target and a masker were colocated or spatially separated, but not on the exact spatial configuration. Moreover, spatial unmasking exceeded that predicted by better-ear acoustics only when modulation cues for detection were weak. Roving increased the colocated but not the spatially separated thresholds, resulting in an increase in spatial release from masking. Differences in both how performance changed over time and the influence of spatial separation support the idea that the cues underlying performance depend on the modulation characteristics of the target and masker. Analysis suggests that detection is based on overall intensity when target and masker modulation and spatial cues are the same, on spatial attributes when sources are separated and modulation provides no target glimpses, and on modulation discrimination in the remaining conditions. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967891]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Rq, 43.66.Mk [RLF] Pages: 2236–2250

I. INTRODUCTION

The extent to which one sound source masks another depends to a large degree on how similar the two sources are in characteristics such as their spectral profile, temporal structure, and spatial location. While a fair amount is known about how these individual characteristics affect the ability to detect and understand a masked target, relatively little is known about how these characteristics interact. In everyday situations, listeners often are faced with the task of understanding one complex, fluctuating signal in the presence of similar, complex signals from different locations, such as understanding one talker in the presence of competing talkers. If we are ever to understand perception in everyday situations, we must explore how source characteristics such as spectral content, amplitude fluctuations over time (modulation), and spatial location jointly affect perception.

This paper considers the individual and combined effects of two stimulus characteristics: modulation structure and spatial location. *A priori*, one might imagine that the two variables are redundant with one other, so that there is no added benefit when spatial cues in a target and a masker differ if they already differ in their modulation structure (and vice versa). Alternatively, it is possible that masking effects related to temporal modulation and spatial location are largely independent of one another and that effects of the two attributes are additive. Finally, it is possible that differences in temporal modulations actually facilitate the effectiveness of spatial cues in releasing masking, or vice versa, resulting in superadditivity of their individual effects. This study investigates these alternative possibilities using a detection task with simple broadband noise targets and maskers by manipulating both temporal and spatial characteristics independently and jointly.

Several previous studies looked at spatial release from masking (SRM) for nonspeech stimuli that fluctuated over time. The target stimuli in these studies ranged widely, including click trains (Saberi *et al.*, 1991; Gilkey and Good, 1995; Good *et al.*, 1997), chirp trains (Lane *et al.*, 2004; Kopco, 2005), and pulsed 1/3-octave bands of noise (Zurek *et al.*, 2004). However, none of these studies looked at how modulation influences SRM.

Other studies examining the relationship between modulation and spatial processing in masked detection tasks differed substantially in approach and the specific questions addressed, making it difficult to compare results across studies. For example, some explored comodulation and binaural masking release (van de Par and Kohlrausch, 1998; Hall *et al.*, 2006) while others looked at monaural and interaural level discrimination (Stellmack *et al.*, 2005), the interaction between modulation detection interference and spatial processing (Sheft and Yost, 1997), or the equivalence of binaural processing of low-frequency fine time structure versus high-frequency envelope structure (Bernstein and Trahiotis, 1994; van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002). Physiological data from the cat inferior colliculus (IC) suggest that binaural cues in the temporal envelope contribute to SRM (Sterbing *et al.*, 2003; Lane and Delgutte, 2005). However, some psychophysical studies sug-

[a)]Portions of this work were presented at the 149th and 151st meetings of the Acoustical Society of America.
[b)]Permanent Address: Department of Cybernetics and AI, Technical University, Košice, Slovakia. Electronic mail: kopco@bu.edu
[c)]Electronic mail: shinn@bu.edu

 0001-4966/2008/124(4)/2236/15/$23.00

gest that the stimulus temporal envelope does not affect SRM. For example, binaural detection thresholds obtained for a harmonic tone complex and broadband noise targets are very similar, despite dramatic differences in their envelopes (van de Par *et al.*, 2004). Overall, these studies do not provide a consistent account of how spatial cues and modulation jointly affect detection of a target embedded in noise.

Some work suggests that the influence of modulation on masked target detection depends on whether the target or the masker is modulated. For example, when listeners must detect a target embedded in maskers, reaction times depend less strongly on the number of distractors when the target is amplitude modulated and the maskers are unmodulated than when the target is a pure tone and the maskers are amplitude modulated (Asemi *et al.*, 2003). This asymmetry suggests that the modulated target is more likely to "pop out" of the background of unmodulated maskers than the reverse, making detection of a modulated target robust to the addition of interferers. In comodulation masking release (CMR) studies, adding off-target-frequency components that are modulated identically with the on-frequency masker improves the detectability of an unmodulated target (Hall *et al.*, 1984; van de Par and Kohlrausch, 1998; Winter *et al.*, 2004). However, we know of no studies reporting a corresponding benefit of increasing masker bandwidth when the target, rather than the masker, is modulated, so it is possible that there is a perceptual asymmetry between modulating the target versus modulating the masker in such situations, as well.

II. EXPERIMENTS AND HYPOTHESES

Two experiments were performed to study how modulation and spatial location of the target and masker affect target detection. Both target and masker were broadband noises that were either unmodulated or sinusoidally amplitude modulated (SAM). As a result, across-channel processing and across-frequency grouping were likely to contribute to performance. Moreover, for these broadband targets and maskers, listeners could not detect the target by using spectral sidebands (as might be the case when the target is a SAM tone; Dau and Ewert, 2004) and the opportunity to use profile analysis (Green, 1988) was minimized (because of the similarity of the target and masker spectral profiles).

A single modulation frequency (40 Hz) was used throughout the study, chosen both because humans are fairly sensitive to modulation at this frequency (Viemeister, 1979) and because responses of space-sensitive IC neurons are affected by modulation at this frequency (Lane and Delgutte, 2005).

Spatial separation of a broadband target from a broadband masker results in a frequency-dependent change in the target-to-masker energy ratio (TMR) at the ears. The resulting TMR profile as a function of frequency varies from one target/masker configuration to another, so that TMR should affect performance differently for different spatial configurations of the target and masker. The contribution of binaural processing to target detection should therefore depend on spatial configuration. In particular, if the TMR profile is such that the most favorable TMRs are at low frequencies, then interaural time difference (ITD) processing is likely to contribute to detection (Kopco and Shinn-Cunningham, 2003). On the other hand, if the most favorable TMRs are at high frequencies, then the contribution of ITD processing to performance is likely to be smaller. Finally, the contribution of across-frequency integration to detection, if any, is likely to be larger when the TMR is similar across frequency than when the TMR is very large in one band and small in others. As a result, the relative contribution of different detection cues (e.g., changes in overall energy and interaural decorrelation) also is likely to vary from one target/masker configuration to another.

Three different spatially separated configurations were included in this study to evaluate whether the interaction of modulation and spatial cues depends on the specific target/masker configuration. Specifically, in one of the chosen configurations the maximum in the TMR profile was in a low-frequency region, while in the remaining configurations it was at high frequencies.

As described above, the way in which modulation and spatial configuration interact is poorly understood. The current experiments were designed to explore how these cues jointly affect performance. If the processing of the two cues is strictly serial then the effects of the cues should be additive. This would occur if (1) spatial processing improves the effective TMR of the signal prior to any modulation processing, (2) modulation processing operates on the output of the spatial processing stage, and (3) detection is based on the output of the modulation processing. If the two cues both work to help listeners perceptually segregate the target from the masker, then the cues may be redundant. Specifically, if differences in modulation of the target and masker are sufficient to segregate the target and masker, then providing additional spatial cue differences in the target and masker might not improve performance. In this case, the benefits of modulation and spatial cue differences would be less than additive. Alternatively, if spatial cue differences are necessary for modulation differences to be useful (or vice versa), then the effects of differences in the two cues may be superadditive.

In addition to exploring whether the two cues are additive, subadditive, or superadditive, we tested two specific hypotheses about how source modulation structure and source location affect detection for broadband signals.

H1. The effect of modulation on SRM will depend on whether the target, the masker, or both target and masker are modulated (e.g., see the results of Asemi *et al.*, 2003).

H2. The effect of modulation on detection threshold will depend on spatial configuration because the relative importance of individual cues changes with spatial configuration. (1) When the best TMR occurs in low frequencies, ITD processing will be relatively influential on performance. (2) If perceived location rather than ITD processing is the critical factor in determining how spatial cues contribute to detection, performance will depend on whether or not the target and masker are spatially separated, but not on the exact spatial configuration. (3) When TMR is relatively constant with frequency, across-frequency integration is likely to contribute to detection.

Experiment 1 was performed with the masker noise presented at a fixed level. However, overall stimulus level may be the primary cue for detection when the target and masker are similar in their spectrotemporal structure and spatial cues, and therefore likely to be perceived as one unitary object from a particular location. To reduce the efficacy of overall level, Experiment 2 roved the masker level from interval to interval within each trial.

III. METHODS

A. Subjects

Seven subjects (four female and three male, including author N.K.) participated in Experiment 1. Seven subjects (three female and four male, two of whom participated in Experiment 1) participated in Experiment 2 (Experiment 2 was conducted almost a year after Experiment 1, so it is unlikely that learning from Experiment 1 transferred to Experiment 2 for the two subjects who performed both experiments). All subjects had normal hearing (confirmed by an audiometric screening), with ages ranging from 23–32 years.

B. Stimuli

The target and masker stimuli were both broadband noises with flat spectrum between either 0.3 and 8 kHz (target) or 0.2 and 12 kHz (masker), generated using a MATLAB implementation of the Butterworth bandpass filter (39th order for target and 33rd order for masker) with a stopband attenuation of 60 dB and stopband frequencies of 0.2–10.05 kHz (target) and 0.1–14 kHz (masker). The 200-ms-long target $s_T(t)$ was temporally centered on the masker $s_M(t)$, which had a duration of 300 ms. Both target and masker were ramped at onset and offset by 30 ms $\cos^2$ ramps. Modulation, if present, was sinusoidal with a frequency of 40 Hz and depth $m=0.5$ and had a random initial phase ϕ chosen from ten possible phases ($\phi=2\pi j/10$, $j=1,\ldots,10$). The stimuli were of the form

$$s_{i,k}(t) = A_i[1 + m_i \cos(2\pi 40t + \phi_{i,k})]n_{i,k}(t),$$

where $i=T$ for the target and $i=M$ for the masker, k is the trial number, $n_{i,k}(t)$ is a random bandpass-filtered noise token, and A_i is a scaling factor that determines the stimulus presentation level. The same five modulation conditions were explored in both experiments: no modulation ($m_T=m_M=0$), in-phase comodulation ($m_T=m_M=0.5$; $\phi_{M,k}=\phi_{T,k}$), target-only modulation ($m_T=0.5$; $m_M=0$), masker-only modulation ($m_T=0$; $m_M=0.5$), and pi-out-of-phase modulation ($m_T=m_M=0.5$; $\phi_{M,k}=\phi_{T,k}+\pi$).

Modulation increases the long-term rms energy of a signal by a factor of $(1+m^2)^{-0.5}$. For the modulation depth and form used here, modulation increases the rms energy of the modulated signal by approximately 0.5 dB. All results were corrected for this rms energy effect by scaling the measured thresholds and reporting thresholds in units of TMR.

Space was simulated using pseudoanechoic nonindividualized head-related impulse responses (HRIRs) recorded at four locations (−45°, 0°, 45°, and 90°, left to right) at a distance of 120 cm from the center of the head, using miniature microphones placed at the entrance of the ear canals of a female listener who did not participate as a subject in this study (see Shinn-Cunningham *et al.*, 2005, for a full description of these HRIRs). Five spatial configurations were explored in Experiment 1: two with the sources colocated at 0° or −45° and three with the sources spatially separated [(T at 90°, M at 0°), (T at 0°, M at 90°), and (T at 45°, M at −45°)]. An additional colocated condition (90°) was added in Experiment 2 to create three matching pairs of colocated and separated spatial configurations.

In both experiments, the average level of the masker was the same in all trials, prior to processing by the HRIRs (which altered the level of the signals reaching the ears). Therefore, because of HRIR processing, there were frequency-dependent variations in the signals reaching the ears across the different masker locations (graphs in Fig. 2 can be used to estimate how the received masker level changed at the two ears). For the masker at 0°, the maximum masker level received at the ears was 61 dB sound pressure level (SPL). In Experiment 1, the masker level was constant across the three intervals within a trial, while in Experiment 2 the masker level was roved independently in each interval by a value uniformly distributed between ±5 dB (the target, if present, was roved with the masker, which kept constant the TMR measured prior to HRIR processing).

Stimulus files, generated off-line at a sampling rate of 50 kHz, were stored on the hard disk of a control computer (IBM PC compatible). Ten random noise tokens were pregenerated to be used as targets and another ten tokens were produced to be used as maskers in this study (i.e., target and masker were always independent samples of noise). These 20 tokens were bandpass filtered (10 by the target filter and 10 by the masker filter, which had a slightly wider passband), modulated (by 1 of 10 modulation envelopes, differing in initial phase), and HRIR filtered (by an HRIR corresponding to locations of −45°, 0°, 45°, or 90°) to produce 440 target stimuli [10 tokens × (10 modulation envelopes + no modulation) × 4 locations] and 440 similar masker stimuli. On each trial, three different masker tokens and one target token were randomly selected, scaled, and concatenated into a stimulus file that contained three masker intervals with the target randomly added to the second or the third interval.

TDT System 3 hardware was used for D/A conversion. The result was amplified through a TDT headphone buffer and presented via Etymotic Research ER-1 insert earphones (with approximately flat frequency response in the range 100 Hz–15 kHz). No filtering was done to compensate for the transfer characteristics of the playback system. A simple alphanumeric interface in MATLAB was used to give instructions to subjects, gather responses, and provide feedback. The subject indicated the perceived target interval by hitting the appropriate numeric key ("2" or "3") on the computer keyboard. Experiments were performed in a single-walled sound-treated booth.

C. Experimental procedure

Each trial consisted of three intervals, each of which contained a masker. Either the second or the third interval

(randomly chosen with equal probability on each trial) also contained the target. The intervals were separated by 50-ms-long silent gaps. Subjects performed a two-alternative, forced-choice task in which they were asked to identify which interval, the second or the third, contained the target. Correct-answer feedback was provided at the end of each trial.

A three-down-one-up adaptive procedure was used to estimate detection thresholds (Levitt, 1971), defined as the 79.4% correct point on the psychometric function. Each run started with a description of the measurement condition of the run (e.g., written instructions might read "In this run, the target is modulated and the distractor is not modulated, the target comes from an azimuth of 0° and the distractor from 90°. Next, you will hear a sample of the noise distractor that you should ignore, followed by the target that you should identify. Hit RETURN to hear the sample."). The subject could listen to the sample repeatedly until he/she was confident that he/she understood the task.

The staircase measurement procedure started with the target presented at a clearly detectable level and continued until 11 "reversals" occurred. The target level was changed by 4 dB on the first reversal, 2 dB on the second reversal, and 1 dB on all subsequent reversals. For each adaptive run, detection threshold was estimated by taking the average target presentation level over the last six reversals.

Each of the two experiments consisted of six 1 h sessions performed on different days (the first session of each experiment was a practice session, serving to familiarize the subjects with the experimental procedure). In each session, the thresholds were measured for all combinations of spatial and modulation conditions (25 thresholds in Experiment 1 and 30 in Experiment 2), with the order of conditions randomized between sessions and between subjects. One adaptive run took approximately 2–3 min to complete.

Informal interviews of the listeners confirmed that at moderate to high TMRs, listeners found it very easy to interpret the two simulated stimuli as a target noise and a distractor noise coming from the indicated locations with the described modulation characteristics (as opposed to hearing them as one combined noise). This was likely the case because of the following: (1) at the beginning of the experiment, the subjects were given a detailed description of the stimulus combinations they should expect; (2) prior to each adaptive run, listeners had the opportunity to familiarize themselves with the target and masker stimuli presented separately before they heard them combined; and (3) the procedure started with both the target and the masker clearly audible. It is difficult to know whether or not the listeners perceived the two stimuli as separate objects when the target level was near the threshold. However, none of the subjects reported any difficulty performing the task (for example, none of them reported being confused about what to listen for in order to detect the target).

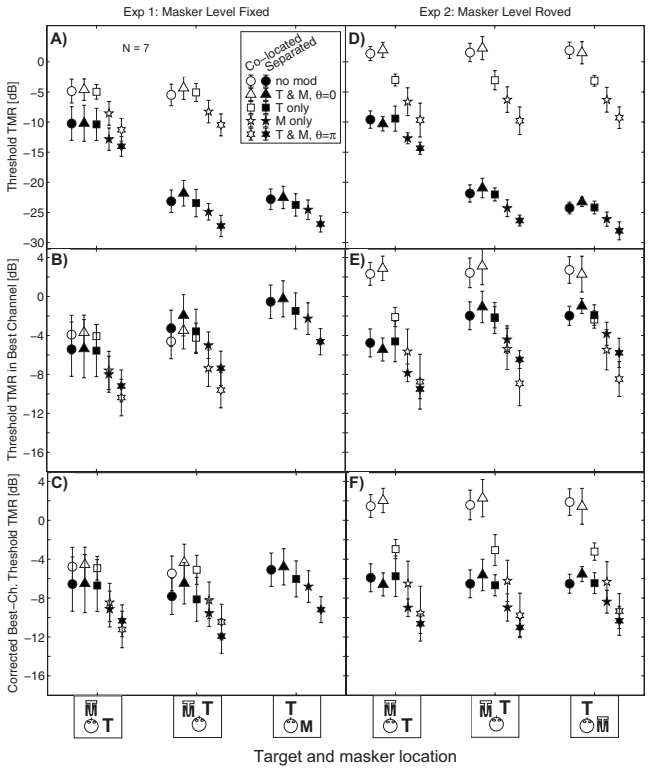


FIG. 1. Raw data plotted as a function of the masker location measured with the masker level fixed (Experiment 1; panels A, B, and C) and roved (Experiment 2; panels D, E, and F). All graphs show the across-subject mean and standard deviations in measured threshold TMRs: panels A and D show the raw threshold TMR energy ratios, panels B and E show the threshold TMRs in the best channel, and panels C and F show the threshold TMRs in the best channel after correcting for the frequency-dependence of the threshold TMR sensitivity.

IV. RESULTS AND ANALYSIS

A. Experiment 1: Fixed masker level

1. Overall results

Panels A, B, and C in Fig. 1 present the data collected in Experiment 1, with the masker level fixed (Panels D, E, and F show the data from Experiment 2, discussed in Sec. IV B). The data are plotted as a function of the masker location (indicated by the position of the letter "M" in the icons along the abscissa). Two spatial configurations are plotted for each masker location, one with the target and masker colocated (open symbols) and one with the target displaced from the masker (filled symbols).[1] The spatially separated target was at the location indicated by the filled letter "T" in the icons along the abscissa. The thresholds for different modulation conditions are represented by different symbols.

Figure 1(a) shows the across-subject mean and standard deviation of the TMR at detection threshold (lower values correspond to better performance). Thresholds varied by more than 20 dB, depending on the spatial configuration and modulation condition. For a given modulation condition and masker location, performance when the target and masker were spatially separated (filled symbols) was always better than when they were colocated (open symbols), revealing robust SRM. The colocated thresholds for target and masker at 0° and −45° were nearly identical, suggesting that the exact spatial configuration of the target and masker was not

important as long as the sources were colocated (this observation, based on the two configurations in Experiment 1, is further supported by the results of Experiment 2 in which three colocated thresholds were measured). In contrast, the spatially separated thresholds were strongly influenced by the specific target and masker locations: performance was worse with the masker at 0° than with the masker at −45° or 90° [compare the leftmost group of filled symbols in Fig. 1(a) to the center or the rightmost groups].

Within each spatial configuration, the no-modulation, in-phase comodulation, and target-only modulation (circles, triangles, and squares, respectively) thresholds were generally comparable, and these thresholds were higher (performance was worse) than the remaining thresholds. Masker-only modulation yielded improvements in performance (pentagrams fall below circles), while out-of-phase modulation of the target gave the lowest thresholds (hexagrams tend to fall below pentagrams).

A three-way repeated-measures analysis of variance (ANOVA) was performed with factors of modulation, spatial separation (colocated versus separated), and masker location (0, −45°), paralleling the layout of Fig. 1(a). The (*M* 90°, *T* 0°) configuration was omitted because it had no corresponding colocated measurement. This statistical analysis found a significant modulation × separation interaction ($F_{4,24}=7.63$, $p=0.0004$), a significant separation × masker location interaction ($F_{1,6}=950$, $p<0.0001$), and significant effects of all three main factors ($p<0.0001$). Notably though, neither the interaction between modulation and masker location nor the three-way interaction was significant ($p>0.1$). These results suggest that, although overall performance and the effect of separation depend on spatial configuration, at least for the spatial configurations explored in this study, the effect of modulation on the thresholds is similar within each spatial configuration rather than varying with target and masker locations.

2. Energy effects in 1/3-octave bands

One factor contributing to the large spatial benefits and to the dependence of these improvements on spatial configuration is the better-ear advantage, arising from the changes in the level at which the stimuli are received at the left and right ears when target and masker are spatially separated. In general, spatial separation of the target and masker sources produces a larger TMR at one of the ears (the "better ear"), and a smaller TMR at the other ear, compared to when the sources are colocated (where the TMR is equal at the two ears). To explore the extent to which changes in TMR at the acoustically better ear could account for the observed spatial unmasking, we calculated the TMR in each of the signals reaching the listeners' two ears as a function of frequency.

For each spatial configuration, we selected a target and a masker processed by the appropriate HRIRs and filtered both target and masker into 22 log-spaced 1/3-octave signals per ear (ANSI, 1986). In this analysis, the target and masker were set to have the same level prior to spatial processing. (Note that the effects of spatial processing on the TMR at the ears are identical for all modulation conditions.) The resulting frequency-dependent TMRs show the proper correction

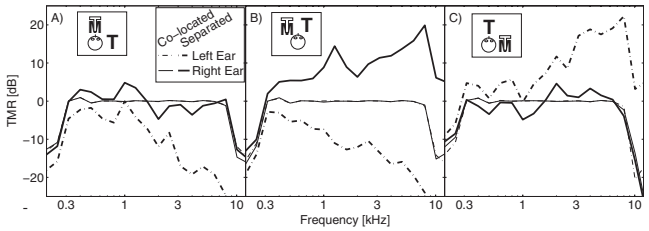


FIG. 2. TMR in 1/3-octave frequency bands in the six target/masker spatial configurations used in this study. Each panel shows the left- and right-ear TMRs in the colocated and separated spatial configurations for one masker location (indicated by the inset) as a function of the center frequency of the third-octave filters.

needed, at each frequency, to calculate the TMR at detection threshold in each of the 22 frequency bands.[2] The results of this analysis are plotted in Fig. 2.

Each panel in Fig. 2 shows the TMRs for one fixed masker location (indicated by the inset icon), with each combination of the ear (solid versus dashed lines for right versus left ear, respectively) and the spatial configuration (thin versus thick lines for colocated versus spatially separated) plotted separately. (Note that the dashed and solid thin lines lie nearly on top of each other, so only the solid thin lines are easily visible.)

TMRs for the colocated configurations (thin lines) were approximately zero (or less than zero at the edges where no target energy was present), independent of the masker location (across panels) or the ear (solid versus dashed thin lines). The spatially separated TMRs were frequency dependent and varied both with the ear (solid versus dashed thick lines within each panel) and with the masker location (panel A versus panel B versus panel C). The largest improvement in TMR with spatial separation was approximately 5 dB in panel A (right-ear channel centered at 1 kHz), approximately 20 dB in panel B (right-ear channel centered at 8 kHz), and approximately 22 dB in panel C (left-ear channel centered at 8 kHz). Assuming that the listeners detect the target by detecting its presence due to the energy effects in the frequency channel with the most favorable TMR, detection performance with spatial separation is expected to improve due to the spatial configuration by an amount equal to the maximum TMR shown in each panel of Fig. 2. Note that this analysis assumes that, in each condition, performance is determined solely by the single frequency channel with the most favorable TMR and that the threshold TMR calculated in 1/3-octave band is the same for all frequency channels. Therefore, this analysis ignores possible contributions of across-frequency integration and binaural processing. Moreover, the exact TMRs computed in this way will depend on the detailed shapes of the peripheral auditory filters used, as well as how they change with center frequency, so that slightly different corrections would be found with different filter assumptions. However, this analysis provides a first-order correction for the wide variation in TMR with frequency caused by HRIR processing.

Figure 1(b) shows the threshold TMRs in the best frequency channel, determined by adding the best-channel correction (i.e., the peak values from Fig. 2) to the respective thresholds in Fig. 1(a). Colocated thresholds [open symbols

in Fig. 1(b)] were essentially unchanged, as the TMR correction was near zero at all frequencies. However, correction of the spatially separated configurations reduced the effect of spatial separation to the point that many spatially separated thresholds (e.g., all thresholds with masker at −45°) were actually higher (performance was worse) than the corresponding colocated thresholds. Although this correction removed a good portion of the spatial effects on performance, ANOVA performed on the better-ear, best-frequency corrected thresholds found the same significant main factors and interactions as did the uncorrected thresholds [Fig. 1(a)] suggesting that the correction, while reducing the dependence of thresholds on the masker location, did not account for all of the variation in performance with spatial configurations.

3. Additional correction for frequency dependence of threshold TMR

The better-ear best-frequency correction yielded threshold TMRs that were much more similar than the uncorrected TMRs. To the extent that this correction was sufficient to account for the behavioral results, it suggests that (a) the threshold TMR is the same in all channels independent of frequency, (b) a simple 1/3-octave filter is an adequate representation of auditory filtering for the current analysis, and (c) there is no contribution of across-frequency integration or binaural processing to performance. The effect of any deviation from these assumptions is likely to depend on the spectral profiles of the target and masker signals, which differ with spatial configuration (see Fig. 2).

We now examine the assumption that threshold TMR in 1/3-octave band is constant as a function of frequency. In a previous study that measured SRM for broadband chirp-train signals masked by noise, threshold TMRs for narrowband targets were not constant as a function of frequency; instead, threshold TMRs were lower for higher-frequency targets (Kopco, 2005). When listening in a 9 kHz channel, best-channel analysis based on 1/3-octave filtering yielded thresholds that were nearly 4 dB lower than threshold TMRs using a 1 kHz channel. A simple frequency-dependent linear correction fit these earlier results relatively well (Kopco, 2005). The same correction, derived from the empirical fit to the data in this previous study, was applied to the current results:[3]

$$\mathrm{TMR}_{\mathrm{corrected}} = \mathrm{TMR}_{\mathrm{uncorrected}} + k_1 \mathrm{CF} + k_2. \quad (1)$$

Here, $\mathrm{TMR}_{\mathrm{uncorrected}}$ are the data from Fig. 1(b), CF is the center frequency of the best-TMR filter in Hz, the constant k_1 was fitted to Kopco's (2005) data (k_1 was estimated to be -4.9×10^{-4} dB/Hz), and the constant k_2 was arbitrarily set to 1.34 dB to minimize the offset of the corrected data from the raw colocated data. (Note that the constant k_2 does not influence relative comparisons, as it shifts all data points by the same amount, but simply accounts for the absolute value of the TMR threshold). The frequency-corrected best-TMR model uses the same assumptions as the best-channel TMR correction shown in Fig. 1(b), except that it relaxes the assumption of a constant frequency-independent threshold TMR sensitivity. Instead, threshold TMR is assumed to decrease linearly with increasing center frequency.

Figure 1(c) shows the thresholds corrected by Eq. (1). Compared to the graphs in Fig. 1(b), the corrected spatially separated thresholds [filled symbols in Fig. 1(c)] were always better than or equal to the corresponding colocated thresholds (open symbols). Thresholds were roughly equal across all masker locations [in Fig. 1(c), the *M* 0°, *T* 90° thresholds were approximately equal to the corresponding *M* −45°, *T* 45° thresholds, as well as to the *M* 90°, *T* 0° thresholds; the trend was confirmed by data shown in Fig. 1(f) from Experiment 2]. Because the same correction was applied to all thresholds for a given spatial configuration, independent of the modulation condition, colocated thresholds still changed more as a function of the modulation condition than did the spatially separated thresholds. (Supporting these observations, ANOVA performed on the corrected data only found one significant interaction, modulation × separation, $F_{4,24}=7.65$, $p<0.0005$; all three main effects were significant, with $p<0.05$.) With these corrections, the spatially separated thresholds were only consistently lower than colocated thresholds in the no-modulation, in-phase modulation, and target-only modulation conditions [filled versus open circles, triangles, and squares in Fig. 1(c)]. Colocated and spatially separated thresholds were statistically indistinguishable in the masker-only modulation and out-of-phase modulation conditions for all spatial configurations.

Given the similarity of the corrected best-channel threshold TMRs at different masker locations [Fig. 1(c)], there only appears to be a modest effect of across-frequency integration in this study (i.e., there are no large differences across different spatial configurations, even though the best frequency and the overall shape of the better-ear TMR as a function of frequency vary dramatically with spatial configuration). Similarly, spatial processing only appears to contribute when the masker is modulated in a way that does not provide glimpses of the target (in the no modulation, in-phase modulation, and target-only modulation conditions).

In all of the following sections, the frequency-corrected best-channel TMR thresholds [from Figs. 1(c) and 1(f)] are used because (1) this correction accounts for the dependence of the thresholds on the masker location; (2) even though consideration of binaural processing and across-frequency integration could also produce corrections that explain some of the variability as a function of the masker location,[4] parsimony argues that these factors played only minor roles in this experiment; and (3) the fact that spatially separated configurations produce thresholds that depend less on the modulation condition than do colocated configurations is independent of the method used to account for energy effects or of the masker location. (However, note that it is currently not clear what causes the frequency dependence of the 1/3-octave filtered threshold TMRs.)

4. Results collapsed across the masker location

To better assess the interaction between modulation and separation, Fig. 3 shows the data collapsed across masker location. Figure 3(a) plots the across-subject mean threshold TMRs in the best 1/3-octave channel (and within-subject standard deviation, chosen here because it removes the between-subject differences from the computation of stan-

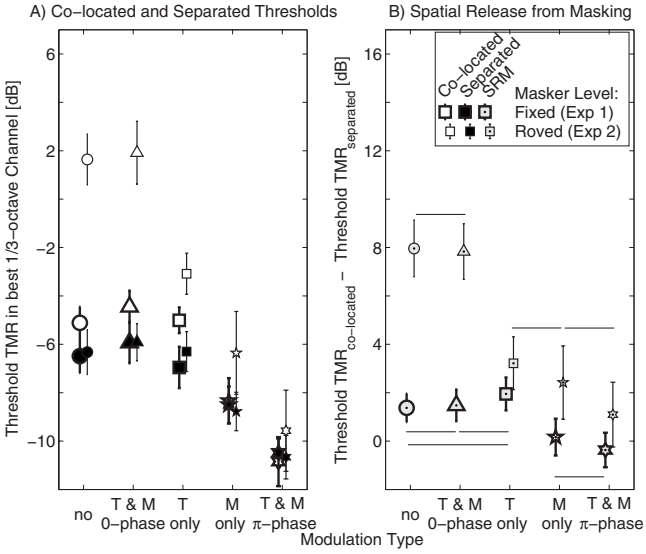


FIG. 3. Threshold TMRs in the frequency-corrected best 1/3-octave channel (panel A) and SRM (panel B) as a function of the modulation type, averaged across the masker locations (error bars give the within-subject standard deviation). The horizontal lines in panel B indicate SRMs that were not significantly different at the 0.01 level in a t-test after correcting for multiple comparisons (lines below large symbols for Experiment 1; lines above small symbols for Experiment 2). Different symbols are used to identify the modulation type, as in Fig. 1. The legend in panel B applies to both panels and all modulation conditions.

dard deviation[5]) as a function of the modulation type. The large filled and open symbols represent the spatially separated and colocated thresholds, respectively (the small symbols represent the results of Experiment 2, discussed in Sec. IV B).

The effect of modulation on performance was similar for colocated and separated spatial configurations. Thresholds were essentially the same for the no-modulation, in-phase comodulation, and target-only modulation conditions [compare large open and filled circles, triangles, and squares in Fig. 3(a)]. Performance with masker-only modulation (pentagrams) and out-of-phase modulation (hexagrams) was better, with lower thresholds.

Although the rank ordering of thresholds was the same for colocated and spatially separated conditions, the dependence of the thresholds on modulation was slightly stronger when the sources were colocated than when they were spatially separated (large open symbols span a range of nearly 7 dB, while the large filled symbols span a range of about 4 dB), suggesting that spatial separation affects performance differently for different modulation conditions. This SRM [the difference between the open and filled symbols in Fig. 3(a)] is plotted as a function of the modulation condition in Fig. 3(b). This panel shows the across-subject mean (and the within-subject standard deviation[5]) of the difference between the spatially separated and corresponding colocated thresholds from panel A.

One-way repeated-measures ANOVA found a significant effect of modulation on SRM ($F_{4,24}=11.44$, $p<0.0001$). The results of Bonferroni-corrected *post hoc* pairwise t-tests (which account for heterogeneity of variances; e.g., Ury and Wiggins, 1971) as implemented in the CLEAVE package (Herron, 2005) are also shown in Fig. 3(b). The horizontal lines under the large symbols in Fig. 3(b) indicate those pairs of conditions in Experiment 1 that did not differ at the $p<0.01$ significance level (all other pairs were significantly different from one another). The no-modulation, in-phase modulation, and target-only modulation SRMs were not significantly different from one another. Similarly, the masker-only modulation versus out-of-phase modulation SRMs were not significantly different from one another. However, the modulation type had a small but significant effect on the SRM: compared to no-modulation, in-phase modulation, or target-only modulation [circle, triangle, and square in Fig. 3(b)], modulating only the masker (pentagram) or modulating the target and masker stimuli with opposite phases (hexagram) decreased the SRM by roughly 1.5–2 dB ($p<0.01$), resulting in no benefit of spatial separation in the latter modulation conditions.

Finally, as discussed in the Appendix, learning affected SRM: in the first of the five repeats of this experiment, the SRM was essentially the same for all types of modulation (the largest difference was less than 1 dB). However, by the fifth repeat, the difference between the target-only modulation and the out-of-phase modulation grew to more than 4 dB. Thus, the average effect plotted in the data collapsed across the repeats is smaller than might be seen after extensive training.

B. Experiment 2: Masker level roved

To isolate the contribution of the overall level cue to performance, Experiment 2 was performed with the masker level roved between the intervals within a trial, a strategy used extensively in the profile analysis literature (Mason *et al.*, 1984; Kidd *et al.*, 1989). The (T 90°, M 90°) colocated condition was added to balance the number of colocated and spatially separated conditions; otherwise, Experiment 2 was identical to Experiment 1, except with a random ±5 dB intensity rove added from interval to interval.

1. Overall results

Panels D, E, and F in Fig. 1 present the results of Experiment 2 in a format identical to Experiment 1 (see Sec. IV A). The raw data in Fig. 1(d) followed the same trends as in Experiment 1. The spatially separated thresholds (filled symbols) were almost identical to those found in Experiment 1. The colocated thresholds for the no-modulation (circles) and in-phase modulation (triangles) conditions tended to be worse than in Experiment 1. However, the level rove had little effect on the remaining colocated configurations (a direct comparison is presented below). This result suggests that overall level was the main cue used for detection only in the colocated configurations in which the target and masker had identical temporal envelopes, a conclusion that was confirmed by a comparison of the data in panels E and F to respective panels B and C. (ANOVAs performed on the raw and corrected Experiment 2 data from panels D, E, and F found the same significant main effects and interactions as the respective ANOVAs performed on the Experiment 1 data.)

2. Results collapsed across the masker location

In order to analyze the interaction between modulation and spatial separation, the data were collapsed across the masker locations. To allow a direct comparison of the effect of masker level uncertainty, Fig. 3 shows the results for Experiment 2 (the small symbols slightly offset to the right) plotted alongside the data from Experiment 1 (larger symbols).

The filled symbols in Fig. 3(a) show the spatially separated thresholds. Roving the masker level had essentially no effect on any of the spatially separated thresholds (compare the small and large filled symbols from Experiments 2 and 1, respectively). In contrast, all colocated thresholds were larger in Experiment 2 than in Experiment 1 (the small open symbols fell above the corresponding large open symbols). The largest increase (around 7 dB) was observed when the target and masker had identical temporal envelopes (i.e., in the no-modulation and in-phase comodulation conditions; circles and triangles). In the three remaining modulation conditions, the masker-level rove increased thresholds by approximately 2 dB.

Figure 3(b) shows that, as a consequence of the effects of the level rove on the colocated configurations, the SRM was much larger in Experiment 2 than in Experiment 1 in the conditions in which the target and masker had the same temporal envelope. A one-way repeated-measure ANOVA revealed a significant effect of modulation on the SRM ($F_{4,24}=35.55$, $p<0.0001$). Bonferroni-corrected *post hoc* pairwise t-tests found no significant differences between unmodulated and comodulated SRMs, target-only and masker-only modulated SRMs, or the masker-only and out-of-phase modulated SRMs (see the horizontal bars above the pairs of small symbols that were not significantly different; $p>0.01$). All other pairs of modulation conditions showed statistically significant differences.

The results in Fig. 3 suggest that overall level was used to detect the target when the colocated target and masker had the same envelope. For wideband noise, the smallest detectable intensity change ΔI is proportional to the base line intensity, I, so that $\Delta I/I$ is approximately constant with values between −9 and −6 dB over a large range of I (20–100 dB above the absolute thresholds; Moore, 2003). The results for colocated identically modulated stimuli in Experiment 1 match these data well, with TMR thresholds of approximately −5 dB [large open circles and triangles in Fig. 3(a)]. If overall level was the only available cue in this two-alternative forced-choice task and the external noise of the 10 dB rove dominated performance, then the TMR at detection threshold would be 1.07 dB for an ideal observer (Durlach *et al.*, 1986; Green, 1988), which is remarkably close to the actual thresholds observed for the identically modulated and in-phase modulated conditions, where threshold TMRs were around 2 dB. In most previous studies of the effect of rove on profile analysis, the rove yielded performance that was worse than was predicted for an optimal observer (Spiegel *et al.*, 1981; Mason *et al.*, 1984). Thus, even the fact that thresholds are slightly higher than the ideal-observer prediction is consistent with past work. Moreover, the no-modulation and in-phase comodulation thresholds were very similar to each other, suggesting that the fluctuating envelope in the latter condition did not make it harder to judge the levels in the different intervals.

In conditions for which target and masker were colocated but had different temporal envelopes, performance was much better than would be predicted if the main cue used for target detection was overall intensity, showing that some other nonlevel cue was the main feature used to detect the target. Nevertheless, in such conditions, the rove interfered slightly with performance, a result that suggests that the intensity rove made it more difficult for listeners to extract whatever feature was the main detection cue when target and masker were colocated.

C. Modulation detection

To understand the effects of modulation on performance, two analyses were performed. First, the instantaneous TMR was analyzed. In this analysis, predictions were based on detecting the target by hearing its effect at the best instant in time. A second analysis assumed that the listeners detected the target+masker interval by detecting a modulation depth that was different from the masker-only modulation (in the nontarget intervals).

1. Listening at peaks and dips: Instantaneous TMR analysis

The presence of modulation in the stimuli caused the instantaneous TMR to change over time. Humans appear to utilize these changes and detect the target in moments when the TMR is most favorable, both in monaural (Buus *et al.*, 1996) and binaural (Buss *et al.*, 2003) listening tasks, even though this ability can differ across subjects (e.g., see Buss *et al.*, 2007). Of course, given that the ability to utilize these cues is limited by the temporal resolution of the auditory system, factors like forward masking are likely to influence the ability to listen in dips (Widin *et al.*, 1986; Wojtczak and Viemeister, 2005). While the present analysis does not consider these limitations, it does provide an upper limit on how much the listeners could have benefited from changes in the instantaneous TMR. Specifically, if one assumes that the peak TMR produced after temporal integration over some fixed time window predicts performance, the current analysis gives the limit of performance if temporal resolution is infinitely precise, leading to an effective time window that is infinitely narrow. Conversely, the overall-TMR analysis shown in Fig. 3(a) shows predictions for an infinitely long time window. Any finite-length time window must produce results intermediate between these two extremes.

In the colocated conditions with identical modulation (no modulation and in-phase comodulation; circles and triangles), the TMR was constant over the duration of the stimulus. In the conditions with different target and masker modulations, the difference between the long-term TMR and the peak instantaneous TMR depended on which stimulus was modulated. Because the modulation envelope was sinusoidal in pressure units, the effect of modulation on the instantaneous sound pressure level was not symmetrical in decibel units. For sinusoidal modulation with a modulation

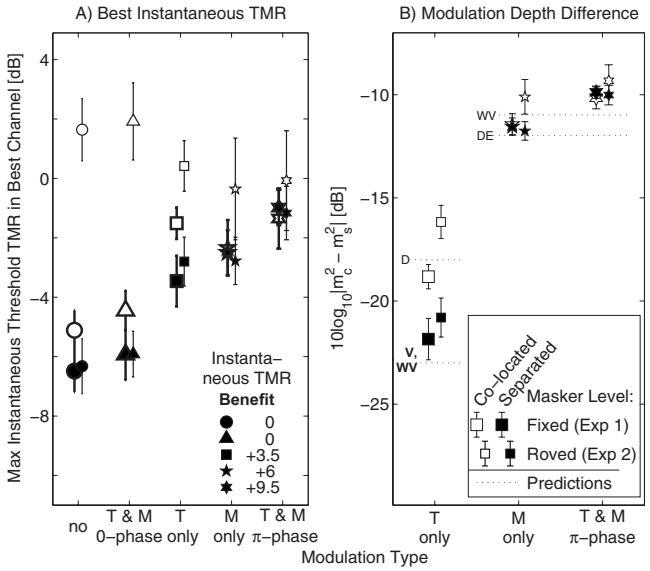


FIG. 4. (A) Peak instantaneous TMRs at threshold in the best 1/3-octave channel [derived from Fig. 3(a) by applying the instantaneous TMR benefit corrections, listed in the inset, to both colocated and separated thresholds of both Experiments 1 and 2] (B) Modulation depth (across-subject means and within-subject 95% confidence interval) at the threshold TMR in the three modulation conditions in which modulation of the target and masker differed. Data are compared to predictions based on the data of Wakefield and Viemeister (1990)—WV, Dau and Ewert (2004)—DE, Viemeister (1979), and Dau (1996)—D. The legend of panel B applies to data in both panels and to all modulation conditions.

depth of 0.5 (used in this study), the instantaneous signal level at the minima of the modulation envelope was 6 dB lower than the level with no modulation, while the level at the peaks of the modulation envelope was 3.5 dB higher than the unmodulated level.

Figure 4(a) plots the best instantaneous TMR in the frequency-corrected best 1/3-octave channel at threshold, determined by adding the instantaneous-TMR-benefit corrections (described above and listed in the inset) to the long-term frequency-corrected TMR thresholds in the best frequency channel [from Fig. 3(a)]. [Note that for each modulation condition, colocated and spatially separated thresholds have the same instantaneous-TMR-benefit correction, so that this correction does not influence SRM, shown in Fig. 3(b).]

As seen in Fig. 4(a), the peak instantaneous TMR at detection threshold falls between −4 and 0 dB for the conditions in which the target and masker envelopes differ [target modulation, masker modulation, and out-of-phase modulation conditions; large open squares, pentagrams, and hexagrams in Fig. 4(a)]. These values are higher than the intensity just noticeable difference (JND) (−9 to −6 dB, as discussed above), suggesting that listeners were unable to make use of the peak instantaneous TMR to detect the target based on changes in overall intensity. Given that the long-term average TMR does not capture the differences in thresholds as a function of modulation type [if it did then the thresholds represented by the large open squares, pentagrams, and hexagrams would be constant in Fig. 3(a)], while the instantaneous TMR predicts performance that is too poor (even though it is approximately constant), it is possible that predictions based on the TMR averaged over an appropriate finite-length time window could account for detection based on changes in intensity. However, if performance were based on the same intensity cue for cases when target and masker had the same envelope and cases when the target and masker envelopes differed, the effect of intensity rove should be similar in all conditions. Instead, intensity rove affected performance in the different conditions very differently, suggesting that some cue other than overall intensity integrated over some finite-duration time window enabled target detection when target and masker envelopes differed.

2. Effect of the target on the masker envelope modulation

One attribute that is affected by the addition of the target to the masker is the shape of the total stimulus envelope (Dau *et al.*, 1997). The salience of any change in the envelope due to the presence of the target depends on the relative levels of the target and masker as well as on the modulation condition. In the target-only-modulated condition, modulation is only present in the target interval and listeners may detect the target by detecting the presence of modulation. In the masker-only-modulated and the target-and-masker-modulated-out-of-phase conditions, the addition of the target decreases modulation depth from the 0.5 depth in the nontarget intervals and listeners may discriminate changes in the modulation depth to detect the target.

Detection and discrimination thresholds for modulation can be expressed as the modulation index $10\log_{10}|m_c^2-m_s^2|$, where m_s represents the modulation depth of the standard (i.e., in the nontarget interval) and m_c is the modulation depth of the stimulus at discrimination threshold (i.e., the modulation depth of the combined target+masker signal in the target interval). The current target-modulated thresholds can be estimated either from previous modulation detection data (Viemeister, 1979; Dau, 1996) or from discrimination data using a standard with a very low modulation depth (Wakefield and Viemeister, 1990, and Dau and Ewert, 2004; summarized in Fig. 2 of Dau and Ewert, 2004).[6] For modulation detection, the modulation index at threshold is in the range from −23 dB (Viemeister, 1979) to −18 dB at threshold (Dau, 1996). The results from modulation discrimination experiments (Wakefield and Viemeister, 1990) suggest that modulation index thresholds are near −23 dB for standard depths less than −30 dB.

Thresholds from a previous modulation discrimination study (e.g., Dau and Ewert, 2004) can be linearly approximated as $10\log_{10}(m_c^2-m_s^2)=10\log_{10}m_s^2-4$, from which the predicted threshold for a decrease in modulation from the standard of $m_c=-6$ dB can be estimated as $10\log_{10}(m_c^2-m_s^2)=-11$ dB [thresholds from Wakefield and Viemeister (1990) are approximately 1 dB larger than the Dau and Ewert (2004) thresholds when analyzed in this way].

In order to compare the current data to these predictions, the relationship between the threshold TMRs and the modulation depth of the combined stimulus was examined for our stimuli. However, combining a SAM noise and an unmodulated noise does not produce a stimulus with sinusoidal amplitude modulation. The relation between the threshold

modulation and threshold TMR was estimated by determining the maximum and minimum amplitudes, of the combined stimulus envelope and then finding the modulation depth of a SAM stimulus that would give the same maximum and minimum (although the exact shape of the modulation envelope differs, the difference is relatively small, especially near threshold). The resulting relationships for the three differential modulation conditions in this study (and for the target and/or masker modulation of 0.5) are as follows.

In target-only modulated,

$$m = \frac{\sqrt{1 + 1.5^2\mathrm{TMR}^2} - \sqrt{1 + 0.5^2\mathrm{TMR}^2}}{\sqrt{1 + 1.5^2\mathrm{TMR}^2} + \sqrt{1 + 0.5^2\mathrm{TMR}^2}}.$$

In masker-only modulated,

$$m = \frac{\sqrt{1.5^2 + \mathrm{TMR}^2} - \sqrt{0.5^2 + \mathrm{TMR}^2}}{\sqrt{1.5^2 + \mathrm{TMR}^2} + \sqrt{0.5^2 + \mathrm{TMR}^2}}.$$

In stimuli modulated out of phase,

$$m = \frac{\sqrt{1.5^2 + 0.5^2\mathrm{TMR}^2} - \sqrt{0.5^2 + 1.5^2\mathrm{TMR}^2}}{\sqrt{1.5^2 + 0.5^2\mathrm{TMR}^2} + \sqrt{0.5^2 + 1.5^2\mathrm{TMR}^2}},$$

where TMR is the threshold TMR in the best channel (from Fig. 3) in pressure units and m is the threshold modulation depth of an equivalent SAM noise. These equations can be inverted to estimate the target+masker modulation depth at target detection threshold for the measured results.

Figure 4(b) shows data for the three modulation conditions in which target modulation is different from the masker modulation, expressed as the difference in modulation depth between the target+masker interval and the reference masker-alone interval (the modulation conditions for which the target and masker have the same envelope were not included in this analysis because there is no change in modulation with addition of the target). Also shown are the predictions estimated from results of Viemeister (1979), Wakefield and Viemeister (1990), Dau and Ewert (2004), and Dau (1996; see dashed lines).

The thresholds for the colocated stimuli with fixed masker levels (open large symbols) generally match the previous detection and discrimination data fairly well for all three types of modulation, suggesting that the listeners detected changes in modulation depth in these conditions. The spatially separated thresholds are only lower (detection is easier) than the colocated thresholds in the target-modulation condition, when the listeners do not ever get a good "glimpse" of the target (large filled versus open squares). At first glance, the fact that the spatially separated thresholds fall within the range of the previous modulation detection data (i.e., between the dotted lines marked by D and V, VW) seems to suggest that the listeners did not benefit from spatial cues in this condition. However, given the large difference between the D and the V, VW thresholds, and given that there is a consistent difference between the colocated and spatially separated thresholds in the current study, it is clear that the listeners did use the spatial separation cue, in addition to modulation, here.

Finally, although the effect is small, colocated roved thresholds (open small symbols) consistently fall above the range of thresholds observed in previous studies which did not rove the stimulus presentation level. This shows that overall level rove impaired the listeners' ability to detect or discriminate modulation in the current study.

V. DISCUSSION

Noise-on-noise threshold TMRs changed over a range of 30 dB [Figs. 1(a) and 1(d)], and were influenced by the spatial configuration of the target and masker, the type of modulation present in the stimuli, and a rove of the masker level. Moreover, as discussed in the Appendix, these differences appear to increase with experience. A large part of the variability in performance across the tested conditions (as much as 20 dB) came from the changes in the target and masker energy levels received at the ears when the target and masker locations changed. Specifically, if one considers the TMR within the best 1/3-octave frequency channel in the acoustically better ear, threshold TMRs ranged only over 5 dB across different spatial configurations. If one then corrects these detection thresholds based on the detection threshold differences across frequency,[3] threshold TMRs were even closer, spanning a range of only about 1 dB across the different spatial configurations for a given modulation condition.

As shown in Fig. 2, the way in which TMR varies with center frequency differs dramatically across the spatial configurations used in this study. Therefore, any contributions of ITD and across-frequency processing to performance are likely to depend on masker location. However, no large differences were observed after applying frequency-dependent corrections to the TMR in the best frequency channel. Thus, for the broadband stimuli used here, both binaural and across-frequency contributions to performance appear to be modest. Frequency-dependent TMR thresholds could also explain the results of a previous related experiment without considering any across-frequency integration or binaural processing (Lane *et al.*, 2004). Together, these results suggest that low-level binaural processing does not contribute very much to spatial unmasking when detecting a broadband target in a broadband masker (although it can contribute significantly when the target is narrowband; e.g., see Kopco and Shinn-Cunningham, 2003).

The benefit of spatial separation found in the current results is similar for all spatial configurations, even though the best frequency channel is sometimes in a low-frequency region where binaural processing is expected to provide a large benefit and sometimes in a high-frequency region where binaural processing typically provides much more modest benefits (Zurek, 1993; Kopco and Shinn-Cunningham, 2003). This suggests that differences in the perceived *spatial attributes* of the stimuli (which depend both on low-frequency ITDs as well as high-frequency interaural level differences and spectral cues) are responsible for the spatial unmasking not explained by changes in the TMR at the better ear, rather than *binaural processing* that operates primarily at low-frequencies (unmasking caused by interau-

ral decorrelation; Colburn, 1977); (see Freyman *et al.*, 1999, for another study contrasting how spatial perception and binaural processing contribute to spatial unmasking).

Both modulation and intensity rove influenced the SRM, defined as the difference between the best-channel threshold TMRs with colocated and spatially separated stimuli. With the masker level fixed, SRM was comparable for no-modulation, target and masker in-phase modulation, and target-only modulation configurations, but SRM was statistically insignificant when only the masker was modulated or target and masker were modulated out of phase [see Fig. 3(b)]. Uncertainty about the masker level increased SRM in all modulation conditions, but the size of this effect depended on the modulation in the stimuli. For the level-roved stimuli, SRM was 7 dB larger when the target and masker have the same temporal envelope, but only 2 dB larger when the stimuli had different modulation. These results can be understood by considering how and when listeners use overall level, modulation, and spatial cues to detect the presence of the target.

A. Overall level

Detection in the colocated, identically modulated conditions [i.e., when neither modulation nor spatial cues were available for target detection; open circles and triangles in Fig. 3(a)] appears to be based on detecting changes in overall intensity. This conclusion is supported by (1) the observed good match between thresholds in these conditions and predictions from previous intensity JND studies (Experiment 1) and (2) the effect of the intensity rove in these conditions (Experiment 2), which increased detection thresholds to just above that expected for an ideal observer using overall level as the detection cue (Green, 1988). (However, note that there were small gating asynchronies and spectral differences between the target and masker signals that could have contributed to the detection of colocated identically modulated targets.)

B. Space cue alone

When stimuli differed in their spatial locations but not in their modulation [filled circles and triangles in Fig. 3(a)], a consistent improvement in performance was observed, showing that spatial separation provided benefits beyond the improvements in the better-ear TMRs. Changes in the spatial attributes of the target+masker versus masker-only stimuli (such as perceived spatial width) likely were used to detect the target at threshold, a conclusion particularly supported by the fact that the threshold was not influenced by the intensity rove [large and small filled circles and triangles are the same in Fig. 3(a)].

C. Modulation cue alone

Differences in the target and masker modulations led to some improvements in detection when the target and masker had the same location, but not in all conditions. Modulation led to lower thresholds when only the masker was modulated and when the target and masker were modulated out of phase, independent of whether the overall level was roved or not [compare open pentagrams and hexagrams to open circles and triangles in Fig. 3(a)]. When the level was roved, modulation also improved detection when only the target was modulated [compare small open square to small open circle and triangle in Fig. 3(a)]. However, when the level was fixed, the target-only modulation did not improve performance compared to when there were no modulation cues to detect the target [compare large open square to large open circle and triangle in Fig. 3(a)].

The intensity rove caused modest degradations in performance when colocated target and masker had different modulation envelopes, hinting that the listeners might have used the overall level cue (selected at the most favorable TMR instances) instead of the modulation cue in these conditions. However, given that the rove effects were much smaller than when target and masker had identical envelopes, and that the thresholds in these cases were better than (i.e., below) those predicted for an ideal observer using intensity increments to detect the target (Green, 1988), it is unlikely that the listeners used overall level to detect the presence of the target in these conditions [small open squares, pentagrams, and hexagrams in Fig. 4(a)]. Instead, it seems that roving overall level made it slightly harder to judge the changes in modulation caused by adding a target to a masker in these tasks. However, in the target-only modulation condition, the long-term TMR threshold is comparable to that for the no-modulation and in-phase modulation conditions when the level is fixed [large open square, triangle, and circle are comparable in Fig. 3(a)]. Moreover, when the level was not roved, the spatial separation improved performance by similar amounts when only the target was modulated and in the cases where the level was clearly the cue for detection (no modulation, in-phase modulation). Thus, for the target-only modulation condition, it is possible that the subjects used an overall level to detect the target when the level was roved and used a modulation to detect the target when the level varied randomly from interval to interval.

Another result hinting that the subjects' behavior might have been more complex than just detecting the modulation depth is that no similar effect of an intensity rove was seen in a previous study that measured modulation discrimination (Stellmack *et al.*, 2006). However, this difference in the effect of an intensity rove in the two studies may be due to the differences in the instructions given to subjects. In the previous study, listeners were instructed to detect changes in the modulation depth of a single stimulus, while in the current study they were presented with examples of the masker and target at the start of each block and instructed to detect the presence of the target. This priming may have enhanced the likelihood that listeners perceptually segregated the target from the masker in the current study, or that they switched cues between the rove and no-rove experiments, rather than detecting the target+masker interval by perceiving a change in masker attributes. However, further experiments are required to explore which of these alternatives is correct.

D. Space and modulation

Spatial separation did not always improve detection beyond performance for colocated sources after accounting for the TMR at the best frequency in the better acoustic ear. Specifically, spatial separation did not improve detection very much, other than by changing TMR, when the masker envelope had dips, providing good glimpses of the target (in the out-of-phase and masker-only modulation conditions). As noted above, in these conditions, listeners appear to have detected the target by detecting changes in the modulation depth between the masker-only and target+masker intervals, and spatial cues did not help in detecting these modulation changes. However, when the intensity rove was added in these conditions, the modulation-based colocated detection performance was impaired, while the spatially separated performance was not. Thus, spatial cues helped, bringing the spatially separated threshold to the no-rove levels, possibly by making it easier to use the modulation cue optimally.

When only the target was modulated, spatial cues provided a significant improvement in performance both when intensity was fixed across intervals in a trial (Experiment 1) and when intensity was roved (Experiment 2). For these stimuli, listeners were never given a good glimpse of the target, because the masker envelope was constant. In addition, the spatially separated thresholds were almost identical to the thresholds in the no-modulation and in-phase modulation conditions, and the size of the spatial benefit in the no-rove experiment was nearly identical to that in the no-modulation cue conditions. There are two possible explanations for the listeners' behavior in the target-only modulation condition when overall level was not roved. One possibility is that when the target and masker were colocated, listeners used an overall level to detect the target, and when target and masker were spatially separated, listeners used a spatial cue to detect the target. If so, then the modulation and spatial cues were subadditive in the target-only modulation case: listeners either used space or modulation. Alternatively, listeners may have used the modulation cue in the colocated target-only modulated condition and a combination of modulation and space cues in the spatially separated condition. If so, then spatial and modulation cues combined additively for this condition, but were combined subadditively in the masker-only and out-of-phase modulation conditions.

E. Final comments

After accounting for the better-ear acoustic benefit of spatial separation, the current study did not find any evidence for superadditive combination of modulation and space cues for detecting a broadband target embedded in a broadband masker. The results are consistent with two interpretations of the behavior when both cues were available and the level was fixed: (1) the subjects always used one of the cues, getting no benefit from the other one, or (2) the combination of modulation and space cues was additive when only the target was modulated, but the space cue contributed nothing to detection in the conditions in which the masker envelope was modulated and provided glimpses of the target. However, when the overall level was roved, spatial cues always helped performance when modulation was the main detection cue.

These results confirm the first of the proposed hypotheses (H1). The combined effect of modulation and spatial separation on detection is asymmetrical in that spatial separation improves detection performance more when the target is modulated and the masker is unmodulated than when the masker is modulated.

The results contradict our second hypothesis (H2). The combined effect of modulation and separation does not depend on the specific location of the target and masker, even though the contribution of binaural and across-frequency processing likely would vary in the different configurations. This result argues that the combined effect of modulation and spatial cues occurs at a stage that is later in the processing stream than the binaural processing occurring in the brainstem.

In contrast to the current stimuli, everyday auditory scenes contain objects that differ along many more dimensions than just their temporal envelopes and locations. It is difficult to extrapolate these findings to predict how modulation and spatial cues may interact for more complex stimuli. Nonetheless, it is likely that the main result, that modulation and space cues tend to contribute to detection subadditively, will also hold true for other stimuli differing in their spatial positions and modulation structure. However, it is also important to consider how our detection results compare to suprathreshold tasks, such as understanding speech embedded in fluctuating maskers. We find it intriguing that there is essentially no evidence for across-frequency integration in our experiments. In contrast, across-frequency integration is the basis of models that predict speech intelligibility in noise (e.g., see Zurek, 1993). We believe that the key difference between these results is that in our simpler detection task, any glimpse of the target (at any frequency) is sufficient for detection. In contrast, understanding speech requires the integration of information from different frequency bands and estimation of the absolute spectrotemporal content of the speech target. Thus, while the current results may be helpful in predicting how listeners detect a complex signal embedded in a competing fluctuating masker, they are only a first step in understanding how we analyze and understand the content of a complex signal in a setting containing multiple sound sources.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Institute on Deafness and Other Communication Disorders (5R01DC005778-03) and a grant from the Slovak Science Grant Agency (VEGA 1/3134/06). Rich Freyman gave a number of extraordinarily helpful suggestions as editor. In addition, two anonymous reviewers provided very helpful feedback on earlier versions of this work. The authors wish to thank H. Steven Colburn, Constantine Trahiotis, Les Bernstein, Bertrand Delgutte, Chris Mason, Eric Thompson, and Gin Best, for their helpful comments, and Jackie Jacobsen for help with data collection.

APPENDIX: LEARNING

Previous studies show that modulation detection performance improves with training over the course of hours (Wakefield and Viemeister, 1990; Dau and Ewert, 2004; Fitzgerald and Wright, 2005). In the current study, subjects did not receive extensive training prior to the experiment; each performed only one practice session in which thresholds for all conditions were measured once each (25 combinations of modulation and spatial configuration in Experiment 1 and 30 combinations in Experiment 2). To evaluate how learning influenced the results, data were analyzed as a function of the experimental session.

A three-way repeated-measure ANOVA was performed for both experiments on the data collapsed across masker locations [as in Fig. 3(a)], with factors of repeat (five levels), modulation type (five levels), and spatial separation (two levels). For Experiment 1, all two-way interactions were significant (repeat × modulation: $F_{16,96}=2.11$, $p=0.0134$; repeat × separation: $F_{4,24}=6.03$, $p=0.0017$; modulation × separation: $F_{4,24}=230$, $p<0.0001$), as were the main effects of modulation and separation ($p<0.0001$). For Experiment 2, the results were very similar (repeat × modulation: $F_{16,96}=1.69$, $p=0.062$; repeat × separation: $F_{4,24}=20.96$, $p<0.0001$; modulation × separation: $F_{4,24}=212$, $p<0.0001$; main effects of modulation and separation: $p<0.0001$). These results show that performance changes over time, and that the change depends on the specific combinations of modulation and of spatial separation.

Post hoc inspection reveals that the largest changes in SRM over time arose when only the target was modulated and when the target and masker were modulated out of phase. Panel A of Fig. 5 shows the thresholds for these conditions (target-only shown as squares; out-of-phase target and masker modulation shown as hexagrams), collapsed across the masker location and plotted as a function of the repeat, for both spatially colocated (open) and separated (filled) conditions. Panel B shows the SRM. The left-hand and right-hand panels show data from Experiments 1 and 2, respectively. Each symbol represents the across-subject mean (and within-subject 95% CI) of the thresholds obtained for one combination of repeat, spatial configuration, and modulation types.

Overall, TMR thresholds generally improved over time, as illustrated by the downward trend in all the graphs in panel A. However, a more detailed inspection shows that the size of this learning effect differed in the different conditions, and that these differences were consistent across the two experiments. When the stimuli were spatially separated, the target-only modulated thresholds (filled squares) improved by 2–3 dB over the five repeats, while the out-of-phase modulated thresholds (filled hexagrams) improved by 1 dB or less. On the other hand, when the stimuli were colocated, there was a roughly 3 dB improvement in the out-of-phase modulated thresholds (open hexagrams), while the improvement was negligible in the target-only modulated thresholds (open squares). As a result, the SRM tended to increase across sessions for target-only modulation stimuli

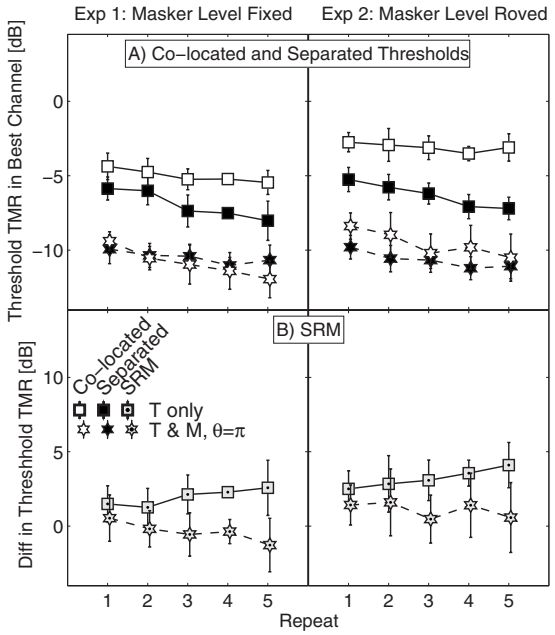


FIG. 5. Threshold TMR in the frequency-corrected best 1/3-octave channel (panel A) and the SRM (panel B) as a function of the measurement repeat in Experiment 1 (left-hand panels) and Experiment 2 (right-hand panels). Panel A: For each repeat, the data represent the across-subject mean (and within-subject 95% confidence interval) of the thresholds collapsed across the corresponding spatially separated or colocated conditions. Panel B: SRM, determined as the difference of the respective thresholds from Panel A.

but to decrease when the target and masker were modulated out of phase (panel B). Thus, while the SRMs for these two conditions differed by only about 1 dB in the first repeat, they differed by more than 4 dB by the fifth repeat.

At first glance, these changes seem difficult to understand. However, as discussed in the main text, spatial cues are generally not helpful for the out-of-phase conditions (hexagrams); in those conditions, performance is based on detecting (nonspatial) changes in modulation. The only effect of spatial cues in the out-of-phase modulation conditions was to make it easier to focus on this change in modulation (e.g., ignoring the distracting effects of intensity rove). Consistent with this, the main effect of learning in the out-of-phase modulation conditions is to improve how well listeners do when there are no spatial cues present and it is most difficult to focus attention on the modulation cue that underlies detection (open hexagrams).

In contrast, in the target-only modulation condition (squares), spatial cues provide a real advantage in target detection and allow detection at lower thresholds than when only monaural modulation and/or level cues are available. In these conditions, listeners improve most in their ability to use this subtle spatial cue (filled squares). However, listeners show little improvement in their ability to detect nonspatial changes in modulation or level with practice (open squares), perhaps because detection of modulation or detection of changes in level increases is a relatively simple task in which near-asymptotic performance is reached much faster (compared to the discrimination of modulation depth or detection of subtle spatial changes). As a result, SRM grows with time for the target-only modulation condition.

[1]The colocated spatial configuration with the target and masker at 90° was not measured in Experiment 1.

[2]In the case of a sinusoidal target, this correction can be computed by considering only the TMR change at the target frequency (Shinn-Cunningham *et al.*, 2005). If the relative contribution of each frequency to task performance is known for a broadband signal, the frequency-dependent TMR function can be used to predict performance (e.g., Zurek, 1993).

[3]This simple linear correction is purely phenomenological, rather than based on theoretical considerations. To the extent that this is the right correction to apply, it may reflect systematic deviations in the degree to which 1/3-octave filters approximate peripheral filtering as a function of frequency, differences in the internal noise of different frequency channels, or other systematic effects of frequency.

[4]Binaural and across-frequency processing may explain some of the dependence of the uncorrected thresholds on the masker locations. Specifically, in the spatially separated configuration of Fig. 2(a), the largest TMRs occur at low frequencies (below 2 kHz, full thick line) and the TMR profile in the right ear is relatively flat as a function of frequency. On the other hand, in the configurations of Figs. 2(b) and 2(c), the best frequency channel is at high frequencies and the TMRs vary significantly with frequency. These differences in the dominant spectral region suggest that binaural and across-frequency processing may contribute more to performance for the conditions of Fig. 2(a) than in the other two configurations, consistent with results in Figs. 1(b) and 1(e) (filled symbols in the *M* 0°, *T* 90° configuration are below the filled symbols for the other two configurations). However, while the binaural and across-frequency processing may explain why threshold TMRs tend to be lower when the masker is at 0° compared to the other configurations [leftmost versus middle and rightmost plots of Fig. 1(b)], they are not analyzed because (1) these factors cannot explain why some spatially separated thresholds are worse than the corresponding colocated thresholds in Figs. 1(b) and 1(e), and 2 the correction based on the frequency-dependent best-channel TMRs accounts for these differences, without considering binaural and across-frequency processing.

[5]Within-subject standard deviations are computed by subtracting out the mean performance (averaged across conditions) for each subject prior to the computation of variability. This method for computing variability is analogous to using subject as a factor in ANOVA analysis. In particular, the remaining variability shows how variable the across-condition results are after removing differences in overall performance across subjects. See the Appendix of Kopco *et al.* (2007) for further descriptions of this analysis.

[6]Comparisons of the current and previous results should be made with caution, as there are important differences in experimental procedures: for instance, none of the previous studies (Viemeister, 1979; Wakefield and Viemeister, 1990; Dau and Ewert, 2004; Dau, 1996) used the 40 Hz modulation frequency adopted in the present study. In addition, the current stimuli differ from the stimuli in the previous studies in their spectral content as they are filtered by the HRIRs.

ANSI (**1986**). *Specifications for Octave-Band and Fractional-Octave-Band Analog and Digital Filters, S1.1 (ASA 65-1986)* (American National Standards Institute, Inc., New York).

Asemi, N., Sugita, Y., and Suzuki, Y. (**2003**). "Auditory search asymmetry between pure tone and temporal fluctuating sounds distributed on the frontal-horizontal plane," Acta. Acust. Acust. **89**, 346–354.

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.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2003**). "The masking level difference for signals placed in masker envelope minima and maxima," J. Acoust. Soc. Am. **114**, 1557–1564.

Buss, E., Hall, J. W., III, 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.

Buus, S., Zhang, L., and Florentine, M. (**1996**). "Stimulus-driven, time-varying weights for comodulation masking release," J. Acoust. Soc. Am. **99**, 2288–2297.

Colburn, H. S. (**1977a**). "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. (**1977b**). "Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise. Supplementary material," J. Acoust. Soc. Am. AIP document no. PAPS JASMA-61–525–98.

Dau, T. (**1996**). "Modeling auditory processing of amplitude modulation," (Universität Oldenburg, Germany).

Dau, T., and Ewert, S. D. (**2004**). "External and internal limitations in amplitude-modulation processing," J. Acoust. Soc. Am. **116**, 478–490.

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.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). "Towards a model for the discrimination of broadband stimuli," J. Acoust. Soc. Am. **80**, 63–72.

Fitzgerald, M. B., and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," J. Acoust. Soc. Am. **118**, 3794–3803.

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.

Gilkey, R. H., and Good, M. D. (**1995**). "Effects of frequency on free-field masking," Hum. Factors **37**, 835–843.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Erlbaum, New York), pp. 349–376.

Green, D. M. (**1988**). *Profile Analysis. Auditory Intensity Discrimination* (Oxford University Press, New York).

Hall, J. W., III, Buss, E., and Grose, J. H. (**2006**). "Binaural comodulation masking release: Effects of masker interaural correlation," J. Acoust. Soc. Am. **120**, 3878–3888.

Hall, J. W., III, Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectrotemporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Herron, T. (**2005**). "C Language Exploratory Analysis of Variance with Enhancements," (Janunary 30, 2005 version. URL: http://www.ebire.org/hcnlab/software/cleave.html, date last viewed: May 19, 2008).

Kidd, G., Jr., Mason, C. R., Brantley, M. A., and Owen, G. A. (**1989**). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. **86**, 1310–1317.

Kopco, N. (**2005**). "Across-frequency integration in spatial release from masking," *Forum Acusticum* (OPAKFI, Budapest, Hungary), pp. 1607–1612.

Kopco, N., Best, V., and Shinn-Cunningham, B. G. (**2007**). "Sound localization with a preceding distractor," J. Acoust. Soc. Am. **121**, 420–432.

Kopco, N., and Shinn-Cunningham, B. G. (**2003**). "Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment," J. Acoust. Soc. Am. **114**, 2856–2870.

Lane, C., Kopco, N., Delgutte, B., Shinn-Cunningham, B., and Colburn, H. (**2004**). "A cat's cocktail party: Psychophysical, neurophysiological, and computational studies of spatial release from masking," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. d. Cheveigne, S. McAdams, and L. Collet (Springer, Dourdan, France), pp. 405–413.

Lane, C. C., and Delgutte, B. (**2005**). "Neural correlates and mechanisms of spatial release from masking: Single-unit and population responses in the inferior colliculus," J. Neurophysiol. **94**, 1180–1198.

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

Mason, C. R., Kidd, G., Hanna, T. E., and Green, D. M. (**1984**). "Profile analysis and level variation," Hear. Res. **13**, 269–275.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego, CA).

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**). "Free-field release from masking," J. Acoust. Soc. Am. **90**, 1355–1370.

Sheft, S., and Yost, W. A. (**1997**). "Binaural modulation detection interference," J. Acoust. Soc. Am. **102**, 1791–1798.

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. (**2005**). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. **117**, 3100–3115.

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.
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers," J. Acoust. Soc. Am. **118**, 2507–2518.
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2006**). "Discrimantion of depth of sinusoidal amplitude modulation with and without roved carrier levels (L)," J. Acoust. Soc. Am. **119**, 37–40.
Sterbing, S. J., D'Angelo, W. R., Ostapoff, E.-M., and Kuwada, S. (**2003**). "Effects of amplitude modulation on the coding of interaural time differences of low-frequency sounds in the inferior colliculus. I. Response properties," J. Neurophysiol. **90**, 2818–2826.
Ury, H. K., and Wiggins, A. D. (**1971**). "Large sample and other multiple comparisons among means," Br. J. Math. Stat. Psychol. **24**, 174–194.
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. (**1998**). "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579.
van de Par, S., Kohlrausch, A., Breebaart, J., and McKinney, M. (**2004**). "Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 398–404.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Wakefield, G. H., and Viemeister, N. F. (**1990**). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," J. Acoust. Soc. Am. **88**, 1367–1373.
Widin, G. P., Viemeister, N. F., and Bacon, S. P. (**1986**). "Effects of forward and simultaneous masking on intensity discrimination," J. Acoust. Soc. Am. **80**, 108–111.
Winter, I. M., Neuert, V., and Verhey, J. L. (**2004**). "Comodulation masking release and the role of wideband inhibition in the cochlear nucleus," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 321–327.
Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hochberg (College-Hill Press, Boston, MA).
Zurek, P. M., Freyman, R. L., and Balakrishnan, U. (**2004**). "Auditory target detection in reverberation," J. Acoust. Soc. Am. **115**, 1609–1620.

On the ability to discriminate Gaussian-noise tokens or random tone-burst complexes[a)]

Tom Goossens[b)]
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Steven van de Par
Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

Armin Kohlrausch
Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands and Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

(Received 2 June 2007; revised 27 June 2008; accepted 16 July 2008)

This study investigated factors that influence a listeners' ability to discriminate Gaussian-noise stimuli in a same-different discrimination paradigm. The first experiment showed that discrimination ability increased with bandwidth for noise durations up to 100 ms. Duration had a nonmonotonic influence on performance, with a decrease in discriminability for stimuli longer than 40 ms. Further experiments investigated the cause for this performance decrease. They showed that discriminability could be improved when using frozen-noise tokens and by instructing listeners to focus on the stimulus endings. A final experiment, using a stimulus consisting of 5 ms Hanning-windowed tone-bursts randomly distributed over time, investigated whether stimulus duration and amount of information differently affect the processing capacity of the auditory system. Results showed that the number of degrees of freedom in the stimulus, not its duration, predominantly influenced the ability to discriminate. Overall, the results suggest that the discrimination performance for acoustic stimuli depends strongly on the amount of information per critical band and the capacity to process this information. This capacity seems to be limited in the temporal dimension, while extending the signal over more auditory filters does have a positive effect on performance.
 [DOI: 10.1121/1.2973184]

PACS number(s): 43.66.Fe, 43.66.Ba, 43.66.Lj [RLF] Pages: 2251–2262

I. INTRODUCTION

In informal listening, two tokens of noise generated by the same statistical process generally sound very similar although their waveforms are completely independent. When asked in a formal experiment to judge whether two presented noise tokens are the same or different, human listeners can respond with above-chance performance, but usually performance will not be perfect.

The ability to perform such a noise discrimination task is a function of the statistical properties of the noise. For instance, the ability to discriminate between tokens of Gaussian noise improves with increasing noise bandwidth for durations up to 100 ms (Hanna, 1984). This improvement is in line with the increase in details in the internal spectrotemporal excitation in the auditory periphery when more auditory channels are excited by the stimulus.

In contrast, the ability to discriminate noisy stimuli does not increase monotonically with an increase in stimulus duration. For example, a 400 ms noise stimulus leads to a longer internal excitation than a 25 ms noise stimulus, and thus the internal representation of a longer stimulus contains more stimulus details than that of a shorter stimulus. One might therefore expect that the ability to discriminate 400 ms noise stimuli is higher than that for 25 ms noise stimuli. Hanna (1984), Heller and Trahiotis (1995), and Sheft and Yost (2004) showed that initially, the ability to discriminate noisy stimuli does increase with duration, but only up to a certain duration. This duration was around 25 ms in the case of Gaussian noise (Hanna, 1984). Beyond this duration, discrimination ability decreases, even though the amount of peripherally represented stimulus details becomes larger. The access to details in the peripheral spectrotemporal excitation pattern is apparently impaired for longer stimuli, which may be related to limitations in more central processes. The nature of this impairment is not well understood at the moment.

A performance impairment for longer stimuli contrasts with observations from signal detection experiments. The temporal and spectral integration of stimulus energy that occurs in a detection task was investigated by van den Brink and Houtgast (1990) using Gaussian tone-burst targets in the presence of a continuous noise masker. They found that this integration of target energy is less efficient than pure energetic integration, specifically for broadband signals. Nevertheless, for all bandwidths, van den Brink and Houtgast (1990) found a temporal integration effect; i.e., for a stimulus with a fixed level, detectability increased with increasing duration. Apparently detection of a known target stimulus and

[a)]Part of this research has been presented at the 14th International Symposium on Hearing (ISH) 2006 (Goossens *et al.*, 2007).
[b)]Electronic mail: tomgoos@gmail.com

discrimination between independent noise tokens are governed by different integration processes, leading to a different dependence on stimulus duration.

We will now more carefully consider how stimulus details are transformed by auditory processing. As a formal definition of information we take the base-two logarithm of the number of stimuli (e.g., noise tokens) that can be distinguished by the most optimal discrimination device (cf. Shannon, 1948). In a same-different task, one could assume that the larger the number of distinguishable stimuli, and hence the amount of information, the more likely it is that two randomly generated stimuli can be distinguished.

Note, however, that on the stimulus level, without assuming any source of uncertainty, e.g., in the form of internal noise, this definition of information is not very useful. Without uncertainty, any two independent tokens of noise will be distinguishable even when only a single sample is considered; thus the amount of information in this situation would be infinite. For example, one single sample of a waveform can be distinguished from another sample, that is an infinitesimally small step higher or lower in level. In real situations, of course, some uncertainty is always present. In the context of auditory stimulus discrimination, the uncertainties in the auditory periphery limit the amount of information to a specific finite amount. To avoid assumptions on the uncertainty at the stimulus level, we will use the number of degrees of freedom of the stimulus rather than the information as defined by Shannon (1948) to characterize the number of potentially distinguishable stimuli. In band-limited noise, the number of degrees of freedom is proportional to the product of bandwidth and duration (Hartley, 1928; Nyquist, 1928).

Within the auditory periphery of a listener the stimuli undergo a series of linear and nonlinear transformations, such as critical-band filtering, hair-cell transduction, auditory nerve encoding, etc., which results in an afferent activation pattern (Dau *et al.*, 1996; Näätänen and Winkler, 1999). Moreover, it is often assumed that some source of internal noise limits the fidelity of this pattern (e.g., de Boer, 1966; Buus, 1990; Dau *et al.*, 1996). The internal noise introduces uncertainty into a discrimination task which reduces the number of distinguishable stimuli. In addition, nonlinear transformations affect the number of distinguishable stimuli. In the remaining data, the number of discriminable stimuli at the level of the afferent activation pattern will be smaller than at the level of the stimuli. We will refer to the amount of information at the level of the afferent activation as *peripheral information*.

In higher stages of the auditory pathway, feature extraction from the afferent activation pattern results in the emergence of sensory feature traces containing cues about, e.g., pitch, loudness, and spatial location (Näätänen and Winkler, 1999). It is likely that also at this level of processing the number of discriminable stimuli is further decreased. This can, for example, be caused by neural processing providing robustness against pitch and duration variation (Patterson *et al.* 2007) or by limitations of the focus of attention (Cowan, 2005).

On a perceptual level, it is useful to speak about cues that can lead to perceptual differences. Some examples of such cues are level, spectral shape, spatial location, spatial compactness, envelope distribution (van de Par and Kohlrausch, 1998; Verhey *et al.*, 2007) and modulation spectrum (Dau *et al.*, 1999). The number of degrees of freedom in a stimulus may influence the type of perceptual cues the listener uses for discrimination. For instance, the shape of the mean spectrum envelope of a very short burst of noise will greatly vary from token to token and hence may provide a good cue for discrimination. However, when the duration is much larger, and hence the number of degrees of freedom is increased, there will be less variability of the mean spectrum of the noise burst. For longer stimuli, the mean spectrum thus provides a less salient cue, which could lead to poorer discrimination performance. Other cues representing more local stimulus properties, such as instantaneous pitch or short-term envelope fluctuations, may therefore be more salient for long-duration stimuli. The cues that are available to the listeners represent what we will refer to as *perceptual information*.

The present study describes a number of discrimination experiments in order to better characterize the nature of perceptual information processing. In particular, this study investigates the decrease in discrimination ability for Gaussian noise with durations above approximately 25 ms.

The nonmonotonic duration dependence observed by Hanna (1984) suggests that an increase in the number of degrees of freedom has a negative influence on discrimination performance for noise tokens longer than 25 ms. Such an interpretation would be in line with the idea that stimulus discrimination is based on cues that reflect more global stimulus properties such as mean spectral envelope. Therefore, we will investigate whether performance increases when listeners are instructed to listen to only a short part of a long-duration stimulus, thus ignoring the rest of the stimulus. In addition, a limitation in the ability to retain the increased amount of peripheral information may be a cause for the impairment in discrimination ability. We therefore studied whether listeners are able to better retain stimuli when they are presented more often.

Since stimulus duration and degrees of freedom are coupled in Gaussian noise, an additional experiment using a stimulus consisting of 5 ms Hanning-windowed tone bursts randomly distributed over time investigated the role of stimulus duration versus number of degrees of freedom by decoupling the two factors. This last experiment thus also addressed whether the nonmonotonic discrimination performance can be related to auditory memory phenomena such as memory decay as a function of elapsed time (see, e.g., Durlach and Braida, 1969).

II. EXPERIMENT 1: TEMPORAL AND SPECTRAL DEPENDENCE OF GAUSSIAN-NOISE DISCRIMINATION

The first experiment is a replication of one of the experiments of Hanna (1984) to verify that listeners in the present study perform similarly. In addition, extra duration conditions were included to get a better indication of the stimulus duration at which discrimination performance is maximal.

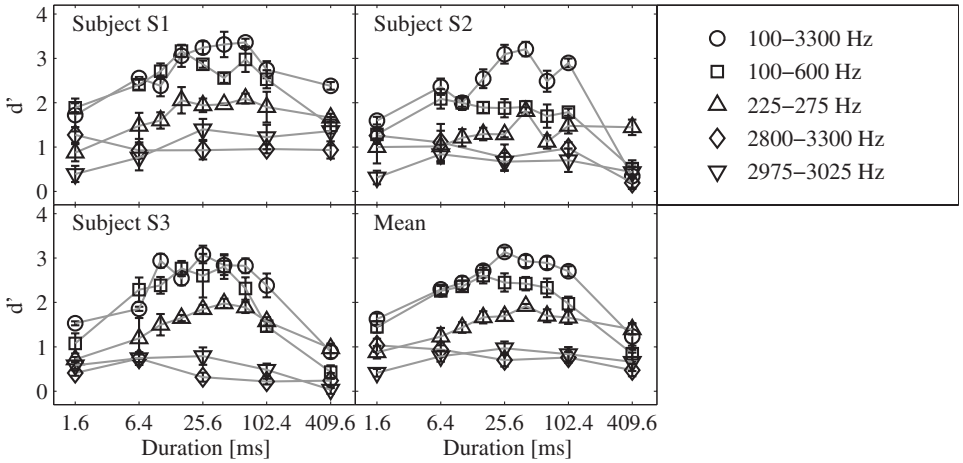


FIG. 1. Mean d' values as a function of stimulus duration for noise with passbands of 100–3300 Hz (circles), 100–600 Hz (squares), 225–275 Hz (triangles), 2800–3300 Hz (diamonds), and 2975–3025 Hz (downward triangles) of individual subjects (upper- and bottom-left panels) and across subjects (bottom-right panel). The error bars indicate plus and minus one standard error of the mean.

A. Method

The experimental method was a same-different procedure where in each trial, two noise tokens were presented to the listener. These noise tokens were either identical or independent. For each trial, new noise samples were generated. Subjects were given feedback about the correctness of their answer after each trial.

Three male subjects participated, including the first (S1) and second authors (S2). The experiments were divided into sessions of maximally 1 h. Each experimental condition (combination of stimulus frequency band and duration) was presented in four blocks of 50 trials (subjects S1 and S2) or three blocks of 100 trials (subject S3). This excludes the training trials. Within a block, half of the trials were same trials and the other half were different trials. The trials within a block were presented in random order. The blocks were also presented in random order.

For each block of trials, the responses of the listeners were transformed into the sensitivity index, d', by calculating percentages correct for the *same* and the *different* trials. These percentages correct were converted to z-scores. Finally, d' was calculated by adding the z-scores of same and different presentations. It sometimes occurred that a subject gave correct answers for all same (or different) trials within a block, resulting in an infinite d' value. In this case an extra artificial incorrect same (or different) trial was added to the block, thus providing a noninfinite d' that could be used for calculating mean d' values and standard errors. For each subject mean d' values and standard errors of the mean were obtained by pooling all d' values of the measured blocks. Similarly, across-subject mean d' values and standard errors of the mean were obtained by pooling all the d' values of the measured blocks of all subjects.

At chance performance, the d' value equals zero. Above-chance performance results in positive d' values, e.g., 69% correct for both same and different trials results in a d' value of approximately 1 and 84% correct for both same and different trials results in a d' value of approximately 2.

Because we observed some training effects, the first 2500 trials for each subject were omitted. In the remaining data, for all subjects, the mean d''s of each repeated set of all conditions were within a margin of ± 0.2 d' relative to their mean d'.

B. Stimuli

The bandpass noise stimuli had five different frequency bands and nine durations. The −3 dB bandpass ranges were 100–3300, 100–600, 225–275, 2800–3300, and 2975–3025 Hz. The specified durations before filtering were 1.6, 6.4, 10.2, 16.1, 25.6, 40.6, 64.5, 102.4, and 409.6 ms. For the two narrowband conditions including 3000 Hz, a subset of these durations was used. The spectrum level was 40 dB sound pressure level (SPL) and the stimuli were presented diotically.

Noise tokens were produced by digitally generating broadband noises of the specified duration and spectrum level with an inter-stimulus interval of 500 ms. The inter-stimulus interval was defined as the temporal separation between the offset of the first stimulus and the onset of the second stimulus within a trial. Subsequently, the tokens were filtered with a Chebyshev type II digital filter with slopes of 100 dB/octave for the broadband and 500 Hz wide bands and approximately 200 dB/octave for the 50 Hz wide bands. The filters were designed using the MATLAB filter design and analysis toolbox. Note that, in the study of Hanna (1984), the filtering was done with analog filters. The stimuli included the ringing of the filters in order to avoid audible truncation effects. The stimuli were presented from a PC through a high-quality soundcard (RME DIGI96/8 PAD) at 16 bit, 44.1 kHz sampling resolution using headphones (Beyerdynamic DT990Pro).

C. Results

Figure 1 shows mean d' values (ordinates) as a function of stimulus duration (abscissas). Curves are shown for five bandwidth and center frequency combinations (symbols) of individual subjects (upper- and bottom-left panels) and the means across subjects (bottom-right panel). The error bars indicate plus and minus one standard error of the mean. The results are generally in agreement with the results of Hanna (1984). However, the average d' was about 0.5 d' units lower in our data.

The curves for the bands containing low-frequency energy (100–3300, 100–600, and 225–275 Hz) show discrimination performance that initially increases with increasing duration. We found a plateau of best performance in the

range 16.1–102.4 ms. The precise location of the plateau depended on subject and spectral range of the noise. For durations above this maximum, discrimination performance decreased with increasing duration.

Discrimination ability for high-frequency conditions (spectral center at 3 kHz, diamonds and downward triangles) was overall poorer than for low-frequency conditions (spectral center at 250 Hz, squares and triangles) and did not show so much evidence for a mid-duration peak. This is not completely in agreement with the data of Hanna (1984), who found that the conditions with low-frequency 50 Hz wide bands and high-frequency 50 Hz wide bands gave very similar results.

In general, for each duration (with the exception of 409.6 ms) an increase in the number of critical bands covered by the noise resulted in higher discrimination performance. The highest performance occurred for the noise with the greatest bandwidth (100–3300 Hz, circles).

For a duration of 409.6 ms, the mean results did not show the highest performance for broadband stimuli. In fact, subject S2 showed a higher performance for the low-frequency 50 Hz wide band. The fact that at a duration of 409.6 ms performance was better for 50 Hz wide bands than for 3200 Hz wide bands was also the case for two of the experiments in the study of Hanna (1984). Hence, there appears to be some evidence that performance is worse for the 100–3300 Hz bands than for the 225–275 Hz bands at a duration of 409.6 ms, or at least that performance is not higher.

III. EXPERIMENT 2: INTERONSET INTERVAL DEPENDENCE OF GAUSSIAN-NOISE DISCRIMINATION

The study of Hanna (1984) and the current study showed that the ability to discriminate broadband noises decreases for durations above 40 ms. A possible explanation for this performance decrease might lie in the increasing temporal separation between corresponding features, e.g., the onsets, in the two noise bursts within a trial. This is a direct consequence of using a fixed offset-onset interval of 500 ms. Arguably, the degradation of discrimination ability for durations above 40 ms shown in experiment 1 may be due to this larger temporal separation of the stimulus features. In the next experiment, the temporal separation of the stimulus features was varied while keeping the stimulus duration fixed at 40.6 ms. The results are compared with results of experiment 1 in order to investigate if the degradation of discrimination ability can be accounted for by the temporal separation.

A. Method

The experimental method was identical to the method of experiment 1. Five subjects, including those from experiment 1, participated in this experiment. All subjects performed each condition in four randomized blocks of 100 trials, of which 50 were same trials and 50 were different trials. The blocks were presented in randomized order.

Because subjects S4 and S5 did not participate in experiment 1, their missing data for variable IOI conditions from Experiment 1 were obtained in a separate session. The conditions in this separate session were presented in four blocks of 100 trials.

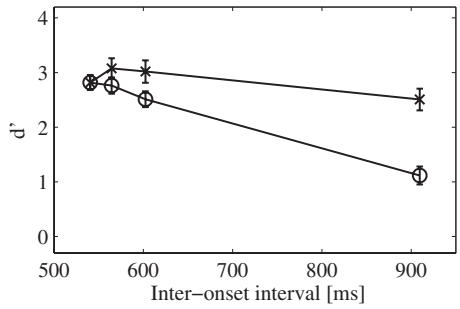


FIG. 2. Across-subject mean d' values as a function of the IOI. The *x symbols* show data for the 100–3300 Hz band with a fixed duration of 40.6 ms and IOIs of 540.6, 564.5, 602.4, and 909.6 from experiment 2. The *circles* show data for the 100–3300 Hz band with durations of 40.6, 64.5, 102.4, and 409.6 ms from experiment 1. The IOIs for these conditions are equivalent to the IOIs for the conditions from experiment 2. The error bars indicate plus and minus one standard error of the mean, based on the results of five subjects.

B. Stimuli

The 40.6 ms, 100–3300 Hz stimulus that resulted in high discrimination performance in experiment 1 was used but was presented with varying interonset intervals (IOIs). The spectrum level was again 40 dB SPL. IOI is defined as the temporal separation between the onset of the first stimulus and the onset of the second stimulus within a trial. In experiment 1 the pause between the two bursts was fixed at 500 ms while varying the duration of the noise bursts. Therefore, the IOIs in each condition of experiment 1 were 500 ms plus the duration of the stimulus. In the current experiment the pause between the two bursts was varied while keeping the stimulus duration fixed at 40.6 ms. In doing so, the IOIs in the current experiment can be chosen to be identical to the IOIs of experiment 1 but with a fixed stimulus duration of 40.6 ms. The conditions in experiment 2 used IOIs of 540.6, 564.5, 602.4, and 909.6 ms, which are equivalent to the IOIs of the 40.6, 64.5, 102.4, and 409.6 ms duration conditions of experiment 1.

C. Results

Figure 2 shows the results of experiment 2 (x symbols) and the data for the 100–3300 Hz band from experiment 1 (circles), plotted as function of their IOI. The 540.6 ms IOI condition is the only condition in which both stimulus duration and IOI were the same in the two experiments. For all other IOIs, the overall stimulus durations are different between the curves. For IOIs above 540.6 ms, the conditions of the current experiment show consistently higher d' values than the conditions with varying durations. Although performance decreases slightly with increased IOI, the intrinsic larger temporal distance between corresponding features in experiment 1 appears not to be a sufficient explanation for the degradation of discrimination ability for stimuli with durations larger than 40.6 ms.

TABLE I. Results of the selective listening experiment, where listeners were asked to focus on the beginning or the end of the stimuli. Data are expressed as mean and standard error of the mean (between brackets) d' values of subjects S1–S6 as well as across-subject means. Gaussian-noise stimulus −3 dB cutoffs were at 100 and 3300 Hz.

	S1	S2	S3	S4	S5	S6	Mean
40.6 ms[a]	3.3 (0.3)	3.2 (0.2)	2.8 (0.2)	2.3 (0.1)	2.4 (0.3)	2.7 (0.3)	2.8 (0.1)
409.6 ms[a]	2.4 (0.1)	0.3 (0.3)	0.9 (0.2)	0.9 (0.1)	0.8 (0.2)	2.1 (0.1)	1.3 (0.2)
409.6 ms begin	2.2 (0.4)	0.3 (0.2)	1.0 (0.2)	0.8 (0.1)	1.6 (0.3)	1.7 (0.2)	1.3 (0.2)
409.6 ms end	2.6 (0.4)	1.5 (0.3)	1.2 (0.3)	1.2 (0.1)	1.5 (0.2)	2.0 (0.3)	1.7 (0.1)

[a]From experiment 1.

IV. EXPERIMENT 3: GAUSSIAN-NOISE DISCRIMINATION WITH SELECTIVE LISTENING

It is remarkable that listeners are unable to perform the discrimination task better for 400 ms stimuli than for 40 ms stimuli, even though there are more degrees of freedom in the longer stimulus. If the decrease in performance cannot be explained by the larger temporal distance between the features, as the previous experiment showed, a surplus of peripheral information might be impairing performance on the discrimination task. Possibly listeners use a suboptimal strategy by trying to retain peripheral information of the complete stimulus. If this were the case, they might be able to improve their performance by focusing on a smaller part of the stimulus when there is too much peripheral information resulting from the entire stimulus.

A relevant experiment in this context is that of Coble and Robinson (1992) using noise bursts that were identical on same trials. On different trials, the bursts were identical except for τ (in milliseconds) where the bursts were independent. They showed that discrimination performance for such partially independent Gaussian noise was dependent on the temporal location of the independent noise parts. Noise discrimination performance was better when the independent part (i.e., the part that supported discrimination) was located at the end of the stimulus than when it was located at the beginning of the stimulus. However, performance for partially independent noise was always lower than for fully independent Gaussian-noise tokens.

In the next experiment, it was investigated whether subjects were able to improve their discrimination performance by concentrating on a part of the stimulus. In contrast to Coble Robinson (1992) the presented stimuli were either fully identical or fully independent. Thus, the stimuli were the same as in the first noise discrimination experiment. Only the instructions to the subjects were different. They were asked to listen only to either the beginning or the ending of the stimulus, which essentially required them to ignore some of the available peripheral information resulting from the stimulus.

A. Method

The experimental method was identical to the method of experiment 1 except that the subjects were explicitly instructed to focus on the beginning (d'_{begin}, begin focus) in one set of blocks and to focus on the end of the stimulus (d'_{end}, end focus) in another set. Six male subjects, including those of experiment 1, participated in this experiment. The two conditions were each assessed in four randomized blocks of 100 trials, of which 50 were same trials and 50 were different trials. The blocks were presented in alternating order.

The results of this experiment were compared to some of the results of experiment 1. However, subjects S4, S5, and S6 did not participate in experiment 1. The results for the 40.6 and 409.6 ms stimuli with a bandpass range of 100–3300 Hz for subjects S4 and S5 were taken from experiment 2. The results for the 40.6 and 409.6 ms stimuli with a bandpass range of 100–3300 Hz for subject S6 were obtained in a separate session, where they were presented in four blocks of 100 trials.

B. Stimuli

The bandpass Gaussian-noise stimuli had −3 dB cutoffs at 100 and 3300 Hz, a duration of 409.6 ms, and a spectrum level of 40 dB. As before, in same trials the two noise tokens were identical and in different trials the two noise tokens were independent. The stimuli were the same as in experiment 1. There was no special begin or end section, nor was there any (visual or acoustic) indication of stimulus sections.

C. Results

The columns of Table I show the individual and across-subject mean d' values and standard errors of the focus experiment. The first two rows show the results for the 40.6 ms and 409.6 ms conditions, with a frequency band of 100–3300 Hz, from experiment 1. The other two rows show the results for the 409.6 ms conditions, with a frequency band of 100–3300 Hz, where listeners focused on the beginning or on the end of the stimuli.

The lowest performance was found for the original 409.6 ms duration condition and the begin-focus condition, both resulting in a mean d' of 1.3. The d' for the end-focus condition was 0.4 higher. The highest d' of 2.8 was achieved for the original 40.6 ms condition. An analysis of variance (ANOVA) combined with *post hoc* Tukey HSD multiple comparisons revealed significant differences ($F_{5,86}=15.9$, $p<0.05$) between all conditions, except between the original 409.6 ms condition from experiment 1 and the condition in which listeners focused on the beginning of the stimulus. Listeners were treated as random effects; thus, differences in baseline performances for the individual listeners were taken into account.

When asked to focus their attention deliberately on the end of the stimulus, most listeners performed the discrimination task better than the original discrimination experiment. Apparently more perceptual information can be retrieved from the stimulus than is typically done by the listeners. When asked for their introspection, listeners reported that the focus experiment was more difficult to perform than the normal discrimination experiment. This could be a reason why most listeners were not inclined to develop this listening strategy independently. Only S1 and S6 may have developed a similar strategy already in the first experiment, as indicated by the small performance improvement when asked to focus. Interestingly, these two subjects also show the highest performance in the original 409.6 ms condition.

Although, on average, the listeners were able to achieve better performance by concentrating only at the end of the stimulus, the effect was relatively small and the overall level of performance was still significantly lower than for the 40.6 ms stimuli of experiment 1. This indicates that listeners could not use the peripheral information that was conveyed by the end of the stimulus in the same way as if it had been presented in isolation in a short stimulus. Focusing on the stimulus beginning did not lead to a significant change in performance relative to the original discrimination experiment.

It is interesting to compare the results of the current experiment with those of Coble and Robinson (1992). They measured discrimination performance for noise tokens (with durations in the range from 25 to 150 ms) where only the beginning, middle, or end of the two noise tokens was independent in a different trial; the rest of the noise tokens was identical. In a same trial the two noise tokens were completely identical. Performance was highest when the stimuli differed at the end. Overall, performance in these conditions was poorer than when the entire stimulus differed. Our experiment showed, however, that performance improved slightly when listeners concentrated on the end of the stimulus. The essential difference between our experiments and those of Coble and Robinson (1992) is that our stimuli were independent across the entire duration of a different trial, while for Coble and Robinson (1992) only a part of the stimulus was independent. Apparently the presence of differences in the unattended part of the stimulus in our experiment influenced discrimination performance when attending to the end part of the stimulus. This finding supports our suggestion that listeners cannot selectively process only a part of the stimulus, but are always influenced by the peripheral information resulting from the entire stimulus.

V. EXPERIMENT 4: FROZEN-GAUSSIAN-NOISE DISCRIMINATION

The selective listening experiment showed that subjects' performance could be somewhat increased when they were instructed to focus their attention on only part of a stimulus. Evidently, the auditory system can retrieve perceptual information more effectively from the stimulus when adopting a better listening strategy. Possibly this improvement in performance is related to a limitation in retaining the peripheral information that is present in a long-duration stimulus (e.g., Cowan, 2001). By instructing the listeners to concentrate on only a part of the stimulus this limitation was partly avoided and performance increased.

In order to investigate the role of limitations in the capacity to retain perceptual information, the next experiment presented the same reference stimulus repeatedly in one block to determine whether listeners are able to form and maintain a consistent internal representation of the reference stimulus. If so, their discrimination performance should increase relative to their performance in the running noise experiment.

A. Method

The experimental method was identical to that of experiment 1. Five male subjects participated, including the subjects from experiment 1. The conditions were assessed in four randomized blocks of 100 trials of which 50 were same trials and 50 were different trials. The blocks were presented in random order.

B. Stimuli

The bandpass Gaussian-noise stimuli had −3 dB cutoffs at 100 and 600 Hz or at 225 and 275 Hz. These bandwidths were selected because listeners were well able to perform the task in the first experiment and their performance was not so high that it would immediately saturate at perfect performance (approximately a d' value of 4). The specified durations before filtering were 1.6, 6.4, 25.6, 102.4, 409.6, and 1638.4 ms. In the previous experiments we replicated the experiments of Hanna (1984) as closely as possible. For the frozen-noise experiment we chose to increase the spectrum level from 40 to 60 dB SPL because mainly the 50 Hz wide stimuli were not so loud.

The stimuli were presented in several repetition configurations that differed in the degree to which they were reused across trails. (1) Running noise: In every trial, new noise tokens were generated. These conditions were a replication of some of the conditions from experiment 1, but with a spectrum level of 60 dB SPL instead of 40 dB SPL. (2) Semifrozen: The first token of each trial within a 100 trial block was always the same (frozen), while the second token was either identical to the first one or a newly generated token. For each block of 100 trials a new frozen-noise token was generated. (3) Frozen: The first token of each trial within a block of 100 trials was always the same, while the second token was either identical to the first one or a different frozen token which remained the same for the entire block of 100 trials. Thus, effectively, only two different noise tokens were used in a block of 100 trials. For each block of 100 trials, two new frozen-noise tokens were generated. All subjects were presented with the same frozen-noise tokens.

C. Results

Figures 3 and 4 show mean d' values (ordinates) as a function of stimulus duration (abscissas). Individual (panels 1–5) and across-subject (bottom-right panel) data are shown for 500 Hz wide (Fig. 3) and 50 Hz wide (Fig. 4) Gaussian-noise bands. The symbols indicate the repetition configura-

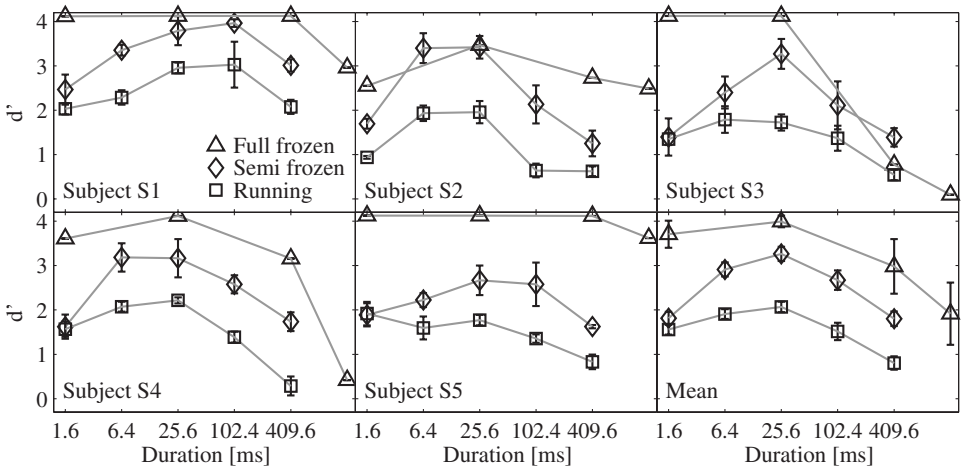


FIG. 3. Discrimination performance for 100–600 Hz Gaussian-noise bands with three repetition configurations: running noise (squares), semifrozen noise (diamonds), and frozen noise (triangles). Mean data across subjects are shown in the bottom-right panel, and data for individual subjects are shown in the other panels. The error bars indicate plus and minus one standard error of the mean.

tions of the stimuli (running, semifrozen, and frozen). The error bars indicate plus and minus one standard error of the mean.

The results show that, for both bandwidths, the ability to discriminate Gaussian noise is directly related to the repetition configurations of the stimuli. Discrimination ability is highest for frozen noise (triangles), lowest for running noise (squares), and intermediate for semifrozen noise (diamonds). It can be concluded that the use of repeated stimuli improved performance in the discrimination task.

Again, as in experiment 1, there was a duration at which discrimination ability was maximal. This duration was approximately 25.6 ms for both bandwidths and did not depend on the repetition configurations. This is consistent with the assumption that the duration at which the maximum occurs is dependent on the number of degrees of freedom per critical band in the stimulus and that the d' for this maximum can be influenced by letting the listener obtain a more accurate representation of the stimulus. Note that the maximum performance was not observed at 40.6 ms because this duration was not used in this experiment. It seems that, in combination with the results of experiment 1, optimal discrimination performance for Gaussian-noise stimuli occurs for durations in the range 16.1–40.6 ms.

Interestingly, for four out of five listeners, the semifrozen-noise tokens with a frequency band of 225–275 Hz and duration of 6.4 ms led to a lower d' than the semifrozen-noise tokens with a duration of 1.6 ms in the same frequency band. Possibly, the presented frozen tokens, which were the same for all the subjects, were by coincidence more difficult to discriminate than the average of the population of possible noise tokens because, for the first stimulus, only four frozen-noise tokens were used in the semifrozen conditions.

The discrimination maximum in experiments 1 and 4 seems to be related to stimulus duration, with about 40 ms being the duration resulting in maximum performance. However, given that stimulus information (number of degrees of freedom, cf. Introduction) increases with duration, the duration at which maximum performance occurs may not be related directly to stimulus duration, but merely an effect of the correlation between the amount of stimulus information and stimulus duration. In the next experiment we wanted to obtain more insight into how the number of degrees of freedom of the stimulus and its duration influence discrimination performance when they were varied independently.

VI. EXPERIMENT 5: DISCRIMINATION OF RANDOM TONE-BURST COMPLEXES

The next experiment used a stimulus comprised of a number of 5 ms Hanning-windowed tone bursts randomly distributed over time, for which, in one set of conditions, the bursts were presented at the same frequency. In another set

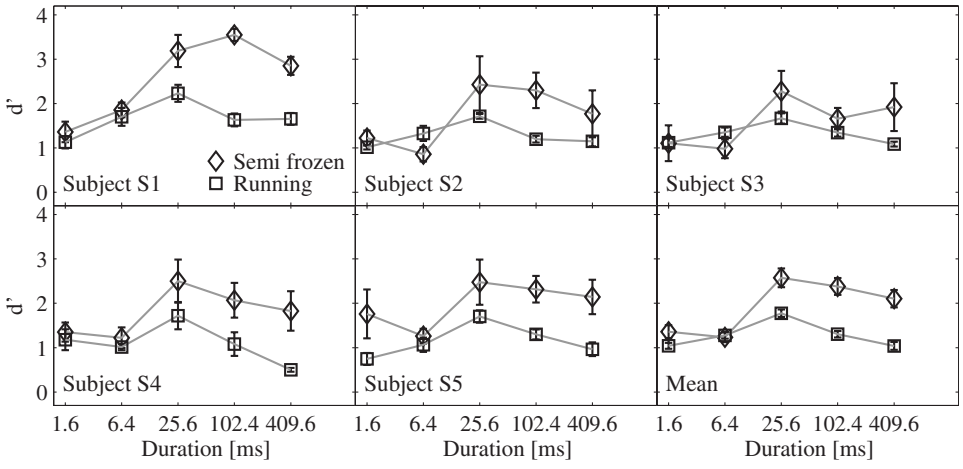


FIG. 4. Discrimination performance for 225–275 Hz Gaussian-noise bands with two repetition configurations: running noise (squares) and semifrozen noise (diamonds). Mean data across subjects are shown in the bottom-right panel, and data for individual subjects are shown in the other panels. The error bars indicate plus and minus one standard error of the mean.

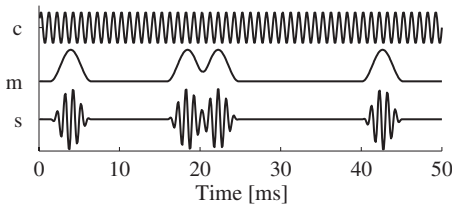


FIG. 5. A random tone-burst signal (s) is generated by multiplying a carrier (c) with a modulation envelope (m) that consists of a number of Hanning windows additively placed at random temporal positions within the duration of the stimulus.

of conditions, seven burst frequencies were used. This stimulus will be referred to as a random tone-burst complex. The number of tone bursts and the number of tone-burst frequencies in these stimuli were varied in order to investigate the influence of number of degrees of freedom on discrimination ability.

Note that whereas for a Gaussian-noise token the number of degrees of freedom is proportional to the product of bandwidth and duration (Hartley, 1928; Nyquist, 1928), for the tone-burst complex, the number of degrees of freedom is proportional to the number of tone bursts and is decoupled from duration.

A. Method

The experimental method was a same/different experiment, identical to the method of experiment 1. Five male subjects participated in this experiment including the subjects from experiment 1. For each subject, the conditions were assessed in four randomized blocks of 100 trials of which 50 were same trials and 50 were different trials. All blocks were presented in random order.

B. Stimuli

Two types of stimuli were used: random tone-burst complexes with tone bursts of only one frequency and random tone-burst complexes with seven frequencies. The stimulus generation is sketched in Fig. 5. For each frequency, tone bursts (s) were produced by multiplying a sinusoidal carrier (c) with an envelope (m). The envelope was comprised of a number of Hanning windows, each with a total duration of 5 ms. The starting points of the Hanning windows were randomly distributed within the full duration of the stimulus, which was either 51.2 or 409.6 ms. It was ensured that the tone bursts fell entirely within these stimulus durations. The random tone-burst complexes with seven frequencies were generated by adding seven independent tone-burst realizations, each with a different carrier frequency.

The peak level of the tone-burst envelopes was 70 dB SPL. In the one-frequency conditions the tone bursts had a (nominal) frequency of 607 Hz (ERB_N number of 12, Glasberg and Moore, 1990). In the seven-frequency conditions the tone bursts had frequencies of 208, 314, 444, 607, 808, 1057, and 1367 Hz (ERB_N numbers of 6 up to and including 18 with a spacing of 2). In both sets of conditions, the numbers of tone bursts were 2, 4, 8, 16, 32, 64, 128, and 256 *per frequency*, with the exception that, for the 51.2 ms duration, the 128 and 256 tone bursts per frequency conditions were not used. This means that, when tone bursts were distributed over seven frequencies, the total numbers of bursts were 14, 28, 56, 112, 224, 448, 896, and 1792.

We chose a relatively short tone-burst duration in order to limit the amount of temporal overlap. Especially for the low-frequency tones, this creates some spectral overlap within the auditory filters that are centered around the different burst frequencies.

C. Results

Figure 6 shows the mean d' values (ordinates) for the random tone-burst complexes as a function of the *total number* of tone bursts (abscissas). This total number is the number of spectral components in the stimulus multiplied by the number of tone bursts per frequency. Data are shown for durations of 51.2 ms (dashed lines) and 409.6 ms (solid lines) with tone bursts of either one frequency (circles) or seven frequencies (x symbols). The upper panels and the bottom-left panels show the individual d' means and the bottom-right panel shows the mean d' across subjects. The error bars indicate plus and minus one standard error of the mean.

In the mean results it can be observed that, in general, discrimination performance for each frequency and duration combination decreased with increasing number of tone bursts. However, for the short-duration conditions, with only one-frequency component (circles, dashed lines), discrimina-

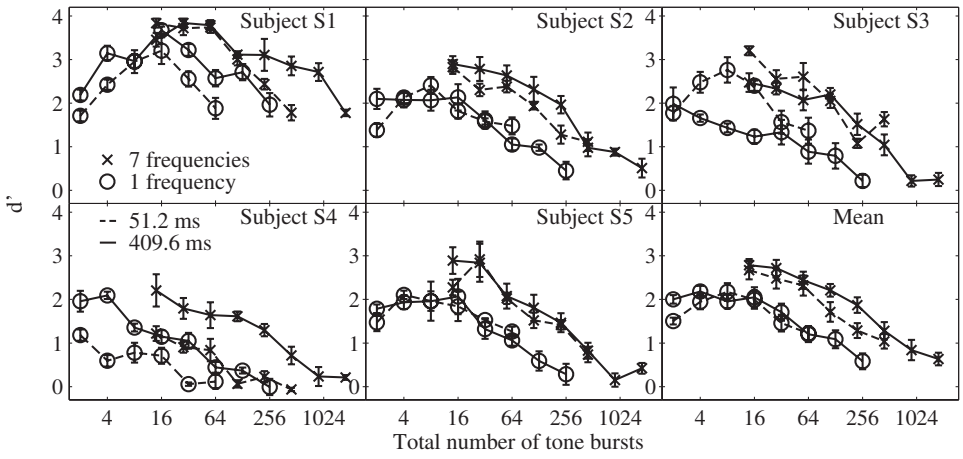


FIG. 6. Mean d' values as a function of the total number of tone bursts for random tone-burst complexes with one frequency (circles) or seven frequencies (x symbols). Stimulus durations were 51.2 ms (dashed lines) or 409.6 ms (solid lines). Means across subjects are shown in the bottom-right panel and individual results in the other panels. The error bars indicate plus and minus one standard error of the mean.

tion performance first increased with increasing number of tone bursts. A comparison of the conditions with circle symbols and with cross symbols for the same total number of tone bursts reveals that discrimination performance was overall higher when the tone bursts were spread over seven frequencies than when they were concentrated at one frequency.

When comparing the solid lines with the dashed lines in Fig. 6, for equal number of frequencies (circles or x symbols), the data for short-duration stimuli showed large overlap with the data for long-duration stimuli, indicating that there is little influence of duration. In the individual data, there were a few deviations from this observation. Specifically, the short-duration stimuli with 4 up to and including 16 tone bursts for subject S3 led to higher performance than for the long-duration stimuli. Furthermore, subject S4 showed generally higher performance for the longer-duration stimuli than for the short-duration stimuli. However, an ANOVA did not reveal a significant influence of duration on discrimination ability for random tone-burst complexes. Significant effects were found for the number of tone bursts ($F_{1,2}=94.2$, $p<0.011$) and the number of frequencies ($F_{1,2}=30.8$, $p<0.031$).

These experiments show that an increase in the number of degrees of freedom (i.e., number of tone bursts) leads to a reduction in discrimination performance, resembling what was seen in experiment 1 for durations in excess of 40 ms. When comparing conditions with equal bandwidth and number of tone bursts but different durations, we see that on average discrimination performance is very similar. Apparently, discrimination performance depends primarily on the number of degrees of freedom and not so much on the duration per se. A similar result was found in a study of Watson *et al.* (1990, p. 2638) in which "… it was shown that listeners' ability to detect *spectral-temporal changes* in randomly generated patterns of multiple nonoverlapping tones is primarily a function of the number of tones in the pattern with very little effect of component duration or of total pattern duration." However, in that study, only one or two tones of each pattern were altered and the patterns were serial, so none of the tones overlapped in time. Therefore, not all tones contributed to discriminability in that study. Rather, the varied number of tones in the pattern affected the relative duration of the informative tones with respect to the total duration of the pattern. This relative duration governed performance in a large variety of tonal pattern paradigms (Kidd and Watson, 1992).

When comparing conditions with different bandwidths in Fig. 6, i.e., complexes with one (circles)- or seven-frequency components (x symbols), we see that distributing a given amount of tone bursts across more frequencies leads to better performance than presenting them all at one frequency. Figure 7, where the average results are plotted as a function of the number of tone bursts *per frequency*, shows that the curves overlap much more than in Fig. 6. By calculating the coefficient of determination[1] (R^2) it was found that 68% of the variance of the data in Fig. 7 could be explained by the number of degrees of freedom per critical band (only using the data for 2 up to and including 64 tone bursts per frequency because these were available for all curves). The portion of variance explained by the duration was 5% and for number of frequency components this was 8%. This shows that adding more degrees of freedom in the spectral dimension did not have a large effect on discrimination. The largest difference between the one- and seven-frequency curves is seen for the two and four bursts per frequency conditions. For these conditions, there seems to be some spectral integration of peripheral information.

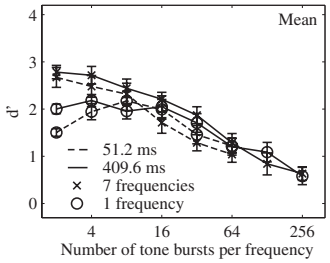


FIG. 7. Mean d' values as a function of the number of tone bursts *per frequency* for random tone-burst complexes with one frequency (circles) or seven frequencies (x symbols). Stimulus durations were 51.2 ms (dashed lines) or 409.6 ms (solid lines).

Interestingly, for short-duration stimuli with one frequency only, there is an initial improvement of performance for small numbers of tone bursts (2–4). This initial improvement is not seen for any of the other conditions. The relatively low performance for the short duration stimuli with one frequency component and two tone bursts may be related to difficulties in perceiving the absolute timing of the tone bursts within the nominal stimulus interval. When only two tone bursts are present within both stimuli of a different trial, it can happen that nearly the same time intervals occur between the two tone bursts within these stimuli, but at a different timing offset relative to the nominal start time of the stimulus. In that case, discrimination purely depends on hearing the absolute timing of the tone-burst *pairs*. For short stimuli with a proportionally larger interstimulus interval, it may be substantially more difficult to hear this difference in timing because of a smaller relative change in timing. For the longer stimulus, absolute timing of bursts can differ more across intervals, thus providing a potentially stronger cue. For stimuli with more tone-burst frequencies, the tone bursts in other frequencies can serve as a reference to compare tone-burst timing.

When the number of bursts is relatively small, the individual bursts are resolved and the most important cue that listeners can use to distinguish different stimuli is given by the timing intervals within a stimulus. However, as the number of degrees of freedom is increased, other cues like temporal envelope modulation and frequency modulation (in the case of seven frequencies) may start to play a role. This is especially so when the number of bursts becomes so large that individual bursts start to overlap. This suggests that the cue a listener uses may be a function of the number of degrees of freedom in the stimulus.

VII. GENERAL DISCUSSION

The first experiment of the current study reproduced the finding of Hanna (1984) that, after an initial increase in discrimination performance with increasing duration up to

40 ms, the ability to discriminate *decreases* for Gaussian-noise tokens with longer durations. As indicated by experiment 2, this decrease could not be understood solely by the larger intrinsic temporal distance that is present between the corresponding features of two noise tokens.

Although there are more degrees of freedom in, e.g., a 400 ms stimulus than in a 40 ms stimulus, the ability to discriminate was higher for the shorter stimulus. This suggests that listeners are better able to retain and compare peripheral information for shorter stimuli. Such an observation is not in good agreement with temporal integration and multiple look models (e.g., Viemeister and Wakefield, 1991; Dau *et al.*, 1996) because such models predict that discrimination ability should increase with available peripheral information in the internal representation. Even if the maximal duration for the accumulation (temporal integration or multiple looks) was restricted to an interval shorter than 400 ms, one would expect that discrimination ability for 400 ms noise tokens should be at least as high as for 40 ms stimuli, but not that it was lower.

Several descriptive models have been proposed for the discrimination of Gaussian noise (Fallon, 1989; Rickert, 1998). However, these models were not aimed at explaining effects of the total stimulus duration and operated at either a fixed total stimulus duration or over a duration range for which there was not a large effect of total duration. These studies were concerned with investigating the effect of temporal location of pieces of uncorrelated noise within a stimulus. One of the major findings was that the uncorrelated parts were more easily detected when they were placed toward the end of the stimulus.

Our experiments bear some resemblance with informational masking studies, which show that the detection of a target stimulus depends on the amount of uncertainty (information) in the masker stimulus (e.g., Watson, 1987; Durlach *et al.*, 2003). In our study there were no maskers and the complete stimulus was the target. Nevertheless, we see the same tendency that when the amount of peripheral information exceeds a certain threshold, discrimination performance decreases.

The duration effect of experiment 1 suggests a limited access to memory for subparts of the stimulus. This is also illustrated by the third experiment, in which listeners were instructed to attend only to the beginning or to the end of a 409.6 ms stimulus. Compared to the condition where no instruction was given to focus on a part of the stimulus (experiment 1), performance was the same when focusing on the beginning of the stimuli and improved only slightly when focusing on the end. However, performance always remained significantly worse than for 40.6 ms duration stimuli. If listeners were able to process an arbitrary 40.6 ms part of the 409.6 ms duration stimulus independently, we would have expected similar discrimination performance as for the 40.6 ms duration stimulus. This limited access to subparts of the stimulus is in line with the idea that, within certain durations, noise bursts are stored in memory as a discrete entity as proposed by Näätänen and Winkler (1999).

In the fourth experiment, when the first noise token in each trial was frozen, discrimination performance improved. This result suggests that the repeated presentation of the "same" stimulus enabled listeners to build up a more accurate internal reference representation. Frozen noise is often used, for example, in detection experiments to investigate the relative contributions of internal and external variabilities (e.g., Buus, 1990). In our experiments, external variability plays a fundamentally different role than in detection experiments. Instead of being a limiting factor for performance, it is the factor that *enables* discrimination. In terms of stimulus variability in itself, one would, on average, not expect a difference in performance between random and frozen noise. The average distance in the internal representations between pairs of frozen noise is the same as between pairs of running noise. The major difference between these two types of experiments is that, for frozen noise, subjects can build up templates of the internal representations of the two noises and also of their difference. In terms of an optimal discrimination process, having a template allows for weighting differences between the stimuli such that they are emphasized at locations where they are expected according to the template. In this way, the influence of the internal noise can be reduced.

The repeated presentation of the same noise links this study to the study of Kaernbach (1993). He investigated the perception of repeated noise, i.e., a continuous noise made by repeating a single piece of noise with a duration of, e.g., 500 ms. For such repeated noise, details of the noisy structure were perceived that would not have been perceived in a nonrepeated presentation. Such details were referred to as "clanks" and "rasping," similar to what listeners reported informally in the current study after doing the experiment with frozen noise.

For Gaussian noise, duration and number of degrees of freedom are inseparably coupled. Therefore, in experiment 5, a stimulus was used consisting of a specified number of tone bursts that were randomly placed within a time frame of either 51.2 or 409.6 ms. In this type of stimulus, the random tone-burst complex, the duration, and the number of degrees of freedom are decoupled, allowing their effects to be studied separately. The results showed that there was not a large effect of duration on the ability to discriminate, but there was a large influence of number of degrees of freedom in the stimulus. In fact, the number of tone bursts per auditory filter appeared to account for the majority of the trends in the results.

An interpretation of the above findings that fits the framework provided by Cowan (2005) is that listeners retrieve or remember the stimuli as stand-alone auditory objects from sensory memory and that there is a fixed and limited amount of resources that can be allocated to retain and process the internal representations of these auditory objects. In working memory, Cowan (2005) attributed this limitation to the focus of attention. The limitation has also been attributed to mechanisms of storage and of processing (cf. Halford *et al.*, 1998).

It is informative to compare the number of degrees of freedom of the stimuli from experiment 1 to those from experiment 5. In Fig. 8 we have plotted the d' values from Fig. 1 (in gray) and from Fig. 7 (in black) as a function of the number of degrees of freedom in the lowest critical band that

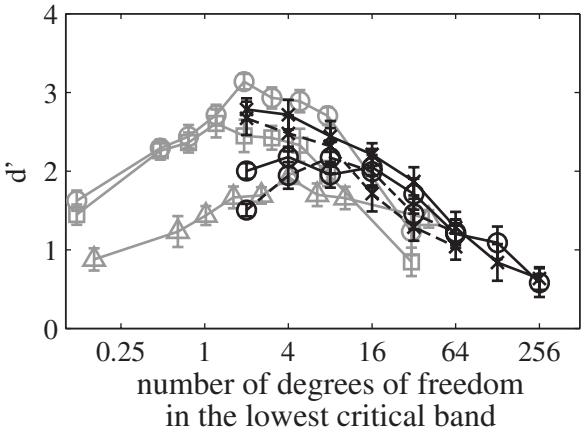


FIG. 8. Mean discrimination performance expressed as d' values as a function of the number of degrees of freedom in the lowest critical band that is excited by the stimulus for Gaussian-noise tokens (from Fig. 1) and random one-burst complexes (from Fig. 7). The gray symbols indicate results for the Gaussian-noise tokens with frequency ranges of 100–3300 Hz (circles), 100–600 Hz (squares), and 225–275 Hz (triangles). The black symbols indicate results for the random tone-burst complexes with one frequency (circles) or with seven frequencies (x symbols) and a duration of 51.2 ms (dashed lines) or 409.6 ms (solid lines).

is excited by the stimulus. For the Gaussian-noise stimuli from experiment 1, the number of degrees of freedom is calculated by taking two times the product of the duration and the width of the lowest excited critical band (cf. Hartley, 1928). The motivation for plotting results as a function of the number of degrees of freedom in the *lowest* critical band was that it had the smallest bandwidth and therefore the lowest number of degrees of freedom. If a large number of degrees of freedom is limiting performance, analyzing this auditory filter should thus lead to the best performance for long-duration stimuli. Note that the duration used in the calculation of the number of degrees of freedom was the duration prior to filtering, which results in a number of degrees of freedom smaller than 1. The actual stimulus durations after filtering were longer. The bandwidth of the lowest excited auditory filter was determined using the ERB_N scale. If this width exceeded the stimulus bandwidth, then the stimulus bandwidth was used. For the random tone-burst complexes from experiment 5, the number of degrees of freedom in the lowest critical band was simply the number of tone bursts per frequency.

Figure 8 shows that the Gaussian noise and the random tone-burst complex stimuli had a comparable range of the number of degrees of freedom, which enables us to compare the conditions of experiments 1 and 5. It appears that, for both types of stimuli, there is a similar upper limit in performance when the number of degrees of freedom per critical band is larger than 8. In this range discrimination performance decreases with increasing number of degrees of freedom with a similar slope. The data may indicate that the number of degrees of freedom per critical band is an important measure that determines the maximum discrimination performance that can be achieved by the listeners. In addition, for both types of stimuli, while there seems to be an advantage of integrating stimulus information across frequency for a low number of degrees of freedom below about eight degrees of freedom, this advantage seems to be lost when the number of degrees of freedom per critical band is too high. This can, for instance, be observed when comparing results for the Gaussian-noise conditions with a bandpass range of 225–275 Hz (gray triangles) to those for the conditions with a bandpass range of 100–600 Hz (gray squares), or, when comparing the results for the one-frequency tone-burst complexes (black circles) with those for the seven-frequency tone-burst complexes (black x symbols). When the number of degrees of freedom was low, discrimination performance was higher for the broadband conditions than for the narrowband conditions. When the number of degrees of freedom was high, discrimination performance was similar for these conditions.

It seems that discrimination performance for these stimuli depends predominantly on the amount of peripheral information of an auditory object and the capacity to process this peripheral information. This capacity seems to be limited in the temporal dimension, leading to a poor discrimination performance when there is a large amount of peripheral information; in the spectral dimension no such limitation was observed. Depending on the number of degrees of freedom in the lowest critical band, performance either increased with increasing number of excited auditory filters or it remained unaffected by additional peripheral information in other critical bands.

ACKNOWLEDGMENTS

We wish to acknowledge Torsten Dau, James F. Juola, Brian C.J. Moore, and three anonymous reviewers for their valuable comments on a previous version on this paper.

[1]Coefficient of determination was calculated with $R^2=1-\Sigma_i(o_i-m_i)^2/\Sigma_i(o_i-\bar{o})^2$, where o_i and m_i are the individual observed and model values, respectively, and $\bar{o}$ is the mean over all observations. As model (m_i), the average of all observations for each value of the variable of interest (i) was taken.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.
Coble, S. F., and Robinson, D. E. (**1992**). "Discriminability of bursts of reproducible noise." J. Acoust. Soc. Am. **92**, 2630–2635.
Cowan, N. (**2001**). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity." Behav. Brain Sci. **24**, 87–185.
Cowan, N. (**2005**). *Working Memory Capacity (Essays in Cognitive Psychology)* (Psychology Press, London).
Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.
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.
de Boer, E. (**1966**). "Intensity discrimination of fluctuating signals," J. Acoust. Soc. Am. **40**, 552–560.
Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.
Durlach, N. I., Mason, C. R., Kidd, G. J., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (l)," J. Acoust. Soc. Am. **113**, 2984–2987.
Fallon, S. M. (**1989**). "Discriminability of bursts of reproducible noise," Ph.D. thesis, Indiana University, Bloomington, IN.
Glasberg, B., and Moore, B. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Goossens, T., van de Par, S., and Kohlrausch, A. (**2007**). "Discriminability of statistically independent Gaussian noise tokens and random tone-burst

complexes," 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, Heidelberg), pp. 343–351.
Halford, G. S., Wilson, W. H., and Phillips, S. (**1998**). "Relational complexity metric is effective when assessments are based on actual cognitive processes," Behav. Brain Sci. **21**, 848–860.
Hanna, T. E. (**1984**). "Discrimination of reproducible noise as a function of bandwidth and duration." Percept. Psychophys. **36**, 409–416.
Hartley, R. (**1928**). "Transmission of information," Bell Syst. Tech. J. **3**, 535–564.
Heller, L. M., and Trahiotis, C. (**1995**). "The discrimination of samples of noise in monotic, diotic, and dichotic conditions." J. Acoust. Soc. Am. **97**, 3775–3781.
Kaernbach, C. (**1993**). "Temporal and spectral basis of the features perceived in repeated noise." J. Acoust. Soc. Am. **94**, 91–97.
Kidd, G. R., and Watson, C. S. (**1992**). "The "proportion-of-the-total-duration rule" for the discrimination of auditory patterns," J. Acoust. Soc. Am. **92**, 3109–3118.
Näätänen, R., and Winkler, I. (**1999**). "The concept of auditory stimulus representation in cognitive neuroscience," Psychol. Bull. **125**, 826–859.
Nyquist, H. (**1928**). "Certain topics in telegraph transmission theory," **47**, 617–644.
Patterson, R., van Dinther, R., and Irino, T. (**2007**). "The robustness of bio-acoustic communication and the role of normalization," in 19th International Congress on Acoustics Madrid, edited by A. Calvo-Manzano, A. Pérez-López, and S. Santiago (unpublished), pp. 1–6.
Rickert, M. E. (**1998**). "Temporal and spectral effects in the auditory discrimination of Gaussian noise samples." Ph.D. thesis, Indiana University, Bloomington, IN.
Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423, & 623–656.
Sheft, S., and Yost, W. (**2004**). "Minimum integration times for processing of amplitude modulation," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Chevegné, S. McAdams, and L. Collet (Springer-Verlag, Berlin), pp. 245–250.
van de Par, S., and Kohlrausch, A. (**1998**). "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579.
van den Brink, W. A. C., and Houtgast, T. (**1990**). "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711.
Verhey, J. L., Rennies, J., and Ernst, S. M. (**2007**). "Influence of envelope distributions on signal detection," Acta. Acust. Acust. **93**, 115–121.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Watson, C. S. (**1987**). "Uncertainty, informational masking, and the capacity of immediate memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hilsdale, NJ), pp. 267–277.
Watson, C. S., Foyle, D. C., and Kidd, G. R. (**1990**). "Limits of auditory pattern discrimination for patterns with various durations and numbers of components," J. Acoust. Soc. Am. **88**, 2631–2638.

Perception of rhythmic grouping depends on auditory experience[a)]

John R. Iversen[b)] and Aniruddh D. Patel
The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, California 92121

Kengo Ohgushi
Kyoto City University of Arts, Kyoto, 610–1197 Japan

(Received 22 June 2007; revised 11 July 2008; accepted 17 July 2008)

Many aspects of perception are known to be shaped by experience, but others are thought to be innate universal properties of the brain. A specific example comes from rhythm perception, where one of the fundamental perceptual operations is the grouping of successive events into higher-level patterns, an operation critical to the perception of language and music. Grouping has long been thought to be governed by innate perceptual principles established a century ago. The current work demonstrates instead that grouping can be strongly dependent on culture. Native English and Japanese speakers were tested for their perception of grouping of simple rhythmic sequences of tones. Members of the two cultures showed different patterns of perceptual grouping, demonstrating that these basic auditory processes are not universal but are shaped by experience. It is suggested that the observed perceptual differences reflect the rhythms of the two languages, and that native language can exert an influence on general auditory perception at a basic level.
 [DOI: 10.1121/1.2973189]

PACS number(s): 43.66.Mk, 43.70.Kv, 43.75.Cd [MSS] Pages: 2263–2271

I. INTRODUCTION

A. Aims

The perception of rhythm is central to how we find structure and meaning in speech and music (Lerdahl and Jackendoff, 1983). In perceiving rhythm, we naturally perceive events as *grouped* into higher-level patterns. Such grouping is an essential step in the interpretation of complex sound sequences. In language, for example, listeners must segment words and phrases from the ongoing speech stream in order to make sense of the incoming signal. In music, comprehending melodic structure involves segmenting tone sequences into smaller coherent chunks (e.g., motives and phrases) in order to discern larger patterns. The human proclivity for auditory grouping is so strong that it is even applied to sequences of physically identical sounds, as, for example, when an electronic metronome is heard as "tick tock" when, in fact, each sound is the same (Bolton, 1894). The fundamental question behind the present work is whether grouping is an innate building block of perception or if instead it is learned from the environment.

The weight of evidence to date has supported the former view, placing grouping together with other basic auditory perceptual mechanisms thought to operate in a bottom-up fashion, reflecting, for example, innate mechanisms that evolved to help animals sort sounds into distinct sources (Bregman, 1990). Two principles, established over a century ago (Bolton, 1894; Woodrow, 1909), are widely accepted and thought to universally govern the subjective grouping of simple sound sequence:

(1) A louder sound tends to mark the beginning of a group;
(2) a lengthened sound (or interval between sounds) tends to mark the end of a group.

These principles have been confirmed in numerous studies since their first description (Vos, 1977; Trainor and Adams, 2000; Hay and Diehl, 2007). Principle 2 is observed also in infants (Trainor and Adams, 2000). Similarities of these principles to Gestalt principles underlying the perception of visual patterns are often noted as additional support for their universality (Wertheimer, 1938; Fraisse, 1982).

The importance of these principles extends beyond the perception of artificial tone sequences and has bearing on language perception and learning. These innate grouping preferences have been proposed to affect aspects of language learning, such as the learning of language segmentation (e.g., Trehub and Trainor, 1993), and to account for the minimal set of rhythmic units in the world's languages (the "iambic-trochaic law;" Hayes, 1995). The principles have recently been shown to apply to speechlike sequences of phonemes in native listeners of English and French, languages with distinct speech rhythms, suggesting universality and relevance to language processing (Hay and Diehl, 2007; Dauer, 1983; Ramus *et al.*, 1999).

The question of whether such grouping principles might instead be determined by experience, which was raised long ago (Jakobson *et al.*, 1952), has only recently been raised again (Drake and Bertrand, 2001) and has received little ex-

[a)]Portions of this work were presented in "Perception of Nonlinguistic Rhythmic Stimuli by American and Japanese Listeners," Proceedings of the International Congress of Acoustics, Kyoto, Japan, April 2004 and "Nonlinguistic rhythm perception depends on culture and reflects the rhythms of speech: Evidence from English and Japanese," Fourth ASA/ASJ Joint Meeting, Honolulu, November, 2006.

[b)]Electronic mail: iversen@nsi.edu

perimental examination. Interestingly, the earliest proposal that nonlinguistic grouping might be shaped by experience had to do with language. In their foundational text Preliminaries to Speech Analysis (1952), Jakobson *et al.* made the following claim: "Knocks produced at even intervals, with every third louder, are perceived as groups of three separated by a pause. The pause is usually claimed by a Czech to fall before the louder knock, by a Frenchman to fall after the louder; while a Pole hears the pause one knock after the louder. The different perceptions correspond exactly to the position of word stress in the languages involved: in Czech the stress is on the initial syllable, in French, on the final and in Polish, on the penult." The groupings suggested by Jakobson *et al.* (1952) can be schematically represented as follows, where each x represents a knock and the upper case X's are louder

$$\begin{aligned} \ldots \text{X x x X x x X} \ldots &= (\text{X x x}) \quad \text{Czech} \\ &= (\text{x x X}) \quad \text{French} \\ &= (\text{x X x}) \quad \text{Polish}. \end{aligned}$$

While this claim is provocative, there has been no empirical evidence to support it. Bell (1977) attempted to test the claim by examining rhythmic grouping perception in tone sequences with every third tone differing in intensity, pitch, or duration. He found no significant differences in grouping perception among native speakers of English, Bengali, French, Persian, and Polish. However, the sample sizes in this study were small, and there were confounds in the stimuli, which made the results hard to interpret. In contrast to the proposal of Jakobson *et al.* (1952), further studies have since confirmed the universality of grouping principles for speakers of English, French, and Dutch (Hay and Diehl, 2007; Vos, 1997).

There are several reasons, however, to suggest that experience could play a role in shaping the perception of rhythmic grouping. The first, and most direct, is an empirical discovery made by Kusumoto and Moreton (1997), who found a difference between American and Japanese listeners with regard to rhythmic grouping in simple tone sequences. In a study conducted in the United States, they presented listeners with sequences of tones alternating in amplitude or duration. Members of both cultures perceived sequences with alternating loud and soft tones as repeating "loud-soft" groups of two elements. In contrast, a cultural difference was found for sequences with tones that alternated in duration: Americans showed a strong bias for hearing a repeating "short-long" group of two elements, while Japanese speakers did not. In fact, many Japanese reported strongly hearing the opposite "long-short" pattern. This challenged the idea that grouping perception follows universal principles and motivated the current study's focus on rhythm perception by English versus Japanese speakers.

Several additional findings provide reasons to suspect that language experience could shape nonlinguistic rhythm perception. The first concerns psycholinguistic research on segmenting words from connected speech. Such research has indicated that experience with a language's rhythm leaves a permanent influence on a listener in terms of rhythmic segmentation strategies. That is, segmentation strategies (such as the "metrical segmentation strategy" in English, whereby a native listener posits a word boundary at every stressed syllable) are a perceptual habit of a listener and are even applied (inappropriately) to languages with other rhythmic patterns (see Cutler, 2000 for a review). This raises the possibility that strategies developed for segmenting meaningful linguistic units from connected speech could carry over and also influence how one segments nonlinguistic rhythmic patterns.

The second concerns evidence for connections between linguistic and nonlinguistic rhythmic patterning. Patel and Daniele (2003) found that timing patterns in British English and continental French speech are reflected in the classical instrumental music of these two cultures (see also Huron and Ollen, 2003). While that finding concerned rhythmic production rather than perception (and timing rather than grouping), the relevant point concerns the mechanism proposed to account for this cross-domain similarity. This mechanism was statistical learning of the rhythms of the native language, i.e., tracking rhythmic patterns in language and acquiring implicit knowledge of their statistical properties, without any direct feedback (see Patel *et al.*, 2006). It was proposed that this implicit knowledge could influence the creation of rhythmic patterns in a nonlinguistic domain (music). The current study broadens this idea to consider whether statistical learning of the rhythms of speech might be responsible for shaping basic auditory grouping biases, which reveal themselves in the perception of tone sequences that are simpler than either speech or music. Thus the issue at hand is whether learning the characteristic rhythms of meaningful units in the auditory environment (which is dominated by speech for humans) can shape low-level rhythm perception mechanisms.

Finally, infants show a remarkable sensitivity to the rhythms of their native language and music (Mehler *et al.*, 1996; Nazzi *et al.*, 1998; Hannon and Trehub, 2005), suggesting that mechanisms for learning rhythmic regularities from language and music exist from birth (Mehler *et al.*, 1996; Nazzi *et al.*, 1998).

Thus, there are several reasons to revisit the question of how experience shapes low-level perception of rhythm. Recently, the issue of top-down influences on basic auditory processing has received renewed attention (see Davis and Johnsrude, 2007). To date, however, only a few studies have addressed how language might influence basic auditory processes, and these studies have focused on the perception of pitch (i.e., Bent *et al.* 2006; see Deutsch, 1991). The current study complements these studies by its focus on rhythm.

B. Overview

In the present work, perceptual grouping is studied in Japanese and English speakers using simple tone sequences in which sounds alternate in a single parameter (amplitude or duration). The use of simple tone sequences follows the tradition in auditory research that first established the grouping principles under study (Bolton, 1894; Woodrow, 1909). In contrast to previous work (which did not test non-Western listeners), the present study finds that even simple grouping

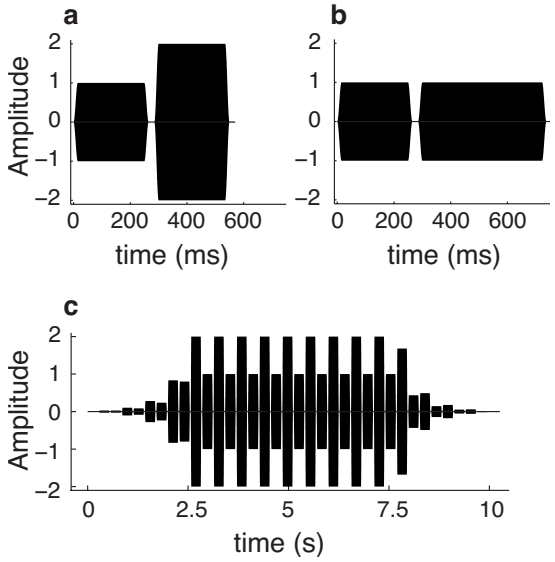


FIG. 1. Amplitude and duration stimulus sequences. Every second tone had a different amplitude or duration. Stimuli consisted of a basic tone (500 Hz, complex; 150 or 250 ms duration) alternating with a second tone of increased amplitude (1.5 or 2 times greater) or increased duration (1.25, 1.5, or 3 times greater). (a) Example of tone pair from an amplitude sequence (basic duration 250 ms, amplitude ratio=2). (b) Example of a tone pair from a duration sequence (basic duration 250 ms, duration ratio 1.5). (c) Example of a complete amplitude sequence. See text for additional stimulus details.

perception depends on culture: Americans and Japanese hear different grouping patterns in identical sequences of sound.

The finding of a cultural difference motivates the question of what the source of the difference might be. Under the assumption that the difference in grouping perception reflects auditory experience (rather than innate factors), this paper proposes the hypothesis that experience with the most common auditory input, speech, underlies this more general perceptual difference.

II. EXPERIMENT: PERCEPTION OF RHYTHMIC GROUPING

A. Introduction

To test the universality of grouping preferences, native speakers of English or Japanese were presented with repeating sequences of two alternating tones in which the second tone had either increased amplitude or duration relative to the first tone, at one of several fixed ratios (Fig. 1). Listeners were asked to report their perceived grouping of the stimulus (e.g., long-short or short-long), and also to indicate how confident they were of their judgment. If grouping principles (1) and (2) above (Sec. I A) are indeed universal, one would predict that all listeners would perceive tones of alternating amplitude as repeating loud-soft groups, with the higher amplitude tone beginning the group, and would perceive tones of alternating duration as short-long, with the longer duration tone ending the group.

As noted in Sec. I, similar research in the past has focused on listeners from Western European cultures and has found cross-cultural agreement in terms of grouping preferences. Kusumoto and Moreton (1997) were the first to test individuals from an East Asian culture and found a cultural difference, motivating the current research. The present research extends these findings by using native Japanese speakers living in Japan and with no foreign living experience, by having larger sample sizes and by collecting confidence ratings on listeners' perceptual judgments. These changes are intended to increase the reliability of the results and enable meaningful individual-listener analysis in addition to group-level analyses.

B. Methods

Stimuli were 10 s sequences in which tones alternated in either amplitude (loud-soft) or duration (long-short) (henceforth "amplitude sequences" and "duration sequences;" see Fig. 1). The basic tone was a 500 Hz complex tone (15 ms rise/fall) with a duration of either 150 or 250 ms. The complex tone was a low-pass square wave consisting of the fundamental and first three odd harmonics and was constructed to match previous work on grouping (Kusumoto and Moreton, 1997; Trainor and Adams, 2000). The alternating tone was constructed by multiplying either the amplitude or duration of the basic tone by one of several small ratios (amplitude: 1.5 or 2; duration: 1.25, 1.75, or 3). The two tones alternated with a gap of 20 ms between them. To mask possible effects of starting tone order, stimuli were slowly faded in (and out) over 2.5 s (double-logarithmic ramp), and each sequence was presented both forward and reversed, yielding a total of 20 sequences (2 base durations $\times$ 2 orders of presentation [forward and reversed] $\times$ 5 ratio parameters [2 amplitude ratios + 3 duration ratios], hence four different sequences per ratio). Stimuli were generated in MATLAB, at CD quality (44.1 kHz sample rate) and were presented in free field in a classroom setting. Sound levels were not measured, but were verified to be easily audible to all participants (see EPAPS).

Listeners were familiarized with the experiment by hearing one example sequence of the two types (but with different parameters than those used in the experiment). The experimental sequences were then presented in random order, and listeners were asked to indicate their perceived grouping by circling pairs of tones on diagrams schematically depicting the stimuli, in answer booklets with one page per sequence. The starting tone (e.g., long or short) of diagrams in the booklets was counterbalanced to avoid any possible visual bias due to the diagrams. Listeners also rated the certainty of each judgment, ranging from 3 (completely certain) to 1 (guessing). The experiment was conducted in a classroom setting, enabling data to be collected in parallel for multiple participants: 43 native American English speakers) (in San Diego; henceforth referred as "English speakers") and 46 native speakers of Japanese (in Kyoto; all were speakers of the Kansai dialect of Japanese; three participants were excluded because they had lived abroad for more than six months). Participants were college aged 17–25 years. Musical training (years studied) was similar in the two groups (English: 5.0 ± 4.7 years, Japanese: 7.0 ± 5.8 years; n.s., $p=0.09$, unpaired t-test). Although study of English is part of compulsory education in Japan, none of the native Japanese speakers rated themselves as "highly proficient"

speakers of English. None of the native speakers of English rated themselves as proficient in Japanese. All instructions, both verbal and written on the answer booklets, were given in the native language. To test the robustness of our findings for Japanese listeners, replication studies were conducted in an identical manner in different regions of Japan (Niigata, $n=38$; Tokyo, $n=54$).

To analyze the behavioral data, each participant's grouping response and confidence rating for each of the 20 stimuli was tabulated. Only the completely certain (level 3) responses (55% of all responses) were analyzed further. [Responses with intermediate certainty (31% of all responses) showed the same pattern of results as the high-certainty responses, but with slightly weaker preference, while those marked as guesses (14%) showed no preference.] Confidence ratings showed no evidence of differences by culture: 91% of Japanese listeners had at least one confident response to the duration task, compared with 87% of English listeners. Listeners varied in the number of high-certainty responses, but the mean number of high-certainty responses was not significantly different between cultures for either type of sequence [amplitude: 3.7 (s.d. 2.6) for English speakers and 3.8 (s.d. 2.6) for Japanese speakers (out of a possible 8); duration: 7.3 (s.d. 3.3) for English speakers and 5.7 (s.d. 3.6) for Japanese (out of a possible 12)].

For both amplitude and duration sequences, individual listener responses were pooled across the four stimulus sequences for each ratio, as neither basic tone duration nor sequence order had a significant effect on responses. For amplitude sequences, the percentage of each participant's high-certainty responses in which a soft-loud grouping was reported was calculated for each of the two amplitude ratios. Similarly, for duration sequences, the percentage of high-certainty short-long responses was found, for each of the three duration ratios. With individual average responses for each participant in hand, group means for each of the stimulus conditions were computed by averaging across participants.

In the analysis, all participants are weighted equally, independent of the number of high-confidence responses they made. To examine any potential bias due to weighting individuals with few high-confidence responses equally with those with many high-confidence responses, all analyses of the original studies were repeated including only participants who gave high-confidence responses to at least half of the stimuli (i.e., high-confidence responses to at least 4 of the 8 amplitude sequences or 6 of the 12 duration sequences). This excluded an average of 42% of participants from each language group (35% of English; 48% of Japanese). Results from this restricted analysis were statistically indistinguishable from those computed on the entire set of participants (all $p>0.38$, Mann-Whitney U), suggesting that including participants with fewer responses does not bias the results. Consequently, we used data from all participants.

C. Results

The perception of amplitude sequences was similar for all listeners, consistent with the universality of the first grouping principle: both Japanese and English listeners preferred a loud-soft grouping [Fig. 2(a)]. For both groups, the preference was highly significant (compared to no preference; Wilcoxon signed-rank test, pooled across ratio: English $p<0.001$; Japanese $p<0.0001$). As a group, the Japanese listeners had a significantly stronger preference for loud-soft grouping (91% of all responses; mean across ratios) than the English listeners [68% of all responses; two-way analysis of variance (ANOVA) on language and ratio; $F(1,139)=17.8$, $p<0.0001$]. There was no significant effect of amplitude ratio [$F(1,139)=0.69$, $p=0.41$). English and Japanese listeners had nearly identical distributions of confidence ratings, with 50% and 51% of responses receiving the highest confidence.

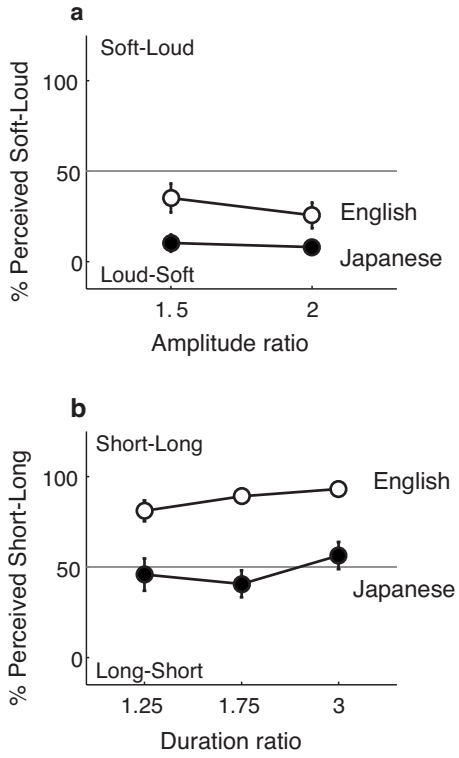


FIG. 2. Grouping preferences as a function of native language. (a) Amplitude sequence grouping preferences of English (open symbols, $n=43$) and Japanese (filled symbols, $n=43$) listeners. The across-subject mean ($\pm$ s.e.) percentage of sequences heard grouped as soft-loud is shown for two amplitude ratios. Both language groups strongly preferred a loud-soft grouping. (b) Duration sequence grouping preferences. There is a large difference between language groups. English listeners strongly preferred a short-long grouping preference, while Japanese listeners on average did not show a preference.

In contrast, a striking difference between English and Japanese listeners was found in the perception of duration sequences [Fig. 2(b)]. English listeners showed a strong preference for short-long grouping (89% of all responses; $p<0.0001$, Wilcoxon signed-rank test), as predicted by the grouping principles, but Japanese listeners as a group showed no preference ($p=0.39$). The differences between the language groups were highly significant [$F(1,219)=85.8$, $p<0.0001$]. On the strength of this result, then, the second grouping principle appears not to be universal.

Additional insight into the patterns of response may be gained by examining the distribution of individual participants' mean responses for the amplitude and duration stimuli

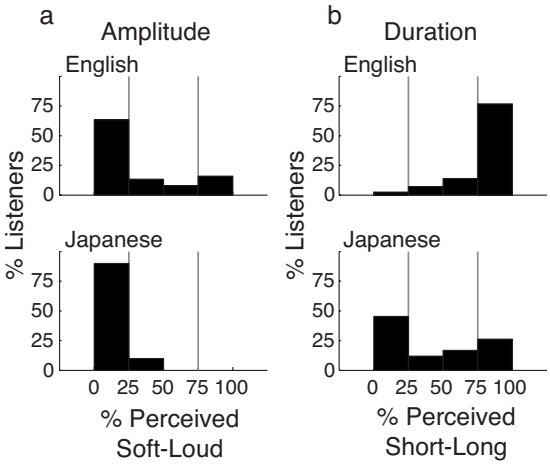


FIG. 3. Distribution of individual listeners' grouping preferences. (a) Distribution of amplitude sequence grouping preferences of English (top) and Japanese (bottom) listeners. The distribution shows each listener's percentage of all amplitude sequences heard as soft-loud. For both English and Japanese listeners, the majority preferred a loud-soft grouping (63% of English listeners; 90% of Japanese listeners), with most participants choosing this preference consistently for all stimuli. (b) Distribution of duration grouping preference for English (top) and Japanese (bottom) listeners. The distribution shows each listener's percentage of all duration sequences heard as short-long. The large majority of English listeners chose a short-long grouping (77%). In contrast, nearly half (45%) of the Japanese listeners chose the opposite long-short grouping, which English speakers almost never chose. Japanese responses were more varied, and 26% of Japanese listeners chose the short-long grouping. Thin vertical lines delineate regions defined as strong preference (0%–25% and 75%–100% preference).

pooled across ratio, base duration, and order of presentation [Figs. 3(a) and 3(b), respectively]. When data are pooled in this way, each participant simply responded to 8 amplitude sequences and 12 duration sequences. [Thus a participant who had high-confidence responses to 10 out of 12 duration sequences, and responded short-long in all such cases, would contribute a point to the 100% column of the histogram in Fig. 3(b)].

For amplitude sequences, the distributions for English [Fig. 3(a), top] and Japanese listeners [Fig. 3(a), bottom] are highly skewed toward a loud-soft grouping, as would have been expected from the group means shown in Fig. 2(a). The majority of participants in both groups had strong preferences for the loud-soft grouping, where strong preference is defined as a preference score in the extreme 1/4 of the response scale (e.g., strong preference for loud-soft is any response score $<=25\%$).

The distribution of individual participant responses to duration sequences is highly illuminating, because it shows that the majority of individual Japanese listeners do have strong grouping preferences [Fig. 3(b)]. Thus the Japanese group result of no preference seen in Fig. 2(b) is not representative of individual perception. Specifically, while the distribution for English listeners shows that a majority (77%) have strong preference for short-long grouping [Fig. 3(b), top] in accord with the group results, Japanese individual listeners, in contrast, had more widely distributed grouping preferences [Fig. 3(b), bottom]. The distribution of Japanese responses is highly significantly different from a uniform distribution (which would be the case if all Japanese had no preference and were responding randomly; Kolmogorov–Smirnov test, $p<0.005$, $df=41$). Instead, most Japanese listeners (71%) had strong preferences for either short-long or long-short grouping. Japanese preferences were not equally distributed, however, but were significantly biased toward long-short grouping ($p<0.001$, $df=41$, Wilcoxon signed-rank, null hypothesis of equal preferences for both groupings). Indeed, 45% of Japanese listeners had a strong preference for the long-short grouping, versus 26% for short-long grouping. The substantial proportion of listeners with a long-short grouping preference is striking given that only a single English listener (<2% of the population) expressed any perception of long-short grouping [Fig. 3(b), top panel].

Two replication studies on an additional 92 Japanese listeners living in different parts of Japan (Niigata and Tokyo) were conducted to test the consistency of the Japanese grouping results. This brings the number of Japanese participants to 135. The results of these studies were indistinguishable from those seen for Kyoto listeners (ANOVA on individual participants' mean response with test location as factor. [amplitude sequences: $F(2,115)=0.094$, $p=0.91$; duration sequences: $F(2,121)=1.42$, $p=0.25$]. This replication suggests that the Japanese result is reliable.

D. Discussion

With respect to the purported universality of grouping principles (Sec. I A), both Japanese and English listeners followed the first principle (grouping based on amplitude), but only English listeners followed the second principle (grouping based on duration). Japanese listeners showed variation across individuals in their duration-based grouping preferences, but the majority of individuals expressed a strong preference for one or the other grouping. A salient finding was that the most common Japanese preference was for long-short grouping, a choice made by virtually no English listeners. This result is incompatible with the universality of rhythmic grouping principles. The current findings confirm and extend those of Kusumoto and Moreton (1997). As they run counter to long-held claims for universality, two additional experiments in Japan were conducted, which confirmed the findings of the initial study. Hence the finding appears robust, leading to the question of how this cultural difference in perception arises.

Before turning to this question in the general discussion, it is worth addressing a possible methodological concern about the task we used to assess grouping. The current study used a metalinguistic task to probe grouping preferences (circling of images in answer booklets), raising the question of how the results might compare to tasks that use other perceptual methods to probe grouping. This question can be answered for English listeners because their duration-based grouping preferences have been explored by Trainor and Adams (2000) using a purely perceptual gap-detection paradigm. The task required listeners to detect a short gap placed within a repeating sequence of short-short-long tones. It was found that the gap was more detectable when following one of the short tones than it was following the long tone. It was inferred that the perceived grouping boundary is after the long tone (assuming that a gap that violates grouping struc-

ture would be more noticeable). The results of that study converge with the current findings in showing a preference for short-long grouping among English listeners. Hence, it is reasonable to expect that perceptual measures of grouping would replicate the findings reported here for English versus Japanese listeners, though further work is needed to test this empirically.

III. GENERAL DISCUSSION

The current study focused on the perception of grouping, a basic aspect of rhythm perception by which sequences are perceptually segmented into higher-order units. In contrast to much previous research, a cultural difference in the perception of grouping in simple tone patterns was demonstrated: Native English speakers consistently perceived sequences of alternating-duration tones as repeating short-long groups; native Japanese speakers most often perceived the opposite long-short grouping (a grouping preference that almost never occurred among English speakers) and were as a group more variable in their preferences. The observed difference in perception between English and Japanese speakers is notable because the principles underlying auditory grouping have long been thought to be innate and universal.

How can the difference be explained? We assume that the key factor is the different auditory experiences of listeners living in America versus Japan (rather than, say, a different genetic background). The most obvious source of cultural differences in auditory experience is the dominant language of the culture. Two questions must be addressed: What is it in the experience of English speakers that causes them to prefer short-long grouping of tones? How does the experience of Japanese speakers create many listeners with the opposite long-short preference? Below, we propose the hypothesis that language experience is the source of differences in grouping perception. The theoretical perspective adopted is that statistical learning of the duration patterns of common rhythmic units in speech is responsible for shaping low-level grouping biases. Such learning presumably occurs automatically, via exposure to the rhythmic patterns of speech (see Patel *et al.*, 2006), where gaining implicit knowledge of these patterns may serve a useful role in learning word order, or in segmentation of meaningful units from the speech stream.

If speech rhythm is the key factor in shaping the duration-based grouping biases, then linguistic rhythms in English and Japanese should predict these biases. In particular, short-long patterns should be prevalent in English, while Japanese would be expected to be biased toward long-short patterns (though perhaps not as strongly, in light of the lower conformity in Japanese responses). This approach follows the style of proposals made by Jakobson *et al.* (1952) and Kusumoto and Moreton (1997) in that language is proposed to be responsible for shaping basic rhythmic biases. However, while these researchers argued that the rhythmic shape of words was the key factor in driving nonlinguistic grouping preferences, the current study focuses on a different hypothesis, namely, that the most important linguistic unit influencing grouping perception is not the word, but rather the phrase. Specifically, it is proposed that a syntactic parameter influencing the word ordering of language ("head direction") ultimately influences grouping perception, because of the way it influences the rhythm of phrases in Japanese and English.

A. Rhythmic differences between English and Japanese

Broadly speaking, head direction refers to the order of various structural constituents in sentences, such as verbs and their objects and function-content word ordering (Baker, 2001; Nespor and Vogel, 1986; Nespor *et al.*, 1996). English, for example, is a head-complement language, in which verbs precede objects; Japanese is a complement-head language in which objects precede verbs. We focus here on the consequence of head direction on function-content word order and suggest that this ordering could drive the perceptual grouping of nonlinguistic sounds.

The distinction between function words and content words is a universal feature of language (Selkirk, 1984; Shi *et al.*, 1998). Function words (functors), such as articles, prepositions, conjunctions, and pronouns, are high-frequency linguistic elements from a small closed set of words that provide a structural framework for the semantically meaningful content words of a sentence. In English, functors typically *precede* their syntactically related content word, as in "the dog," "to eat," etc., Japanese, in contrast, typically places functors *after* the related content word (Baker, 2001). (Examples of function morphemes in Japanese include case marking particles such as "ga" and "wo" or "ni," which indicate whether a noun is a subject, a direct object, or an indirect object). Notably, the functor/content distinction is available to infants at an early age, well before individual words are recognized (Shi, *et al.*, 2006), and sensitivity to the relative order of functors and content words emerges within the first year of life (Gervain *et al.*, 2008).

A key element of the proposal is the suggestion that functor/content ordering will have acoustic consequences that predict grouping preferences. Functors tend to be monosyllabic and phonologically reduced (Selkirk, 1996; Shi *et al.*, 1998; Shi *et al.*, 2006). It seems likely then that the syntactic difference in functor location will result in an acoustic difference in the typical rhythmic shape of functor + content word phrases (Nespor *et al.*, 1996; Christophe *et al.*, 2003). If the functor has a shorter duration than its associated content word, which is plausible, the syntactic difference in placement of functors would lead to characteristic temporal rhythms in the two languages that match perceptual grouping preferences: short-long in English and long-short in Japanese. To illustrate this, consider short phrases corresponding most closely to the two-element rhythms used in the perceptual study, namely, two-syllable phrases consisting of a monosyllabic function word plus a monosyllabic content word, e.g., "the book" in English versus "hon ga" in Japanese (book+subject marker). Given the difference in placement of functors between English and Japanese, one can make a strong prediction that such short two-syllable phrases will have a short-long duration pattern in English but a long-short duration pattern in Japanese.[1] Thus, the proposal is that

the acoustic-level ramifications of the syntactic organization of word ordering shape nonlinguistic grouping. Such an argument extends more generally to larger phrases because phrase and utterance edges are typically marked by functors (e.g., Morgan *et al.* 1987; Gervain *et al.*, 2008).

While we have suggested the rhythm of linguistic phrases as the driving factor in shaping nonlinguistic segmentation biases, we briefly consider the original proposal of Jakobson *et al.* (1952) and of Kusumoto and Moreton (1997) that the rhythm of words contributes to these biases. Initially it seems unlikely that common disyllabic words in English would be associated with a short-long duration pattern, since English is well known to have a bias for word-initial stress, and since duration is an important acoustic correlate of syllable stress in this language (Cutler and Carter, 1987; Delattre, 1966). However, the tendency for word-initial stress applies to nouns, not to verbs (such as "begin") or multisyllable function words (such as "about"), which are final stress and are highly represented (e.g., comprising 28% of all tokens of the 50 most common disyllabic words in the Brown corpus) (Francis and Kučera, 1982). The large representation of final-stress words would be expected to contribute many short-long elements to English. Similarly, there are reasons to expect that disyllabic words in Japanese would have a long-short temporal pattern. Kubozono (personal communication) has commented "Disyllabic words consisting of two moras (Light+Light) tend to turn into three mora words (Heavy+Light) by the lengthening of the first syllable. Moreover, disyllabic words consisting of four moras (H+H) tend to be shortened to H+L sequences as the final syllable becomes monomoraic. These two tendencies are observed in a wide range of linguistic phenomena in Japanese, from historical sound changes to baby words and word games" (see Kubozono, 2003, 2004). To establish the plausibility of word rhythms being a source of the perceptual biases, measurements of syllable duration patterns in infant-directed speech would need to be made.

Finally, this study has focused on language as a force in shaping basic rhythmic processing, but another important form of auditory experience is music. Might differences in the musics of the two cultures underlie the perceptual grouping differences observed in the current work? For example, if it were found that phrases in Western music typically start with a short-long pattern (e.g., a "pick-up note," as in the opening of the song "Greensleeves") and that phrases in Japanese music typically start with a long-short pattern, this could account for the perceptual differences in grouping preference. It has been noted that while pick-up notes often exist in Western music, they traditionally do not occur in Japanese music (Koizumi, 1984). A preliminary analysis of children's songs by our laboratory suggest that most musical phrases in both cultures start with notes of equal duration, suggesting that musical patterns cannot explain the biases seen in study 1 (see EPAPS). Further works analyzing larger corpora of music or quantifying the actual musical experience of infants are clearly called for to more fully test this possibility. An additional consideration is whether familiarity with western music notation might affect the grouping preferences expressed by adults.

B. Language development

The preceding discussion suggested a correlation between the perceptual grouping preferences of listeners and rhythmic patterns in their languages, which is consistent with an experience-based explanation of grouping preferences. However, the question of whether language experience actually causes the perceptual grouping differences awaits more direct study. Developmental studies could illuminate the timing of development of perceptual grouping, and whether it corresponds with language development milestones.

One important question is whether infants start life with inborn rhythmic grouping preferences, or whether these emerge purely as a function of linguistic experience. For example, do all infants begin with a short-long grouping preference, which is then reinforced by some languages and partially overridden by others? Or, do infants begin life with no rhythmic grouping bias and learn the bias while learning language? An answer to this question was recently provided by research on 5–8 month old English-immersed infants (living in Canada) by Yoshida *et al.* (2008). Using nonlinguistic tone sequences based on those in the current study, these researchers found that, like adults, 7–8 month olds prefer short long rhythmic groups (replicating Trainor and Adams, 2000). Interestingly, however, 5–6 month olds do not show a grouping preference. These results suggest that there is no innate bias, but that grouping preferences are learned sometime between 6 and 8 months. This is a period in which many language skills also emerge (including sensitivity to the order of function and content words, Gervain *et al.*, 2008), a clear suggestion that learning of linguistic factors (such as word order and word segmentation) and perceptual grouping interact. Further research will be needed to determine the precise timing and direction of causality, namely, whether grouping perception precedes and informs language processing, or may be a secondary consequence of it.

C. Linguistic and nonlinguistic rhythm in cross-cultural perspective

The syntactic proposal outlined above also unifies what may have been a puzzling result that supported claims for the universality of grouping perception: why native speakers of three distinct European languages (English, Dutch, and French) all show a similar preference in duration-based grouping of tone sequences, namely, a preference for short-long groups, despite these languages being rhythmically heterogeneous (Vos, 1977; Bell 1977; Hay and Diehl, 2007). Notably, these three languages span two traditional linguistic rhythm classes, with English and Dutch being "stress-timed" languages and French being a "syllable-timed" language. Related empirical research has revealed systematic differences in the temporal structure of English and Dutch on the one hand and French on the other, based on the temporal patterning of vowels and consonants in sentences (Ramus *et al.*, 1999; Grabe and Low, 2002; Lee and Todd, 2004). The key point for the current purposes, however, is that all of these languages have a level of linguistic structure at which they are similar, namely, in placing functors before content words

(e.g., the book, "le livre," and "het boek"). According to the current view, it is this syntactic similarity (ultimately expressed in terms of durational contrasts) that is responsible for the common preference for short-long groups across these cultures, via mechanisms suggested earlier.

If this view is correct, it should be possible to predict nonlinguistic grouping preferences in other cultures based on the syntactic structure of the native language. Since most European languages place functors before content words (Dryer, 2005), one can predict that native speakers of these languages will exhibit a preference for short-long grouping. Some European and many non-European languages, however (e.g., in Asia and India), resemble Japanese in placing functors after content words. Accordingly, one can predict that native speakers of these languages (e.g., Turkish, Korean, and Marathi) will prefer long-short grouping.

In making predictions about grouping perception based on native language, two additional factors must be kept in mind. First, the temporal structure of common words may also be important in shaping nonlinguistic grouping biases, as in the original proposal of Jakobson *et al.* (1952). The second factor that should be kept in mind concerns multilingualism. If implicit learning of linguistic duration patterns shapes nonlinguistic grouping, then exposure to or facility in a second language may have an influence on grouping preferences. In terms of the current study, the variability in grouping preferences seen among Japanese listeners may have reflected variation in the degree of proficiency or experience with English. Future cross-cultural work should quantify proficiency in non-native languages, to see if this can account for within-culture variability seen in grouping preferences. It also remains to be studied if grouping preferences are "locked in" during language development (as word segmentation strategies may be; Cutler, 2000) or can be modified by experience later in life.

D. Conclusion

The present work demonstrated that perception of basic grouping in nonlinguistic auditory sequences varies by culture, in contrast to long-held views about universal principles governing such grouping. The source of cultural variation in grouping perception is hypothesized to lie in experience of the rhythms of language, suggesting that learning of language may have consequences for the low-level rhythmic perception of sound.

ACKNOWLEDGMENTS

We thank N. Azechi, J. Fry, B. Hayes, H. Kubozono, E. Moreton, D. Roland, S. Roland, M. Sadakata, S. Shattuck-Huffnagel, N. Warner, and two anonymous reviewers. This work was supported by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where J.R.I. is the Karp Foundation Fellow and A.D.P. is the Esther J. Burnham Fellow. This work was additionally supported by Grants No. 14101001 and No. 19103003 from the Japan Society for the Promotion of Science.

[1]Indeed, the short-long bias of such phrases groups in English is a likely source of the prevalence of "iambic" (weak-strong) meters in English poetic verse (Gall, 1987).

Baker, M. C. (**2001**). *The Atoms of Language* (Basic Books, New York).
Bell, A. (**1977**). "Accent placement and perception of prominence in rhythmic structures," in *Studies in Stress and Accent*, edited by L. Hyman (UCLA Department of Linguistics, Los Angeles), pp. 1–13.
Bent, T., Bradlow, A. R., and Wright, B. A. (**2006**). "The influence of linguistic experience on the cognitive processing of pitch in speech and non-speech sounds," J. Exp. Psychol. Hum. Percept. Perform. **32**, 97–103.
Bolton, T. (**1894**). "Rhythm," Am. J. Psychol. **6**, 145–238.
Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
Christophe, A., Nespor, M., Guasti, M. T., and Van Ooyen, B. (**2003**). "Prosodic structure and syntactic acquisition: The case of the head-direction parameter," Dev. Sci. **6**, 211–220.
Cutler, A., and Carter, D. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.
Cutler, A. (**2000**). "Listening to a second language through the ears of a first," Interpreting **5**, 1–23.
Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," J. Phonetics **11**, 51–62.
Davis, M. H., and Johnsrude, I. S. (**2007**). "Hearing speech sounds: Top-down influences on the interface between audition and speech perception," Hear. Res. **229**, 132–147.
Delattre, P. (**1966**). "A comparison of syllable length conditioning among languages," IRAL **4**, 183–198.
Deutsch, D. (**1991**). "The tritone paradox: An influence of language on music perception," Music Percept. **8**, 335–347.
Drake, C., and Bertrand, D. (**2001**). "The quest for universals in temporal processing in music," Ann. N.Y. Acad. Sci. **930**, 17–27.
Dryer, M. S. (**2005**). "Relationship between the order of object and verb and the order of adposition and noun phrase," in *The World Atlas of Language Structures*, edited by M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie (Oxford University Press, New York), pp. 386–389.
See EPAPS Document No. E-JASMAN-124-045810 for sound examples and music analysis. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).
Fraisse, P. (**1982**). "Rhythm and tempo," in *The Psychology of Music*, edited by D. Deutsch (Academic, London), pp. 149–180.
Francis, W. N., and Kučera, H. (**1982**). *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton Mifflin, Boston).
Gall, S. M. (**1987**). "Versification," in *Poetry in English: An Anthology*, edited by M. L. Rosenthal (Oxford University Press, Toronto).
Gervain, J., Nespor, M., Mazuka, R., Horie, R., and Mehler, J. (**2008**). "Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study," Cogn. Psychol. **57**, 56-74.
Grabe, E., and Low, E. L. (**2002**). "Durational variability in speech and the rhythm class hypothesis," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 515–546.
Hannon, E. E., and Trehub, S. E. (**2005**). "Tuning in to musical rhythms: Infants learn more readily than adults," Proc. Natl. Acad. Sci. U.S.A. **102**, 12639–12643.
Hay, J. S. F., and Diehl, R. L. (**2007**). "Perception of rhythmic grouping: Testing the iambic/trochaic law," Percept. Psychophys. **69**, 113–122.
Hayes, B. (**1995**). *Metrical Stress Theory: Principles and Case Studies* (University of Chicago Press, Chicago).
Huron, D., and Ollen, J. (**2003**). "Agogic contrast in French and English themes: Further support for Patel and Daniele (2003)," Music Percept. **21**, 267–271.
Jakobson, R., Fant, G., and Halle, M. (**1952**). "Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates," Acoustics Laboratory, MIT, Technical Report No. 13.
Koizumi, F. (**1984**). *Research on Japanese Traditional Music 2 (Rhythm). [Nihon dento ongaku no kenkyu 2 (Rizumu)]*. (Ongaku no tomo sha, Tokyo).
Kubozono, H. (**2003**). "The syllable as a unit of prosodic organization in Japanese," in *The Syllable in Optimality Theory*, edited by C. Fery and R. van der Vijver (Cambridge University Press, Cambridge), pp. 99–122.
Kubozono, H. (**2004**). "Weight neutralization in Japanese," J. Japanese Ling. **20**, 51–70.
Kusumoto, K., Moreton, E. (**1997**). "Native language determines parsing of nonlinguistic rhythmic stimuli," J. Acoust. Soc. Am. **105**, 3204.

Lee, C. S., and Todd, N. P. McA., (**2004**). "Toward an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora," Cognition **93**, 225–254.
Lerdahl, F. and Jackendoff, R. (**1983**). "A generative theory of tonal music" (MIT Press, Cambridge, MA).
Mehler, J., Dupuox, E., Nazzi, T., and Dehaene-Lambertz, D. (**1996**). "Coping with linguistic diversity: The infant's viewpoint," in *Signal to Syntax*, edited by J. L. Morgan and D. Demuth (Lawrence Erlbaum, Mahwah), pp. 101–116.
Morgan, J. L., Meier, R. P., and Newport, E. L. (**1987**). "Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases in the acquisition of language," Cogn. Psychol. **19**, 498–550.
Nazzi, T., Bertoncini, J., and Mehler, J. (**1998**). "Language discrimination in newborns: Toward an understanding of the role of rhythm," J. Exp. Psychol. Hum. Percept. Perform. **24**, 756–777.
Nespor, M., and Vogel, I. (**1986**). *Prosodic Phonology*. (Foris, Dordrecht).
Nespor, M., Guasti, M. T., and Christophe, A. (**1996**). "Selecting word order: The rhythmic activation principle," in *Interfaces in Phonology*, edited by U. Kleinhenz (Akademie, Berlin), pp. 1–26.
Patel, A. D., and Daniele, J. R. (**2003**). "An empirical comparison of rhythm in language and music," Cognition **87**, B35–B45.
Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (**2006**). "Comparing the rhythm and melody of speech and music: The case of British English and French," J. Acoust. Soc. Am. **119**, 3034–3047.
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **73**, 265–292.
Selkirk, E. O. (**1984**). *Phonology and Syntax: The Relation Between Sound and Structure*. (MIT, Cambridge, MA).
Selkirk, E. (**1996**). "The prosodic structure of function words," in *Signal to syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, edited by J. L. Morgan and K. Demuth (Lawrence Erlbaum, Mahwah, NJ), pp. 187–213.
Shi, R., Morgan, J., and Allopenna, P. (**1998**). "Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective," J. Child Lang **25**, 169–201.
Shi, R., Werker, J. F., and Cutler, A. (**2006**). "Recognition and representation of function words in English-learning infants," Infancy **10**, 187–198.
Trainor, L. J., and Adams, B. (**2000**). "Infants″ and adults″ use of duration and intensity cues in the segmentation of tone patterns," Percept. Psychophys. **62**, 333–340.
Trehub, S. E., and Trainor, L. J. (**1993**). "Listening strategies in infancy: The roots of music and language development," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams and E. Bigand (Oxford University Press, Oxford) pp. 278–327.
Vos, P. (**1977**). "Temporal duration factors in the perception of auditory rhythmic patterns," Scientific Aesthetics/Sciences de l'Art **1**, 183–199.
Wertheimer, M. (**1938**). in *A Source Book of Gestalt Psychology*, edited by W. Ellis (Routledge and Kegan Paul, London), pp. 71–88.
Woodrow, H. (**1909**). "A quantitative study of rhythm: The effect of variations in intensity, rate and duration," Arch. Psychol. (Frankf) **14**, 1–66.
Yoshida, K. A., Iversen, J. R., Patel, A. D., and Werker, J. F. (**2008**). "Development of Abstract Grammatical Representation," *26th Biennial International Conference on Infant Studies*, Vancouver, BC, Canada.

Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region

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 30 August 2007; revised 10 July 2008; accepted 18 July 2008)

In cochlear implants, or vocoder simulations of cochlear implants, the transmission of envelope cues at high rates (related to voice fundamental frequency, $f0$) may be limited by the widths of the filters used to form the channels and/or by the cutoff frequency, f_{lp}, of the low-pass filters used for envelope extraction. The effect of varying f_{lp} in tone and noise vocoders was investigated for channel numbers, N, from 6 to 18. As N increased, the widths of the channels decreased. The value of f_{lp} was 45 Hz (envelope or "E" filter), or 180 Hz (pitch or "P" filter). The following combinations of cutoff frequencies were used for channels below and above 1500 Hz, respectively: EE, PE, EP, and PP. Results from a competing-talker task showed that the tone vocoder led to better intelligibility than the noise vocoder. The PP condition led to the best intelligibility and the EE condition to the worst. For $N=6$, intelligibility was better for condition PE than for condition EP. For $N=18$, the reverse was true. The results indicate that the channel bandwidths can compromise the transmission of $f0$-related envelope information, and suggest that vocoder simulations of cochlear-implant processing have limitations. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2968678]

PACS number(s): 43.66.Ts, 43.66.Mk, 43.71.Gv [KWG] Pages: 2272–2282

I. INTRODUCTION

The vocoder (Dudley, 1939) has been used widely to simulate the effects of cochlear-implant (CI) processing (e.g., Van Tasell *et al.*, 1987; Shannon *et al.*, 1995). To vocode an audio signal, it is first split into N frequency channels. The passbands of the channels are usually contiguous, although there is often overlap of adjacent channels on the skirts. The number N determines the degree of spectral detail that is preserved, and it can be chosen to be small to simulate the limited effective spectral resolution associated with CIs (Shannon, 1983; Friesen *et al.*, 2001). The envelope of each channel signal is extracted and used to modulate either a noise carrier of the same bandwidth as the channel (noise vocoder) or a sinusoidal carrier at the center frequency of the channel (tone vocoder). The use of a carrier independent from the original signal simulates loss of information about the temporal fine structure of the signal, which is another characteristic of CI signal reception (Moore, 2003a; Moore and Carlyon, 2005).

The choice of the most appropriate carrier has been a matter of debate. A noise band has inherent low-rate random envelope modulation that is independent of the modulation derived from the input signal. The envelope modulation from a band of filtered Gaussian white noise has a triangular power spectral density, with power increasing with decreasing modulation rate (Lawson and Uhlenbeck, 1950). One can think of the envelope modulation as arising from beats between the different frequency components in the filtered noise. There are more closely spaced components than widely spaced components, so there is more modulation power at low modulation frequencies. This spurious low-rate modulation may interfere with or mask the signal modulation (Dau *et al.*, 1997; Whitmal *et al.*, 2007). A sinusoidal carrier does not introduce spurious modulation, but, especially for low N, produces an artificial structure of the vocoded signal that may be distracting. Also, if N is small and the bandwidths of the channels are correspondingly large, spectral sidebands falling within an output channel may be resolved by normally hearing listeners. For a voiced speech sound, these spectral sidebands are separated from the carrier frequency by ± the fundamental frequency ($f0$), and may provide a cue to $f0$ (de Boer, 1956; Gonzales and Oliver, 2005; Whitmal *et al.*, 2007). Such spectral cues would not be available to a person listening via a CI.

Dorman *et al.* (1997) compared the performance of normally hearing listeners using noise and tone vocoders with N varying from 2 to 9, and found that, for speech presented in quiet, the noise vocoder led to slightly poorer performance than the tone vocoder for vowel sounds (a 5%–10% difference), but there was no clear effect of vocoder type for consonants or sentences. After the present work was completed, Whitmal *et al.* (2007) reported a study comparing speech intelligibility for noise and tone vocoders under quiet conditions and when the speech was presented in speech-shaped noise or two-talker babble. They found that, in all conditions, the tone vocoder led to intelligibility about 15% better than for the noise vocoder for sentences and VCV nonsense syllables. They concluded that the poorer performance with the noise vocoder occurred because the intrinsic fluctuations of the carrier interfered with processing of the temporal fluctuations conveying information about the speech. One purpose of the present study was to provide information about the

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

TABLE I. Parameters of the vocoder processing used in several studies. The slope, S, of the low-pass filter applied to the envelope is specified in dB/octave. The cutoff frequency of the low-pass filter is designated f_{lp}. The carrier type is indicated by n (noise), t (tone), and hc (harmonic complex tone). BW denotes bandwidth.

Study	S	f_{lp} (Hz)	Carrier	N
Lorenzi *et al.* (2000)	6	500	n	4
Füllgrabe *et al.* (2006)	6	500	n	4
Fu *et al.* (1998b)	6	50, 500	n	1, 2, 3, 4
Kong and Zeng (2006)	6	50, 500	n	1, 8
Shannon *et al.* (1995)	6	15, 50, 160, 500	n	1, 2, 3, 4
Stickney *et al.* (2004)	12	500	n	2, 4, 8
Qin and Oxenham (2003)	12	300	n	4, 8, 24
Dorman *et al.* (1997)	12	160	n	2, 3, 4, 5, 6, 7, 8, 9
Xu *et al.* (2002)	12	1–512 (octave steps)	n	1, 2, 3, 4, 6, 8, 12
Xu *et al.* (2005)	12	1–512 (octave steps)	n	1, 2, 3, 4, 6, 8, 12, 16
Whitmal *et al.* (2007)	12	300 or 0.5×channel BW	n	6
Shannon *et al.* (2002)	18	220	n	20
Van Tasell *et al.* (1992)	18	20, 150	n	1
Luo and Fu (2004)	24	500	n	4
Baskent (2006)	24	160	n	2, 4, 6, 8, 10, 12, 16, 24, 40
Stone and More (2003)	40	50	n	4, 6, 8, 11, 16
Kwon and Turner (2001)	40	4, 16	n	12
Apoux and Bacon (2004)	48	500	n	4
Fu *et al.* (1998a)	48	500	n	3, 4, 8, 16
Faulkner *et al.* (2000)	48	32, 400	n	4
Van Tasell *et al.* (1987)	48	20, 200, 2000	n	1
Fu and Shannon (2000)	48	2, 5, 10, 20, 40, 160, 640	n	4
Dorman *et al.* (1998)	12	400	t	6, 8, 12, 16, 20
Loizou *et al.* (1999)	12	400	t	2, 3, 4, 5, 6, 8, 10, 12, 16
Loizou *et al.* (2000)	12	400	t	4, 6, 8, 12, 16
Dorman *et al.* (1997)	12	400	t	2, 3, 4, 5, 6, 7, 8, 9
Dorman *et al.* (2002)	12	400	t	4, 6, 8, 10, 12, 16, 20
Whitmal *et al.* (2007)	12	300 or 0.5×channel BW	t	6
Deeks and Carlyon (2004)	24	30	hc	3, 6

relative performance obtained using noise and tone vocoders under conditions where the background is a single competing talker. Reasons why this case is especially interesting are presented later.

Another issue is connected with the range of envelope modulation rates transmitted by a vocoder. Rosen (1992) defined three ranges of modulation rates to characterize the information in speech:

(1) Envelope, defined as "fluctuations in overall amplitude at rates between about 2 and 50 Hz." This range is primarily related to articulator movement.
(2) Periodicity, defined as "rates between about 50 and 500 Hz." This range is related to the rate of vocal-fold vibration, which determines the $f0$ and pitch.
(3) Fine structure, defined as "rates from about 600 Hz to about 10 kHz." This range is related either to the harmonic structure produced by vocal-fold vibration or to the noiselike sound produced by turbulent airflow at constrictions in the vocal tract.

It is often assumed that cues from range (1) are most important for speech intelligibility (Plomp, 1983; Drullman *et al.*, 1994a,b; Shannon *et al.*, 1995). However, cues from range (2) may be important when the listener is trying to understand a target talker in the presence of one of more background talkers (Brokx and Nooteboom, 1982; Qin and Oxenham, 2003, 2005, 2006). The highest envelope modulation rate transmitted via a given channel in a vocoder or a CI depends partly on the bandwidth of that channel. The passbands of the channels used in vocoder simulations and in CI processing are usually minimally overlapping. Thus, as N is increased, the individual channels become narrower. Narrow channels attenuate the sidebands produced by high-rate envelope modulation, and this limits the highest modulation rate that is transmitted by a given channel. To allow modulation at rates corresponding to $f0$ to be passed at full depth, the channel width must be at least $2f0$. Table I gives values of the parameters of vocoder processing for a large number of studies. The value of N covers a wide range, from 1 to 40. For values of N larger than about 10, the transmission of envelope cues related to $f0$ would certainly have been partially limited by the widths of the channels. It should be noted that some earlier studies used a very small number of channels ($N=1$ or 2) to investigate the contribution of envelope and $f0$ cues to speech-reading (lip-reading) (Erber, 1969; Rosen *et al.*, 1981; Grant *et al.*, 1991; Breeuwer and Plomp, 1986). However, the focus here is on studies that included larger values of N, sufficient to support speech understanding without speech-reading.

The characteristics of the low-pass filtering (if any) used to extract the envelope also have a strong influence on the highest modulation rate that is transmitted by a given channel. Both the low-pass filter cutoff frequency, f_{lp}, and the filter slope outside the passband, S (dB/octave), play a role. As shown in Table I, the value of f_{lp} varies from 1 to 2000 Hz, while the value of S varies from 6 to 48 dB/octave. The idea that speech intelligibility depends primarily on low-rate modulations in different frequency channels has come partly from studies of the intelligibility of vocoded speech, as summarized in Table I. Intelligibility remained relatively high when f_{lp} was set to a low value, 50 Hz or lower. However, in many such studies, the value of S was low, either 6 or 12 dB/octave, so the modulation at high rates was gradually attenuated, but was not removed entirely by the low-pass filtering of the envelope.

In the studies summarized in Table I, the value of S was fixed within each study. Thus, the studies give little direct information about the influence of S on performance. This issue was addressed in a recent study of Healy and Steinbach (2007). In their first experiment, they used a three-channel tone vocoder to code speech in quiet, and showed that intelligibility decreased monotonically as the value of S was increased from 6 to 192 dB/octave. They argued that this effect depended on the transmission of modulation with rates falling within the skirt of the low-pass filter, which they called the transition band. Also, intelligibility was much higher for $f_{lp}=100$ Hz than for $f_{lp}=16$ Hz. In a second experiment, they used three- or four-channel noise vocoders. The envelope signal in one of the channels was low-pass filtered at 400 Hz, while the remaining envelope signals were low-pass filtered at 16 Hz, both with $S=48$ dB/octave. The benefit of the higher value of f_{lp} was greatest when the high-rate information was supplied to the highest spectral channel. However, the best performance was obtained when f_{lp} was 400 Hz for all channels. Healy and Steinbach concluded that the use of higher-rate envelope cues was not restricted to high spectral regions.

The main goal of the present study was to examine the effect of providing higher-rate envelope information under conditions when a background talker was present, and also for higher values of N. For $N=1$ and a noise carrier, Van Tasell *et al.* (1987) reported that consonant perception improved when f_{lp} increased from 20 to 200 Hz. For $N=1$ and a tone carrier, Grant *et al.* (1991) reported that sentence intelligibility improved as f_{lp} increased from 4.2 to 67 Hz, but plateaued for higher values of f_{lp} (using the same definition of f_{lp} as Van Tasell *et al.*). Both of these studies used speech in quiet, a necessity given the low value of N. Here, for a speech-in-interference task, it was anticipated that higher-rate envelope information would provide cues for $f0$ and that this might be especially important to allow the target talker to be segregated from the background talker. Also, it was anticipated that a tone vocoder might convey $f0$ cues more effectively than a noise vocoder, so that the difference between vocoder types would be especially apparent when high-rate envelope cues were available. The classification scheme of Rosen (1992), described earlier, was used to guide the design of two low-pass filters: an envelope ("E") filter which primarily passed modulation rates in range (1) (up to 45 Hz), and a pitch ("P") filter which passed modulation rates in ranges (1) and (2) (up to 180 Hz). The P filter passed most of the range of $f0$ values of the talkers used in this study. The effect of providing the higher-rate information in all channels, only in the higher-frequency channels (above 1500 Hz), or only in the lower-frequency channels (below 1500 Hz) was explored. This was done to assess which channels were most effective in conveying $f0$ information, and to assess the benefits of providing $f0$ information over a wide frequency range.

II. EXPERIMENT 1: EFFECTS OF N AND f_{lp} FOR NOISE AND TONE VOCODERS

A. Vocoder design and processing condition

The analysis stage of the vocoder used finite impulse response bandpass filters equally spaced on the ERB_N-number scale (Glasberg and Moore, 1990; Moore, 2003b), using a variable-order design. The passbands of the filters covered the frequency range from 100 to 7800 Hz. The transition bands of each filter were designed to have similar slopes when plotted on a logarithmic frequency scale. The response of each filter was −6 dB at the point where the response crossed that of the adjacent filter. The crossover with the next-but-one filter was below the −50 dB point. The stop band response was below −70 dB. N was equal to 6, 11, or 18. The envelope was extracted from each channel signal by half-wave rectification and low-pass filtering; details of the low-pass filters are given later. The envelope was then used to modulate the channel carrier. As described in Sec. I, two different carrier signals were used. For the noise vocoder, the envelope signal from each channel was used to modulate a band of white noise and the resulting modulated noise was bandpass filtered with the filter appropriate for each channel. The filtered channel signals were then summed. For the tone vocoder, the noise was replaced by a sinusoidal carrier at the center of each channel when expressed on a linear frequency scale. Again, the channel signals were bandpass filtered before being summed. The filters matched those used for the analysis stage. All delays introduced by the different stages of filtering were removed. This preserved the time alignment of envelope features across frequency.

Frequency-response plots for the E and P filters are shown as labeled solid lines in Fig. 1. As Healy and Steinbach (2007) pointed out, a high value of S (above about 24 dB/octave) may lead to undesirable properties in the time domain, namely undershoot, which could cause the modulation index to exceed unity. The value of S used here was chosen to be about 24 dB/octave to minimize such effects. The response of the E filter was −3 dB at a corner frequency of 45 Hz and was negligible above 100 Hz. The P filter had a corner frequency of 180 Hz, and a negligible response above 400 Hz. The dashed lines in Fig. 1 show stylized responses of the envelope low-pass filters used by Shannon *et al.* (1995); see Table I. Based on the results of Healy and Steinbach (2007), it seems likely that the transition bands in

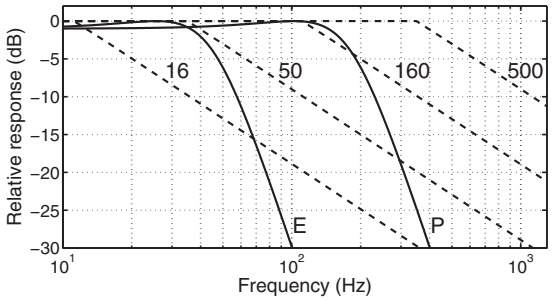


FIG. 1. Frequency response plots for the E and P envelope low-pass filters are shown as solid lines. Dashed lines show stylized responses of the envelope-extraction filters used by Shannon *et al.* (1995), with f_{lp}=16, 50, 160, and 500 Hz, and S=6 dB/octave.

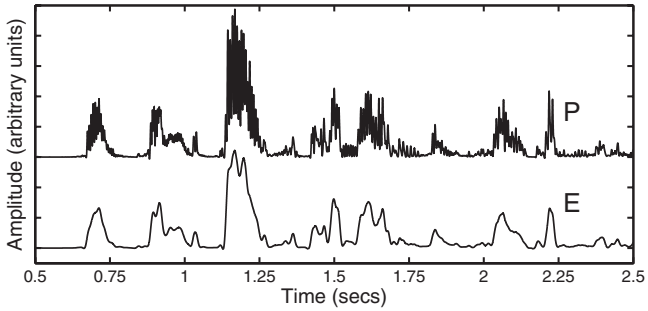


FIG. 2. Example of envelopes at the output of channel six of the 11-channel vocoder (centered at 1413 Hz) following application of the P filter (top) and E filter (bottom). The input signal was a mixture of two talkers at a +6 dB target-to-background ratio.

the study of Shannon *et al.* would have provided considerable information, and some *f*0 information would have been available when f_{lp} was 50 Hz. Of course, the extent to which such information is used depends on the ability of the listeners to detect the envelope modulation and discriminate changes in modulation rate (Grant *et al.*, 1998).

Figure 2 shows the envelopes at the output of a single channel centered at 1413 Hz for the 11-channel vocoder, after low-pass filtering using the P filter (top) and the E filter (bottom). The input signal was a mixture of two talkers, as used in the experiment (see below for details). When the P filter was used, considerable modulation at the *f*0 rate was preserved. The modulation depth was typically above 50%, which is above the detection threshold for amplitude modulation with rates up to 200 Hz, at least for sinusoidal carriers (Kohlrausch *et al.*, 2000). The overall shape of the envelope is similar for the P and E filters. While the envelope rises slightly more sharply at onsets for the P filter, the onsets still appear to be quite well defined for the E filter.

In this study, four combinations of low-pass filters were used:

(1) Condition EE: The E filter was used for all channels, preserving envelope modulation only for rates below 45 Hz, and with reduced representation of rates up to about 100 Hz.
(2) Condition PP: The P filter was used for all channels, preserving envelope modulation for rates below 180 Hz, and with reduced preservation of rates up to about 400 Hz.
(3) Condition EP: The E filter was used for channels centered below 1500 Hz, while the P filter was used for channels centered above 1500 Hz. As a consequence, envelope cues related to *f*0 were available only in channels centered above 1500 Hz.
(4) Condition PE: The P filter was used for channels centered below 1500 Hz, while the E filter was used for channels centered above 1500 Hz. As a consequence, envelope cues related to *f*0 were available only in channels centered below 1500 Hz. However, these cues were partially limited by the bandpass filtering used to form the channels when N was large (see Table II).

TABLE II. Channel edge frequencies and bandwidths used in experiments 1 and 2. A, B, and C are for N=6, 11, and 18 respectively. Bandwidths shown in bold are for those channels where the center frequency was less than 1500 Hz. The line starting with "Limited" shows which channel bandwidths lead to partial removal of the modulations capable of being passed by the E or the P envelope low-pass filters (less than 45 or 180 Hz, respectively), or both.

A: N = 6																			
Edge frequency (Hz)	100			331			725			1396			2538			4485			7800
Bandwidth (Hz)		**231**			**394**			**671**			1143			1946			3315		
Limited		**P**																	
B: N = 11																			
Edge frequency (Hz)	100	211		359	557		822		1176		1650	2283		3130		4262		5776	7800
Bandwidth (Hz)		**111**	**148**		**198**	**265**		**354**	**474**		633	847		1132		1514		2024	
Limited		**P**	**P**		**P**	**P**		**P**											
C: N = 18																			
Edge frequency (Hz)	100	164	240	331	440	570	725	910	1132	1396	1712	2089	2539	3076	3718	4485	5401	6494	7800
Bandwidth (Hz)		**64**	**76**	**91**	**109**	**130**	**155**	**185**	**222**	**264**	316	377	450	537	642	767	916	1093	1306
Limited		**EP**	**EP**	**EP**	**P**	**P**	**P**	**P**	**P**	**P**	P								

The choice of 1500 Hz as the "split" frequency was based on calculations using the Speech Intelligibility Index (SII) procedure (ANSI, 1997), which showed that, for (uncoded) clean speech with a moderate level, the contribution to intelligibility was equal for spectral components above and below this frequency (this was based on the importance function for "average speech," Table I in the standard). Due to the quantizing effect of the channel spacing, the calculated SIIs for the channels centered below the split were 0.47, 0.55, and 0.47 for $N=6$, 11, and 18, respectively.

Table II shows the channel edge frequencies and bandwidths for each value of N. The first line of each part shows the edge frequencies. The second line of each part shows in bold the bandwidths of channels whose center frequencies were below 1500 Hz, the dividing point for conditions EP and PE. The third line of each part shows whether the channel bandwidth led to limitation of the modulation rate over part of the range passed by the E or P filter. Note that, for $N=18$, the range of modulation rates passed by the channels with the three lowest center frequencies was less than the range passed by the E filter. Thus, even relatively low-rate envelope modulations were partially removed as a result of the small channel bandwidths at low frequencies.

B. Speech material

The target (male) talker was a native speaker of British English producing sentences from the IEEE corpus (IEEE, 1969). The recording was made at University College London; see Smith and Faulkner (2006) for details. The $f0$ of his voice ranged from 75 to 150 Hz. The speech used as the background was recorded from a male speaker of British English reading naturally from a script in a large sound-isolated and sound-treated room with low reverberation. The recording was made direct to digital audio tape using a high-quality, low-noise condenser microphone. Using CoolEdit2000™, stammerings, repetitions, and pauses for breath were removed, but natural-sounding pauses between sentences with durations of 100–200 ms were left. The segment of background speech to be used with a given sentence was selected randomly from the continuous speech file. The background started 1000 ms before the target sentence and continued for approximately 1000 ms after the sentence. The background talker produced speech with a very dynamic $f0$, ranging between 100 and 200 Hz. The overlapping range but different means of $f0$ were chosen to ensure that $f0$, where available in the processed signals, would be a strong cue to segregation of the two voices (Brokx and Nooteboom, 1982).

The target-to-background ratio (TBR) was +6 dB for all N. This value was chosen so as to avoid floor and ceiling effects, as far as possible. At this ratio, the listeners were easily able to determine when each of the target sentences was presented. The target and background were mixed prior to vocoder processing.

C. Subjects, training, and experimental design

Two groups of 12 subjects, all university undergraduates or graduates, were selected on the basis of their having audiometric thresholds $\leqslant$15 dB HL at octave frequencies between 125 and 8000 Hz and at 3000 and 6000 Hz. All were fluent speakers of British English and were paid for their attendance. Group NV (three male, nine female, mean age =23.3 years, standard deviation (s.d.)=3.4 years, range =20–30 years) was tested using the noise vocoder and group TV (three male, nine female, mean age=23.6 years, s.d. =5.4 years, range=19–35 years) was tested using the tone vocoder. All subjects attended two sessions, each of which lasted about 1.25 h. The first session was spent entirely on training with stimuli whose difficulty was progressively increased by reducing N and by reducing the TBR. The initial training involved listening to continuous prose passages, without any background sound. The first passage was presented without any processing (i.e., nonvocoded). Then the same passage was presented again. Initially the speech was uncoded, but after 5 s it was cross-faded to vocoded speech, using 18 channels (with the carrier appropriate for the group). Then, the same passage was presented for a third time, but it was entirely vocoded, using 11 channels. Then this whole stage of training was repeated, using a new prose passage, except that the vocoder used 11 channels when first introduced, and six channels in the final stage. After this initial training, subjects listened to approximately 170 vocoded sentences, presented with a background talker. They were asked to identify the target sentence, and could listen to each sentence twice before making a decision. If they were incorrect, the experimenter spoke the correct sentence (via a microphone linked to the headphones of the subject), and the sentence was presented one more time. The second session started with more training for 20 min (using approximately 100 sentences, and again using feedback) and then data collection began.

The 12 processing schemes (three numbers of channels by four envelope-filter conditions) were tested in a counterbalanced design. Within each group, each processing scheme was assessed using 20 sentences (each containing five key words), giving a maximum score of 100 key words correct. The same set of sentence lists was used for each group.

D. Signal presentation

Preprocessed material was replayed through a LynxONE soundcard hosted in a PC, under control of MATLAB. The soundcard output signal was buffered by a Mackie 1202-VLZ PRO mixing desk. Presentation was diotic via Sennheiser HD580 headphones. Subjects were seated in a double-walled, sound-attenuating booth. The overall level of the target plus background was 68 dB SPL for both groups.

E. Results

The scores were transformed into rationalized arcsine units (RAU; Studebaker, 1985) and averaged in this domain. The mean scores, converted back to percentages, are plotted in Fig. 3. The left- and right-hand panels show scores for the noise and tone vocoders, respectively.

To assess the effect of vocoder type, an analysis of variance (ANOVA) was conducted on the RAU-transformed scores with N and envelope-filter condition as within-subject factors and vocoder type as a between-subject factor. The

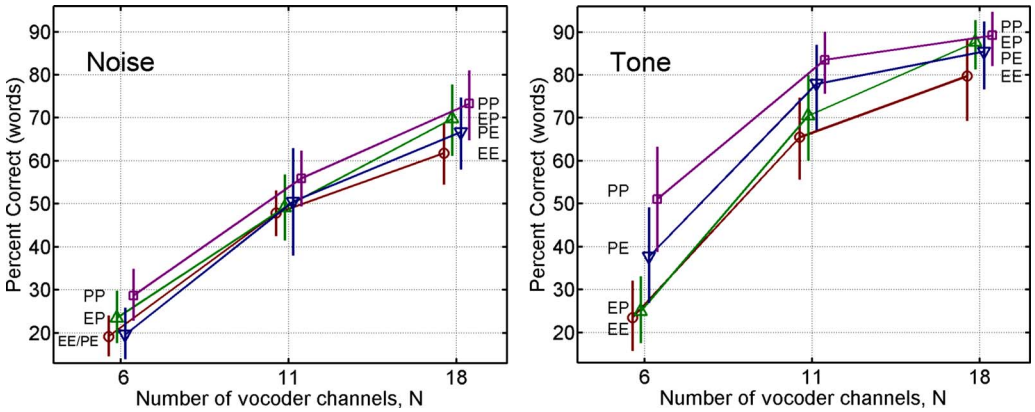


FIG. 3. (Color online) Results of experiment 1. Mean identification scores are plotted as a function of N for the noise vocoder (left) and the tone vocoder (right). PP, EP, PE, and EE denote different envelope low-pass filter conditions, as described in the text. Points for a given N are slightly offset for clarity. Error bars denote ±1 s.d. across subjects. Standard errors would be a factor of 3.5 smaller.

main effect of vocoder type was significant: $F(1,22)=70.9$, $p<0.001$; intelligibility was higher for the tone than for the noise vocoder. There was also a significant interaction of N, filter condition, and vocoder type [$F(6,242)=2.88$, $p=0.01$]. This can be interpreted as reflecting the fact that the tone vocoder gave higher intelligibility than the noise vocoder for all conditions except EE and EP with $N=6$. All other main effects and interactions were also significant, but they will not be discussed further, as the relevant effects are covered by the additional analyses presented later.

For further analyses, within-subject ANOVAs were conducted on the RAU-transformed scores, separately for the two vocoder types, with factors N and envelope-filter condition. The outcomes are shown in Table III. Post hoc comparisons were based on Fisher's least significant difference (LSD) test (Keppel, 1991). The main effects were significant for both vocoder types. As expected, for both carrier types, intelligibility increased as N increased.

For the noise vocoder, there was no significant interaction between N and filter condition, so only results averaged across N are considered. Performance for condition PP was significantly better than for all other conditions, [$t(33)\geq 3.65$, $p<0.001$, two-tailed]. Thus, adding high-rate envelope information was clearly beneficial when it was added to all channels, as found by Healy and Steinbach (2007). Performance did not differ significantly for conditions EP and PE [$t(33)=1.03$, $p>0.05$]. Performance for condition EE was significantly poorer than for condition EP [$t(33)=3.36$, $p<0.01$], but not than for condition PE [$t(33)=1.86$, $p>0.05$]. Thus, when high-rate envelope information was added over a limited spectral range, it produced a significant, but small, benefit only at high spectral frequencies.

As was found for the noise vocoder, performance for the tone vocoder in condition PP was significantly better than for all other conditions [$t(33)\geq 3.56$, $p<0.001$]. However, in contrast to results for the noise vocoder, performance for condition PE was significantly better than for conditions EP and EE [$t(33)\geq 2.59$, $p<0.05$]. Finally, performance for condition EP was significantly better than for condition EE [$t(33)=2.42$, $p<0.05$]. There was a significant interaction between N and filter condition: $F(6,66)=6.12$, $p<0.001$. For $N=6$, post hoc tests showed the following significant differences: PP > PE > (EP and EE) [$t(89.3)\geq 3.41$, $p<0.001$]. A similar pattern of results was found for $N=11$; PP > PE > (EP and EE) [$t(89.3)\geq 2.0$, $p<0.05$]. In other words, the beneficial effect of adding high-rate envelope information was greatest when it was added in all channels, and was greater when it was added to channels centered below 1500 Hz than when it was added to channels centered above 1500 Hz. Possible reasons for this are discussed later: see Sec. IV A. For $N=18$, a different pattern of results was obtained: (PP and EP and PE) > EE [$t(33)>2.0$, $p<0.05$]. In other words, there was a beneficial effect of adding high-rate envelope information, but the benefit did not depend significantly on whether the information was added to all channels, or only to channels tuned to high or low frequencies.

In summary, adding high-rate envelope information in all channels produced a significant benefit for both the noise and the tone vocoders, the benefit generally being greater for the latter. For the tone vocoder, the pattern of results varied with N. For $N=6$ and 11, performance was better for condition PE than for condition EP, indicating a greater benefit of adding high-rate envelope information to channels with center frequencies below 1500 Hz. For $N=18$, there was no significant difference between scores for conditions PE and EP, but this may have happened because of a ceiling effect.

TABLE III. Summary of ANOVA outcomes for experiment 1.

Factor	Noise vocoder	Tone vocoder
N	$F(2,22)=310.1, p<0.001$	$F(2,22)=566.0, p<0.001$
Filter condition	$F(3,33)=17.8, p<0.001$	$F(3,33)=2.69, p<0.001$
Filter condition × N	$F(6,66)=1.04, p>0.05$	$F(6,66)=6.12, p<0.001$

III. EXPERIMENT 2: AVOIDING FLOOR AND CEILING EFFECTS

A. Rationale

In experiment 1, results for the tone vocoder showed a significant interaction between N and envelope-filter condition. This probably happened because, for small N (6 and 11), cues related to $f0$ were conveyed more effectively by

channels tuned to low frequencies than by channels tuned to high frequencies, while for large N (18) the reverse was true. However, there was not a significant difference between conditions PE and EP for $N=18$, possibly because of a ceiling effect. Also, for the tone vocoder with $N=6$, scores were low, especially for conditions EP and EE. A possible explanation for the low score in condition EP is that the TBR may have been too low to allow effective use of envelope cues relating to $f0$.

In experiment 2, TBRs were chosen to make performance more similar across N and to avoid floor and ceiling effects. For $N=18$, the TBR was reduced from 6 to 4 dB. For $N=11$, the TBR was kept at 6 dB to permit comparison of the results with those for experiment 1. For $N=6$, two new values of the TBR were used, 8 and 10 dB, compared to the earlier value of 6 dB. Only the tone vocoder was used, since in experiment 1 a significant interaction between N and envelope-filter condition was found only for this type of vocoder.

B. Subjects and method

All 12 subjects from group TV of experiment 1 took part in this experiment. Since 16 conditions were to be assessed, four further subjects were recruited on a basis identical to that for experiment 1, so as to allow counterbalancing of the order of testing conditions across subjects. The four new subjects were given an identical amount of training and testing as the rest of the group before participating in this experiment. The characteristics of the 16 subjects were: four male, 12 female, mean age=23.5 years, s.d.=5.0 years, range=19–35 years. The IEEE sentence material was again used, but with a different set of lists, to which the subjects had had no prior exposure. All other details of the stimuli and procedure were the same as for experiment 1.

C. Results

The scores were transformed into RAU for statistical analysis. The mean results, converted back into percentages, are shown in Fig. 4. The rank ordering of the scores across filter conditions for each N is the same as found for the tone vocoder in experiment 1 (right-hand panel of Fig. 3). A within-subject ANOVA was conducted on the RAU-transformed scores with factors filter condition and combination (of N and TBR). The main effect of filter condition was significant: $F(3,45)=73.28$, $p<0.001$. All pairwise comparisons of filter condition were significant at $p<0.05$. Performance was always best for condition PP, and always worst for condition EE. The interaction of filter condition and combination was also significant: $F(9,135)=1.98$, $p<0.05$. For ($N=6$, TBR=8), ($N=6$, TBR=10), and ($N=11$, TBR=6), performance was better for condition PE than for condition EP, but for ($N=18$, TBR=4) the reverse was true.

D. Combined analysis of results for experiments 1 and 2

To determine the effect of changing the TBR, and to increase the power of the analysis in revealing differences between filter conditions, the results of experiment 2 were pooled with those of experiment 1 for the tone vocoder. The two experiments were very similar in design, but used (a) different group sizes, (b) a slightly different group of subjects, and (c) nonoverlapping lists from the same speech corpus. To check whether these factors influenced the results, the data were compared from the processing conditions common to the two experiments, namely those for $N=11$ with TBR=6 dB. The mean scores (in RAU) for conditions EE, EP, PE, and PP, respectively, were 64.6, 69.6, 77.5, and 84.1 for experiment 1, and 63.8, 68.1, 75.1, and 87.4 for experiment 2. As the scores were very similar for the two experiments, it seemed reasonable to conduct a two-factor ANOVA on the scores, with one within-subject factor (filter condition), and one between-subject factor (experiment number). There was no significant effect of experiment number: $F(1,26)=0.02$, $p>0.05$. There was a significant effect of processing condition [$F(3,78)=42.7$, $p<0.001$], of which more detail will be given later. Importantly, the interaction between experiment number and processing condition was not significant: $F(3,78)=0.72$, $p>0.05$. This means that the pattern of results was not significantly different across the two experiments, at least for $N=11$.

The results for $N=6$ and 18 were then combined across the two experiments, and the means (averaged in RAU and converted back from RAU to percent), are plotted in Fig. 5. Separate ANOVAs were performed on the data sets for $N=6$ and 18, with factors TBR and filter condition. The summary statistics from the ANOVAs are shown in Table IV. For both values of N, there were significant effects of TBR and filter condition, but there was no significant interaction between the two factors. Intelligibility consistently increased with increasing TBR, as expected.

For $N=6$, all pair-wise comparisons of filter conditions gave a significant effect: PP>PE>EP>EE [$t(123)=3.37$, $p<0.001$; $t(123)=3.37$, $p<0.001$; $t(123)=1.98$, $p<0.05$, respectively]. Performance was consistently better for condition PE than for condition EP, i.e., the high-rate envelope

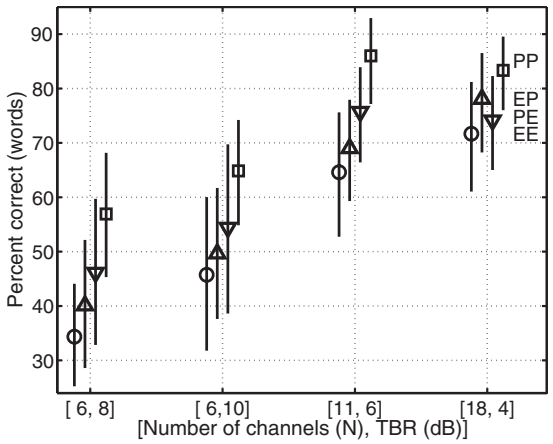


FIG. 4. Mean scores for each condition of experiment 2. Points for the different envelope filtering conditions are slightly offset for clarity. The abscissa shows the combination of N and target-to-background ratio (TBR). The abscissa order was selected based on the rank order of the mean scores. PP, EP, PE, and EE denote different envelope low-pass filter conditions, as described in the text. Error bars denote ±1 s.d. across subjects. Standard errors would be a factor of 4 smaller.

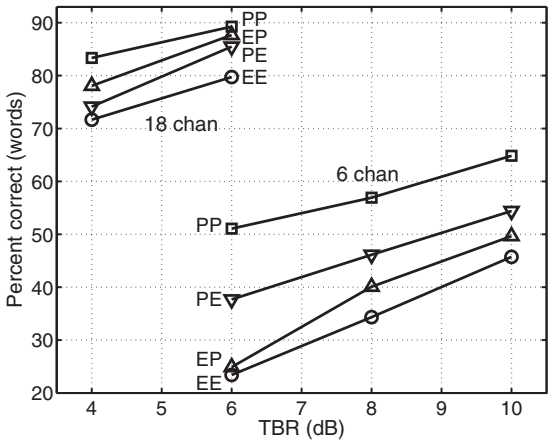


FIG. 5. Combined results for $N=6$ and 18 using the tone vocoder, as obtained in experiments 1 and 2, plotted as a function of TBR. To avoid clutter, error bars are not shown; error bars for all conditions can be seen in Figs. 3 and 4.

information was of more benefit when provided in low-frequency channels than when provided in high-frequency channels. The pattern of results for $N=11$ (Figs. 3 and 4) resembled that for $N=6$: PP > PE > EP > EE [$t(78)=3.46$, $p<0.001$; $t(78)=3.46$, $p<0.001$; $t(78)=2.0$, $p<0.05$, respectively]. For $N=18$, all pair-wise comparisons of filter conditions gave a significant effect: PP > EP > PE > EE [$t(78)>2.0$, $p<0.05$]. In this case, performance was consistently better for condition EP than for condition PE, i.e., the high-rate envelope information was of more benefit when provided in high-frequency channels than when provided in low-frequency channels.

IV. DISCUSSION

A. Influence of *N* on the benefit from high-rate envelope modulation information

Our results show that envelope information at rates above 45 Hz (called *f*0-related information) contributes to intelligibility in a competing-speech task, for *N* ranging from 6 to 18. This extends the results of Healy and Steinbach (2007) for speech in quiet with $N=3$ or 4. The effect found by Healy and Steinbach for speech in quiet presumably occurred mainly because the high-rate information provided a cue to indicate the presence or absence of voicing. The effect in our experiments may depend partly on the same factor, and partly on the contribution of *f*0-related information to the separation of the target and background voices. Our results also show that *f*0-related information is used most effectively when it is present in all channels; scores were consistently higher for condition PP than for any other condition, for both the noise and the tone vocoder. Healy and Steinbach (2007) reported a similar result for speech in quiet.

TABLE IV. Summary of ANOVA outcomes for the combined data of experiments 1 and 2.

Factor	$N=6$	$N=18$
Filter condition	$F(3,123)=52.8, p<0.001$	$F(3,78)=17.0, p<0.001$
TBR	$F(2,41)=16.9, p<0.001$	$F(1,26)=12.9, p=0.001$
Filter condition ×TBR	$F(6,123)=1.39, p>0.05$	$F(3,78)=0.88, p>0.05$

For the tone vocoder, the benefit of providing *f*0-related information in all channels (condition PP versus condition EE) was somewhat greater when *N* was 6 than when it was 18. This may partly reflect a form of trading between spectral and temporal resolution that has been described before; if spectral resolution is poor ($N=6$), then providing extra temporal information has a large effect, whereas when spectral resolution is good ($N=18$), providing extra temporal information has only a small effect (Xu *et al.*, 2005). However, the larger effect when *N* was 6 probably also occurred because the relatively broad channels in this case allowed *f*0-related information to be conveyed more effectively, as discussed in more detail in the following.

For the tone vocoder, when *N* was 6 or 11 (and consequently channel bandwidths were relatively large), *f*0-related information was more useful when it was present in the channels with low center frequencies than when it was present in the channels with high center frequencies (performance was better for condition PE than for condition EP). This may have happened for at least two reasons: (1) *f*0-related temporal information (i.e., information from the temporal envelope) may be extracted more effectively from lower-frequency spectral regions (but regions where spectral sidebands are at least partially unresolved in the auditory system) than from higher-frequency spectral regions (Moore, 2003b; Bernstein and Oxenham, 2005), provided that the channels used in the vocoder are sufficiently wide to pass the spectral sidebands produced by modulation at *f*0 rates.

(2) For $N=6$ and 11, the spectral sidebands produced by modulation of the sinusoidal carrier might have been resolved by the listeners. The auditory filter bandwidth limits the ability to resolve spectral components (Plomp, 1964; Moore and Ohgushi, 1993; Moore *et al.*, 2006); generally, components have to be separated by about $1.25\mathrm{ERB}_N$ for them to be resolved (Moore, 2003b). Hence, such spectral cues would have been available mainly for vocoder channels tuned to lower center frequencies, since the bandwidth of the auditory filter is smallest at low frequencies. The resolved spectral cues might have led to more effective extraction of *f*0-related information (Gonzales and Oliver, 2005; Whitmal *et al.*, 2007). Note that spectral cues of this type would not be usable for the noise vocoder, since the sidebands for a given channel of a noise vocoder are distributed across the whole width of that channel, and out-of-band components are removed by the filtering applied subsequent to application of modulation to that channel. Also, such spectral cues would not be available to a person with a CI.

When *N* was 18 (and consequently channel bandwidths were small), *f*0-related information was more useful when it was present only in the channels with high center frequencies than when it was present only in the channels with low center frequencies (performance was better for condition EP than for condition PE). This probably happened because, with this value of *N*, the vocoder channels tuned to lower center frequencies were too narrow to pass the spectral sidebands produced by modulation at *f*0 rates.

In a CI, the transmission of modulation at $f0$ rates may also be limited by the filtering used to form the channel signals. Again, this limitation would mainly occur for channels tuned to lower center frequencies. However, such a limitation need not necessarily apply. For example, the channel filters could be made to overlap, so that each channel had a sufficiently large bandwidth (say 300 Hz) to pass modulation at $f0$ rates for most talkers. Another processing strategy, which is being used in some CI devices, is to deliver $f0$-related information via the detailed timing of the electrical pulses delivered to channels tuned to low spectral frequencies (Hochmair *et al.*, 2006). Finally, Laneau *et al.* (2006) have described a method of enhancing transmission of $f0$-related information by modulating the channel envelopes sinusoidally at the fundamental frequency ($F0$) of the input signal, with 100% modulation depth and in phase across channels to maximize temporal-envelope pitch cues.

B. High-rate envelope modulation information improves both *f*0-related information and representation of envelope onset

As described earlier, using a single-channel noise vocoder, Van Tasell *et al.* (1987) found better consonant perception with f_{lp}=200 Hz than with f_{lp}=20 Hz. They attributed this to the better representation of $f0$-related information with the former. Using a single-channel tone vocoder, Grant *et al.* (1991) found that performance reached an asymptote for f_{lp}=67 Hz. With this value of f_{lp}, $f0$-related information would have been reduced but not eliminated, as it passed through the tail of the low-pass filter (given their slope of about −15 dB/octave). As shown in Fig. 2, our P filter, which is similar to the 200 Hz low-pass filter used by Van Tasell *et al.*, preserved $f0$-related information better than the E filter and gave a slightly improved representation of envelope onsets. The benefit of increasing f_{lp} found by Van Tasell *et al.* and by Grant *et al.* might therefore have been caused either by improved transmission of $f0$-related information or by improved representation of envelope onsets or both.

Using a noise vocoder, Xu and Zheng (2007) reported that, for vowel or consonant recognition in quiet or in steady noise, when N was 12 or more there was no improvement in intelligibility once f_{lp} was 16 Hz or greater. Even allowing for the shallow slope (−12 dB/octave) of their low-pass filter, $f0$-related information did not appear to play a role for large N. However, when N was small (2 or 4) there was some improvement in performance when f_{lp} was increased above 16 Hz, which is consistent with the results of Van Tasell *et al.* (1987) and Grant *et al.* (1991). Our results show a benefit of the P filter relative to the E filter, even for large N, a result which differs from that of Xu and Zheng (2007). The difference can probably be attributed to our use of two talkers, the target and interferer. Good performance on a two-talker separation task is traditionally associated with the need to have access to $f0$-related information (Brokx and Nooteboom, 1982). We conclude that the observed benefit of the P filter relative to the E filter is probably mainly attributable to the provision of $f0$-related information, although provision of better envelope-onset information may also play a small role.

C. Comparison of noise and tone vocoders

Performance was worse for the noise than for the tone vocoder. Also, the effect of filter condition was greater for the tone vocoder. The effects of carrier type found here were larger than those reported by Dorman *et al.* (1997), but comparable to those found by Whitmal *et al.* (2007). The better performance for the tone vocoder may be due to at least two factors: (1) The inherent low-rate random envelope fluctuations in the noise carrier may disrupt the processing of the envelope modulation related to the input signal. (2) For the tone vocoder, spectral cues resulting from resolution of sidebands falling within a single vocoder channel may have led to improved extraction of $f0$-related information. However, this effect would occur mainly for small N.

The better performance for the tone vocoder cannot be due solely to factor (2), since it occurred even for condition EE, for which there would probably have been no resolvable sidebands in any channel. For example, for N=11, the mean score in experiment 1 for condition EE was about 48% for the noise vocoder and 65% for the tone vocoder. For N=18, the corresponding scores were 62% and 79%. Also, when $f0$-related information was present only in channels above 1500 Hz (condition EP), spectral sidebands produced by modulation at $f0$ rates would not have been resolvable, but performance was still better for the tone than for the noise vocoder. For example, for N=11, the mean score in experiment 1 for condition EP was about 49% for the noise vocoder and 71% for the tone vocoder.

D. Implications for simulations of CI processing

This section considers the issue of what type of carrier would be most appropriate for simulation of the information delivered by a CI. The tone carrier has the advantage that it avoids the inherent random fluctuations produced by the noise carrier. Such fluctuations would not be present in a CI, so the tone vocoder would seem to be more appropriate. On the other hand, when N is small, a tone vocoder may give rise to usable spectral cues in channels with low center frequencies; such cues would not be available in a CI. The value of N chosen for use in CI simulations is often rather small, since the "effective" number of channels in a real CI appears to be small, presumably because of current spread within the cochlea (Shannon, 1983; Friesen *et al.*, 2001; Kwon and van den Honert, 2006). An alternative to noise and tone vocoders is to use a harmonic complex tone as a carrier, as proposed by Deeks and Carlyon (2004). This avoids the random fluctuations associated with a noise carrier, although it still introduces spurious modulation at a rate determined by the $f0$ of the harmonic complex carrier (Deeks and Carlyon actually used an alternating-phase complex; when the components are unresolved, this leads to an envelope modulation rate which is double the $f0$). Overall, it appears that all types of carriers have some disadvantage.

It should be noted that there are several other problems in using simulations of CIs. For example, the ability of people with CIs to process modulation and discriminate changes in modulation rate may be different from that of normally hearing people listening to the simulation (Shan-

non, 1992; Grant *et al.*, 1998). Also, the dynamic range of a person listening via a CI is much less than that of a normally hearing person listening to acoustic stimuli (Moore, 2003a). At best, a simulation may mimic the type of information provided by a CI, but, as the present results show, even this aspect of simulations is problematic.

It is of interest to compare the intelligibility scores obtained here with those obtained from users of CIs under similar conditions. Nie *et al.* (2006) measured the intelligibility of the HINT sentences (Nilsson *et al.*, 1994), spoken by a male talker, in a background of a female talker, using a TBR of 10 dB. The cutoff frequency used for envelope extraction was high (500 Hz or more), but with our use of a background of a male talker, the results can only tentatively be compared with our results for condition PP. The scores obtained by Nie *et al.* were between 30% and 60%, depending on the number of electrodes (4 to 12) and the stimulation rate (1000–4000 pulses/s) used in the CI. In our simulations using the tone vocoder, the mean score for $N=6$ with TBR $=10$ dB was about 65%, slightly better than the best score found with the CI users. This happened despite the fact that the IEEE sentences used here are somewhat more difficult than the HINT sentences used by Nie *et al.* (Stickney *et al.*, 2004). Stickney *et al.* (2004) measured intelligibility for IEEE sentences (IEEE, 1969), spoken by a male talker, in the presence of either the same male talker, a different male talker, or a female talker. The cutoff frequency used for envelope extraction was 500 Hz, so again the results can reasonably be compared with our results for condition PP. For the condition with a different male talker, which is comparable to the condition used in the present paper, scores were about 30% correct for a TBR of 10 dB and 48%–50% for TBRs of 15 and 20 dB. These scores are markedly lower than found here using the tone vocoder. Overall, these comparisons show that CI users perform more poorly than normal-hearing subjects listening via the tone vocoder. This might reflect cues provided by the tone vocoder (such as resolved sidebands) that are not available to implant users. However, it might also reflect the perceptual limitations of the CI users, as described earlier.

V. CONCLUSIONS

High-rate ($f0$-related) envelope information in noise and tone vocoders was found to contribute to intelligibility in a competing-speech task. The information was conveyed more effectively by carriers which were deterministic (pure tones) than random (bands of noise), and it was used most effectively when it was presented to all channels. The $f0$-related modulation information was conveyed less effectively through low-frequency channels when N is large. However, even in this case, performance was better when $f0$-related information was present in all channels (condition PP) than when it was present only in channels with center frequencies above 1500 Hz (condition EP). Overall, our results indicate that vocoder simulations of CI processing should be interpreted cautiously, as there are disadvantages associated with both noise and tone vocoders.

ACKNOWLEDGMENTS

M.A.S. and B.C.J.M. were supported by the MRC (UK). C.F. was supported by a Marie-Curie fellowship from the EU. The authors thank Associate Editor Ken Grant and two reviewers for helpful comments on an earlier version of this paper.

ANSI. (**1997**). *ANSI S3.5-1997, Methods for the Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Apoux, F., and Bacon, S. P. (**2004**). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," J. Acoust. Soc. Am. **116**, 1671–1680.

Baskent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.

Bernstein, J. G., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Breeuwer, M., and Plomp, R. (**1986**). "Speechreading supplemented with auditorily presented speech parameters," J. Acoust. Soc. Am. **79**, 481–499.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature (London) **178**, 535–536.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.

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.

Dorman, M. F., Loizou, P. C., Spahr, A. J., and Maloff, E. (**2002**). "A comparison of the speech understanding provided by acoustic models of fixed-channel and channel-picking signal processors for cochlear implants," J. Speech Lang. Hear. Res. **45**, 783–788.

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

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

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.

Erber, N. P. (**1969**). "Interaction of audition and vision in the recognition of oral speech stimuli," J. Speech Hear. Res. **12**, 423–425.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," J. Acoust. Soc. Am. **108**, 1877–1887.

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.

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998a**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998b**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Gonzales, 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.
Grant, K. W., Braida, L. D., and Renn, R. J. (**1991**). "Single band amplitude envelope cues as an aid to speechreading," Q. J. Exp. Psychol. A **43A**, 621–645.
Grant, K. W., Summers, V., and Leek, M. R. (**1998**). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 1051–1060.
Healy, E. W., and Steinbach, H. M. (**2007**). "The effect of smoothing filter slope and spectral frequency on temporal speech information," J. Acoust. Soc. Am. **121**, 1177–1181.
Hochmair, I., Nopp, P., Jolly, C., Schmidt, M., Schosser, H., Garnham, C., and Anderson, I. (**2006**). "MED-EL cochlear implants: State of the art and a glimpse into the future," Trends Amplif. **10**, 201–219.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**, 225–246.
Keppel, G. (**1991**). *Design and Analysis: A Researcher's Handbook* (Prentice Hall, Upper Saddle River, NJ).
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.
Kong, Y. Y., and Zeng, F. G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**, 2830–2840.
Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.
Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.
Laneau, J., Wouters, J., and Moonen, M. (**2006**). "Improved music perception with explicit pitch coding in cochlear implants," Audiol. Neuro-Otol. **11**, 38–52.
Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Radiation Laboratory Series Vol. **24** (McGraw-Hill, New York).
Loizou, P. C., Dorman, M., Poroy, O., and Spahr, T. (**2000**). "Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution," J. Acoust. Soc. Am. **108**, 2377–2387.
Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–2103.
Lorenzi, C., Dumont, A., and Füllgrabe, C. (**2000**). "Use of temporal envelope cues by children with developmental dyslexia," J. Speech Lang. Hear. Res. **43**, 1367–1379.
Luo, X., and Fu, Q. J. (**2004**). "Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants," J. Acoust. Soc. Am. **116**, 3659–3667.
Moore, B. C. J. (**2003a**). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," Otol. Neurotol. **24**, 243–254.
Moore, B. C. J. (**2003b**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).
Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).
Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.
Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.
Nie, K., Barco, A., and Zeng, F. G. (**2006**). "Spectral and temporal cues in cochlear implant speech perception," Ear Hear. **27**, 208–217.
Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Plomp, R. (**1983**). "The role of modulation in hearing," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin).
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on *f*0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rosen, S. M., Fourcin, A. J., and Moore, B. C. J. (**1981**). "Voice pitch as an aid to lipreading," Nature (London) **291**, 150–152.
Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.
Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.
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.
Smith, M. W., and Faulkner, A. (**2006**). "Perceptual adaptation by normally hearing listeners to a simulated 'hole' in hearing," J. Acoust. Soc. Am. **120**, 4019–4030.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023–1034.
Studebaker, G. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). "Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants," J. Acoust. Soc. Am. **92**, 1247–1257.
Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.
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., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.
Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.

Analysis of facial motion patterns during speech using a matrix factorization algorithm

Jorge C. Lucero[a)]
Department of Mathematics, University of Brasilia, Brasilia DF 70910-900, Brazil

Kevin G. Munhall[b)]
Departments of Psychology and Otolaryngology, Queen's University, Kingston Onatrio K7L 3N6, Canada

(Received 13 January 2008; revised 19 July 2008; accepted 21 July 2008)

This paper presents an analysis of facial motion during speech to identify linearly independent kinematic regions. The data consists of three-dimensional displacement records of a set of markers located on a subject's face while producing speech. A QR factorization with column pivoting algorithm selects a subset of markers with independent motion patterns. The subset is used as a basis to fit the motion of the other facial markers, which determines facial regions of influence of each of the linearly independent markers. Those regions constitute kinematic "eigenregions" whose combined motion produces the total motion of the face. Facial animations may be generated by driving the independent markers with collected displacement records.
 [DOI: 10.1121/1.2973196]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt [BHS] Pages: 2283–2290

I. INTRODUCTION

The general goal of this work is to develop a mathematical model of facial biomechanics for applications to speech production and perception studies. The model must be capable of producing computer-generated animations of speech with an acceptable level of realism and should allow for direct manipulation of facial movement parameters (Munhall and Vatikiotis Bateson, 1998).

The core of such a system must be some mathematical representation of the physiology of the human face. When building models of facial physiology, two general strategies have been followed (cf. Beautemps *et al.*, 2001). One is a theoretical modeling strategy, in which the physiological structure of passive tissues and active muscles is explicitly described commonly by differential equations. The face behavior is simulated by computing the forces that act on the tissues and their resultant deformations. In this way, the dynamics of the system is incorporated into the model. Such models are theoretical in the sense that assumptions are made in choosing the key attributes of the models and parameters are estimated or determined from best available measures from the literature. This strategy was pioneered by the muscle-based facial animation work of Terzopoulos and Waters (1990); Lee *et al.*, (1995); and Parke and Waters (1996), in which some of the facial muscles and some aspects of the soft tissue were modeled. Following their work, a three-dimensional (3D) biomechanical model of the face was developed for producing animations of speech (Lucero and Munhall, 1999). The model was driven by recorded perioral electromyographic signals and, in a later version, by records of facial kinematics (Pitermann and Munhall, 2001). In general, the model was able to generate animations with a reasonable level of visual realism and showed potential as a computational tool to study the facial physiology of speech. However, it has been argued that, in its current stage of development, the strategy is still not appropriate for the intended application to speech perception research (Lucero *et al.*, 2005). Its main drawback is the difficulty of producing a representation of the complex structure of muscles and passive tissues precise enough to capture details of speech movement patterns and, at the same time, that could be easily adapted to different subjects.

Within this general strategy, the application of free form deformations by Kalra *et al.* (1992) may also be included. There each facial muscle is represented as a deformable facial region, whose motion is controlled by the movement of a facial point. The geometry of the facial regions and their deformation properties are defined based on information from available anatomical data. Although the result is not a dynamical model, it is still a theoretical representation of the face (because it is based on average physical and anatomical parameters and it thus represents a generic face), and suffers from the same drawbacks noted above.

A second strategy is empirical modeling. In this case, a relation between various facial parameters measured during speech is constructed. For example, in Kuratate *et al.*'s statistical modeling work (Kuratate *et al.*, 1998), principal component analysis (PCA) is used to decompose a set of measured facial shapes into orthogonal components, and determine a reduced base of eigenfaces. Arbitrary facial shapes and movements are next generated by driving a linear combination of the eigenfaces. The physiology of the face is not modeled explicitly, although it is still present, in implicit form, in the model's equations. Besides facial kinematics, other speech parameters may be incorporated into the model, such as muscle electromyography and acoustics (Vatikiotis-Bateson and Yehia, 1996; Kuratate *et al.*, 1999). Similar em-

a)Electronic mail: lucero@unb.br
b)Electronic mail: kevin.munhall@queensu.ca

pirical modeling strategies using independent component analysis (ICA) have also been proposed (Müller *et al.*, 2005).

In a previous work (Lucero *et al.*, 2005), an empirical model was introduced that was based on decomposing the face surface into a finite set of regions. The total motion of the face was computed as the linear combination of the movement of those regions. The model was built by analyzing the recorded 3D position of a set of markers placed on a subject's face, while producing a sequence of sentences. An algorithm grouped the markers into a set of clusters, which had one primary marker and a number of secondary markers with associated weights. The displacement of each secondary marker was next expressed as the linear combination of the displacements of the primary markers of the clusters to which it belonged. The model was next used to generate facial animations, by driving the primary markers and associated clusters with collected data.

It was argued that the computed cluster structure represented the degrees of freedom (DOF) of the system. The DOF are the independent modes of variation and thus the independent sources of information. Subjects differ in the information transmission from their faces and this may be due in part to differences in their DOF. In the model, the facial clusters define facial eigenregions, whose combined motion forms the total motion of the facial surface.

This empirical model and the PCA (or ICA) approach are based on linear modeling of the data which can result in similar levels of accuracy in reconstruction of the data. However, the resulting solutions are quite different and serve different purposes. The PCA approach extracts global structure and will produce a set of primary gestures. Its DOF are the functional modes of deformation of the face. In the present case, the aim is to identify a set of spatially distinct regions that move independently as the DOF. There presumably is a mapping between these two representations and it could be hypothesized that the nervous system must know what to control in its anatomical DOF to produce its gestural basis set.

There are a number of advantages to focusing on the spatial DOF. This approach focuses the data analysis on the generating mechanism for gestures: The facial musculature. The action of muscles is obviously spatially concentrated and thus facial regions can be found that are associated with individual muscles or synergies of muscles that are in close proximity or whose actions are spatially localized. This muscle-based approach is consistent with a productive tradition in the analysis of facial expression and the study of perception of expressions (Ekman *et al.*, 2002) and this approach has also been a powerful tool in facial animation (e.g., Terzopoulos and Waters, 1990). Another advantage of finding regional DOF in a data set is that these regions can be animated in arbitrary facial configurations. There is considerable interest in speech perception research on the role of individual talker characteristics in speech perception (e.g., Goldinger, 1996). Studies that involve the use of animating a generic face or the animation of one talkers morphology with another talkers motion must solve a registration and morphing problem (Knappmeyer *et al.*, 2003). The identification of key features and spatial regions is one form of solution to this correspondence problem.

This empirical modeling approach is close to the articulatory modeling work of Badin, Bailly, *et al.* (Badin *et al.*, 2002; Beautemps *et al.*, 2001; Engwall and Beskow, 2003). In their work, PCA is used to determine articulatory parameters to control the shape of a 3D vocal tract and face model. For better correspondence to the underlying biomechanics, some of the parameters (e.g., jaw height, lip protrusion, etc.) are defined *a priori*, and their contributions are subtracted from the data before computing the remaining components. The present approach proposes to rely entirely on the data to build the model, with as few prior assumptions as possible. Instead of setting a model by defining the biomechanical properties of skin tissue and muscle structure based on *a priori* theoretical reasons, a possible model is inferred just by looking at the measured motion patterns of the facial surface.

The algorithm presented by Lucero *et al.* (2005) had a preliminary nature and had some drawbacks. For example, it was necessary to define an initial facial mesh, linking nodes corresponding to the initial position of the markers and leaving "holes" for the eyes and mouth. Also, a more solid foundation for the criteria for grouping the markers was desired. In the present paper, an improved version is introduced, which uses a QR factorization technique (Golub and Loan, 1996) to identify a linearly independent subset of facial markers. This subset is next used as a basis to predict the displacement of arbitrary facial points.

II. DATA

The data consist of the 3D position of 57 markers distributed on a subject's face, recorded with a Vicon equipment (Vicon Motion Systems, Inc., Lake Forest, CA) at a 120 Hz sampling frequency. According to calibration data, the position measures were accurate within 0.5 mm. An additional six markers on a head band were used to determine a head coordinate system, with the positive directions of its axes defined as: the x axis in the horizontal direction from right to left, the y axis in the vertical direction to the top, and the z axis in the protrusion direction to the front. The position of the 57 facial markers was expressed in those coordinates, with the approximate location shown in Fig. 1. Neither the morphology of subjects' faces nor the precise placement of the markers is exactly symmetrical and thus the markers seem to be offset from the facial features in the schematic and symmetrical face in the figure.

Data were recorded from 2 subjects (S1 and S2), while they were producing selected sentences from the Central Institute for the Deaf Everyday sentences (Davis and Silverman, 1970), listed in http://www.mat.unb.br/lucero/facial/qr2.html. A large portion of the data for the marker at the left upper eyelid (marker 13) of subject S1 was missing, due to recording errors. Since the motion patterns of right and left upper eyelids seemed very close (by visual assessment), the missing data of the left upper eyelid were copied from the marker at the other eyelid. It will be shown later that the

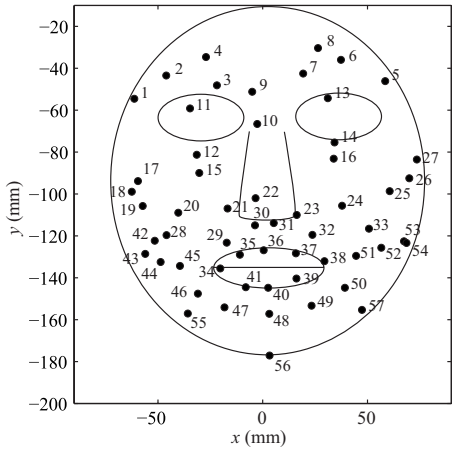


FIG. 1. Spatial distribution of the marker positions superimposed on a schematic face.

algorithm detected this artifact within the data, which serves as a proof of its ability to analyze kinematic patterns.

A total of 40 sentences was recorded from subject S1 and 50 sentences from subject S2. In the case of S2, the set of 50 sentences was recorded twice, forming two datasets which will be denoted as S2a and S2b. In the recording sessions, the subjects were asked to adopt a consistent rest position at the beginning of each sentence. The initial positions of the markers were taken as representative of a rest (neutral) configuration.

III. THE SUBSET SELECTION PROBLEM

This approach for building an empirical facial model is based on the so-called subset selection problem of linear algebra (Golub and Loan, 1996). Assume a given data matrix $A \in \mathbb{R}^{m\times n}$ and the observation vector $b \in \mathbb{R}^{m\times 1}$, with $m \geqslant n$, and that a predictor vector x is sought in the least squares sense, which minimizes $\|Ax-b\|_2^2$. Assume also that the data matrix A derives from observations of redundant factors. Therefore, instead of using the whole data matrix A to predict b, it may be desirable to use only a subset of its columns, so as to filter out the data redundancy. The problem is, then, how to pick the nonredundant columns. In the present case of facial modeling, the data matrix will contain the displacements of the 57 facial markers, arranged in columns. A small subset of markers (columns of the data matrix) must be selected which may be used to predict the motion of any other arbitrary facial point.

The idea of reducing the data set is consistent with the long-held view that the speech production system is itself low-dimensional. As Bernstein (1967) suggested about motor control in general, the nervous system acts to reduce the potential DOF. In speech the muscles and articulators are coupled synergistically during articulation in order to produce particular sounds.

To solve the subset selection problem, the most linearly independent columns of matrix A must be identified. Let A_k denote a subset of k columns of A. A measure of "independency" of the subset is provided by the smallest singular value of A_k, σ_k, which measures the distance of A_k to the set of k-rank singular matrices, in the 2-norm. Thus, it indicates how far A_k is from being a singular matrix. Consequently, the smaller σ_k, the more independent the subset A_k. In principle, the subset selection problem could be solved by testing all possible combinations of k columns from the total of n columns of A. The number of possible combinations is $n!/[k!(n-k)!]$, which could be prohibitively large. In the present case, with $n=57$ and adopting $k=10$ as an example, the number of combinations is 4.32×10^{10}. So far, it appears that an exhaustive search is the only means to compute the optimal solution to the problem (Lawson and Hanson, 1987; Björck, 2004).

A possible solution to the subset selection problem is provided by the algorithm of QR factorization with column pivoting (Golub and Loan, 1996; Chan and Hansen, 1992). That algorithm decomposes A in the form $A\Pi=QR$, where $\Pi \in \mathbb{R}^{n\times n}$ is a column permutation matrix, $Q \in \mathbb{R}^{m\times n}$ is an orthogonal matrix, and $R \in \mathbb{R}^{n\times n}$ is an upper triangular matrix with positive diagonal elements[1] The first column of the permutated matrix $A\Pi$ is just the column of A that has the largest 2-norm (euclidean norm). The second column of $A\Pi$ is the column of A that has the largest component in a direction orthogonal to the direction of first column. In general, the kth column of $A\Pi$ is the column of A with the largest component in a direction orthogonal to the directions of first $k-1$ columns [or, equivalently, is the column of A which has maximum distance from the subspace spanned by the first $k-1$ columns (Björck, 2004)]. The diagonal elements of R (r_{kk}), also called the R values, measure the size of those orthogonal components; they appear in decreasing order for $k=1,\ldots,n$, and tend to track the singular values of matrix A. Thus, the algorithm reorders the columns of A to make its first columns as well conditioned as possible. The first k columns of $A\Pi$ may be then adopted as the sought subset of k least dependent columns.

Another useful property of the above algorithm is that it reveals the rank of matrix A. Let us define the following block partitions for R,

$$R=\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}, \tag{1}$$

where $R_{11} \in \mathbb{R}^{k\times k}$, and the dimensions of the other blocks match accordingly. If $\text{rank}(A)=k<n$, then $R_{22}=0$. Letting σ_i, for $i=1,\ldots,n$ be the singular values of A, with $\sigma_1 \leqslant \sigma_2 \leqslant \cdots \leqslant \sigma_n$, it may be shown that $\sigma_{k+1} \leqslant \|R_{22}\|_2$. Therefore, a small value of $\|R_{22}\|_2$ implies that A has at lest $n-k+1$ small singular values, and thus A is close to being rank k (Chan, 1987). As a tolerance value, ϵ, it is usual to consider the level of uncertainties or precision of the data. Hence, a value of $\|R_{22}\|_2 \leqslant \epsilon$ implies that A has ϵ-rank^2 k.

A number of other algorithms have been proposed to find solutions to the subset selection problem, based, for instance, on singular value decomposition (Golub and Loan, 1996), search techniques exploiting partial ordering of the variables, stepwise regression algorithms (Lawson and Hanson, 1987), and backward greedy algorithms (de Hoog and Mattheij, 2007). However, the QR factorization offers a number of advantages (e.g., Golub and Loan, 1996; Björck, 2004; Setnes and Babuska, 2001; Chan and Hansen, 1992; Bischof and Quintana-Ort, 1998): It is computationally sim-

pler and numerically robust. It detects the rank of the data and provides a good technique for solving least squares problems, as shown below. It has been used in numerous technical applications of the subset selection problem, including fuzzy control systems (Setnes and Babuska, 2001), neural network design (Kanjilal *et al.*, 1993), signal bearing estimation (Prasad and Chandna, 1991), noise control systems (Ruckman and Fuller, 1995), very large scale integrated array implementation (Lorenzelli *et al.*, 1994), and wireless communication system design (Migliore, 2006), and is adopted in the present study.

Assume that the matrices Q, R, and Π have been computed, so that $A\Pi=QR$, and let $\Pi=[\Pi_1\ \Pi_2]$, where $\Pi_1 \in \mathbb{R}^{n\times k}$, $\Pi_2 \in \mathbb{R}^{n\times n-k}$. The first k columns of $A\Pi$ are therefore given by $A\Pi_1$ and are selected as a subset of k independent columns. An observation vector b may be predicted by minimizing $\|A\Pi_1 x-b\|_2^2$. Letting $Q^t b=[c\ d]^T$, where $c \in \mathbb{R}^{k\times 1}$, $d \in \mathbb{R}^{n-k\times 1}$ then the minimizer may be easily computed as the solution of the upper triangular system $R_{11}x=c$ (Golub and Loan, 1996).

In the present case, the k columns of $A\Pi_1$ will be used to predict the remaining $n-k$ columns of A, given by $A\Pi_2$, in the least squares sense. The problem may be expressed as the minimization of

$$E=\sum_i \|A\Pi_1 x_i-(A\Pi_2)_i\|_2^2, \tag{2}$$

where the subindex i represents each of the $n-k$ columns of $A\Pi_2$. Using the Frobenius norm (euclidean matrix norm), produces

$$E=\|A\Pi_1 X-A\Pi_2\|_{\mathrm{F}}^2, \tag{3}$$

where X is a $k\times(n-k)$ matrix. Since the norm is invariant under orthogonal transformations, then

$$E=\|Q^T A\Pi_1 X-Q^T A\Pi_2\|_{\mathrm{F}}^2=\|R_{11}X-R_{12}\|_{\mathrm{F}}^2+\|R_{22}\|_{\mathrm{F}}^2. \tag{4}$$

Therefore, the least square minimizer is the solution of the upper triangular system $R_{11}X=R_{12}$, and the residual is $\|R_{22}\|_F$.

IV. ANALYSIS OF FACIAL DATA

The displacement of each marker was computed relative to the initial neutral position. For each subject, the markers' displacements for all sentences where concatenated and arranged in a displacement matrix $A \in \mathbb{R}^{3M\times N}$, where N is the number of markers (57) and M is the total number of time samples of all the concatenated sentences. QR factorization with column pivoting was then applied to data matrix A, using a standard MATLAB implementation.

Figure 2 shows the computed R values, normalized to the size of matrix R, for both subjects. In both cases, the values decrease smoothly. In the case of subject S1, there is a sudden drop for the last value, with $r_{56,56}=22.70$ mm and $r_{57,57}=0.10$ mm. Since the precision of the data is $\epsilon \approx 0.5$ mm, then for $k=56$, $\|R_{22}\|_2=|r_{57,57}|\leqslant\epsilon$ and therefore rank$(A)=56$. This result indicates that data for one marker are redundant. In fact, it reflects the filling of missing values for the left upper eyelid from the data collected for the right one.

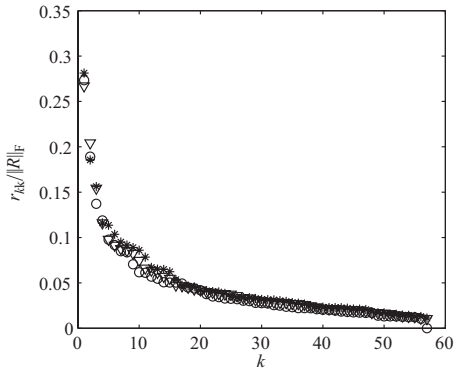


FIG. 2. R values (diagonal elements of matrix R) normalized to the size of R, for subjects S1 (circles), S2a (stars), and S2b (triangles).

In the case of subject S2 (datasets S2a and S2b), there is no gap in the data, and rank$(A)=57$. Thus, any model built from a subset of less than 56 or 57 markers will necessarily result in a loss of information.

Figure 3 shows the first 12 R values normalized to the size of matrix R, for subject S1, when varying the number of sentences in the dataset. The values stabilize for sets with more than approximately 25 sentences. Any larger data set is therefore reliable enough for building a model. The datasets for subject 2 produce similar results, and are therefore not shown.

Table I shows the index of the first 16 columns (or markers) selected by the algorithm, for the various datasets. A set of 30 sentences was used in all analyses.

In all cases, the first selected marker is the 40th, at the center of the lower lip (see Fig. 1), which has therefore the largest displacement (largest 2-norm of the associated column). The second is marker 34, at the lip's left corner. From the third marker, subject S1 shows a different pattern than subject S2. In the case of subject S1, the next four markers are lip's right corner (38), right eyebrow (2), upper lip center (36), and left eyebrow (6). Subject S2, on the other hand, besides the lip' right corner (38), incorporates the eyelids (13 or 11), and markers at the lower-right portion of the face (42, 43, 47, 48, or 56), depending on the dataset. The upper lip marker (6) appears later, in position 8th–10th. The four columns of results for subject S2 show similar markers, although some differences in the selected markers and the or-

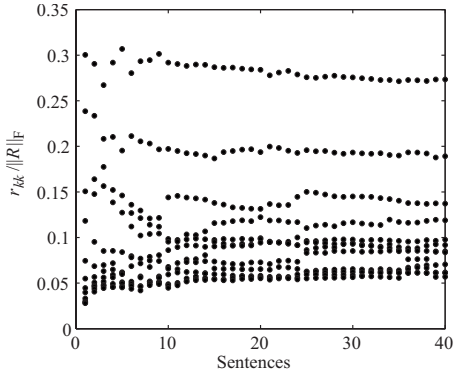


FIG. 3. Normalized values of the first 12 R values vs number of sentences in the data set, for subject S1.

TABLE I. Selected columns (markers) of data matrix *A*. For S1, S2a, and S2b, the first 30 sentences of each dataset were used. In case of S2a* and S2b*, the last 30 sentences (from a total of 50) of the respective datasets were used.

Order	S1	S2a	S2b	S2a*	S2b*
1	40	40	40	40	40
2	34	34	34	34	34
3	38	13	13	13	11
4	2	38	38	42	38
5	36	47	48	38	56
6	6	42	47	43	42
7	20	6	11	47	6
8	49	56	36	11	47
9	11	11	42	36	13
10	52	36	49	6	36
11	54	48	6	49	12
12	47	39	53	12	48
13	22	52	20	48	39
14	32	20	39	16	16
15	39	49	56	20	53
16	48	14	14	52	20

der they appear may be noted. For example, looking only at the first seven markers in the perioral region for S2, we note that 6 of them appear in the four data sets: 34, 36, 38, 40, 42, and 47. The remainder marker is 48 (S2b), 56 (S2a and S2b*), or 43 (S2a*). Markers 48 and 56 are both at the chin and close together, so they may be considered as belonging to the same facial region. Marker 43, on the other hand, belongs from a different region, close to the right border of the face.

In the case of subject S1, the last selected marker is the 13th (left upper eyelid). Since the algorithm already detected a rank 56 for the data (and therefore a redundant data column), then the data column for that marker is redundant. That result provides a validation of the algorithm, showing its capability to detect artifacts introduced into the data.

Once the main columns or markers have been selected, a least square fit of the remaining columns may be computed by solving $R_{11}X=R_{12}$, as explained in the previous section. As a numerical example, a basis of ten independent markers (this number includes up to the lower lip marker in all datasets) was adopted. Figure 4 shows the results of the fit, for subject S1. There the fitting coefficients computed for the secondary markers have been extended to other facial points by cubic interpolation to improve visualization of the results. Note that, in general, the regions associated with each of the markers include both positive and negative subregions, where motion is in the same and opposite direction, respectively, to the marker's motion. Motion of the each kinematic region is therefore determined by the motion of its associated independent marker, and the total motion of the face is composed by the linear combination of all the kinematic regions. The regions seem to distribute in similar numbers and locations on both sides of the face, although they show a large asymmetry. Regarding the eyelids, although only marker 11 (right eyelid) was identified as an independent marker, marker 13 (left eyelid) has a fit coefficient of 1, which indicates identical motion patterns.

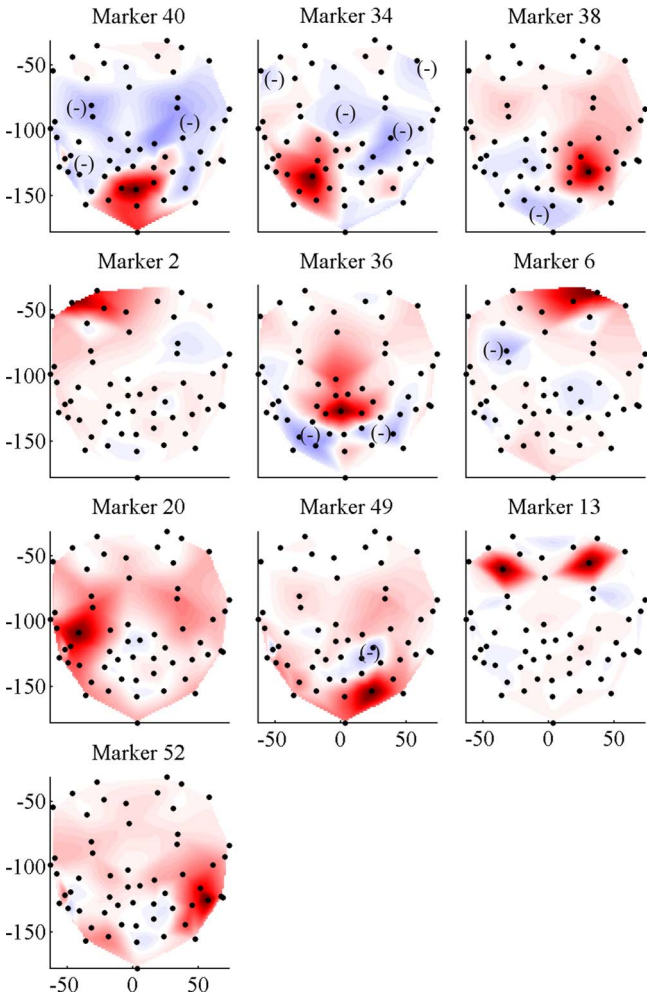


FIG. 4. (Color online) Independent kinematic regions for subject S1, when a basis of ten markers is adopted. The darker the region, the larger the least square fit coefficient of each point relative to the main marker. A minus sign indicates a subregion with negative weight.

For comparison, regions for subject S2 are shown in Figs. 5 (S2a) and 6 (S2b), which correspond to markers at the center of lower lip, both lip corners, and center of upper lip. Particularly, note that although the upper lip marker appears in different positions in the list of independent markers in Table I: 8th for S2a and 10th for S2b, the associated regions have similar shapes.

Recall that each marker is selected by the algorithm in a way such that its motion component in a direction orthogonal to the motion of the already selected markers is maximum. The orthogonal directions are represented by the columns of matrix *Q*, and the components of the markers's motion in each of the orthogonal directions is given by the rows of matrix *R*. Figure 7 shows plots of the first four rows of R for subject S1 (related to markers 40, 34, 38, and 2), extended to the whole facial surface by cubic interpolation. The plots therefore represent regions with motion components in the first four orthogonal directions. The first orthogonal direction is given by the jaw-lower lip motion, which has the largest norm, and clearly dominates motion of the lower half of the face. The left lip corner has the largest motion component orthogonal to the lower lip's motion, and next the right lip corner, with the largest component orthogonal to both the

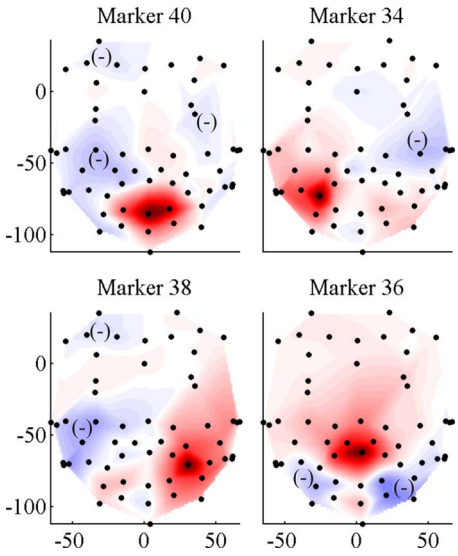


FIG. 5. (Color online) Four independent kinematic regions for subject S2a and for a basis of ten markers. The darker the region, the larger the least square fit coefficient of each point relative to the main marker. A minus sign indicates a subregion with negative weight.

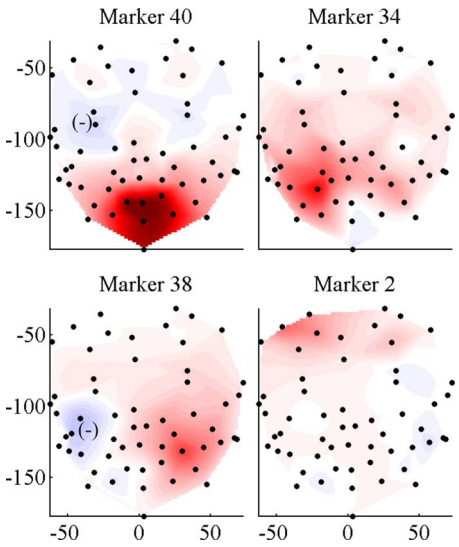


FIG. 7. (Color online) First four orthogonal regions for subject S1. Each plot shows facial points with patterns of motion in the first four orthogonal directions. The darker the region, the larger the motion. A minus sign indicates a subregion with negative weight.

lower lip and the left lip corner. The lip corner regions reflect the mouth widening-narrowing under the combined action of the orbicularis oris, zygomatic major, and other perioral muscles at each side of the face. Naturally, even though motion of both lip corners might be strongly correlated, they are associated with different regions because their motion is controlled by different groups of muscle and happens in opposite directions. This is precisely one of the intended objectives: rather than identifying particular facial gestures such as a mouth widening/narrowing action, the algorithm identifies spatial regions associated with different muscle groups.

V. COMPUTER GENERATION OF FACIAL ANIMATIONS

After the main markers and fitting matrix X have been computed, facial animations of arbitrary speech utterances may be produced by driving the selected independent markers with collected signals. Letting $P_1 \in \mathbb{R}^{n\times k}$ be the displacement matrix of the k main markers (relative to the initial neutral position), then the displacement P_2 of the secondary markers is just $P_2 = P_1 X$. The neutral position of all markers is next added back, to obtain their position in head coordinates. Finally, the position of other arbitrary facial points may be generated by using, e.g., cubic interpolation. Using this technique and the results of Table I, animations were produced for sentences 31–40 for subject S1, and 31–50 for subject S2. A virtual facial surface was generated by cubic interpolation of the recorded markers, built from a grid of 30×30 points. The animations look visually realistic, without any noticeable distortion in the motion pattern. They are available in http://www.mat.unb.br/lucero/facial/qr2.html in AVI format. Figure 8 shows an example of an animation frame.

Figure 9 shows the trajectory of marker 56 at the jaw for subject S1 in one sentence (31), when using a basis of ten

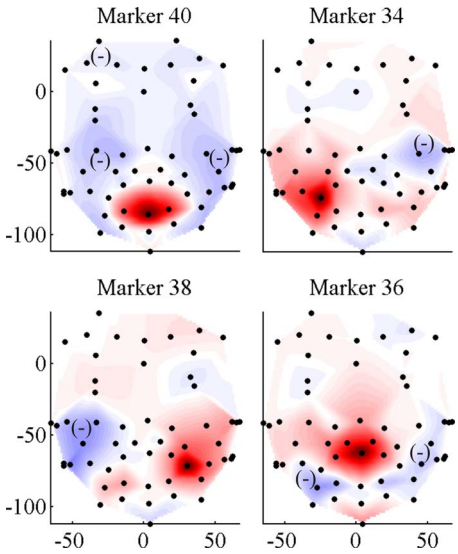


FIG. 6. (Color online) Four independent kinematic regions for subject S2b, and for a basis of ten markers. The darker the region, the larger the least square fit coefficient of each point relative to the main marker. A minus sign indicates a subregion with negative weight.

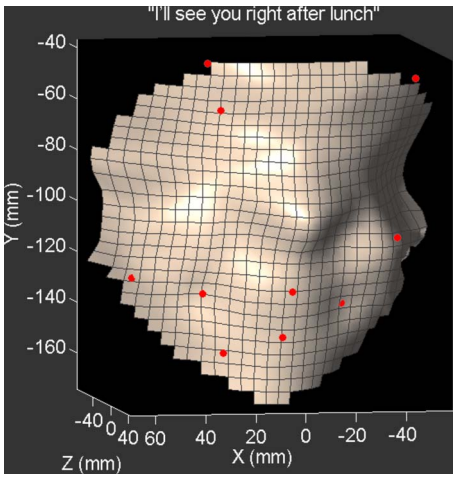


FIG. 8. (Color online) Example of a facial animation frame.

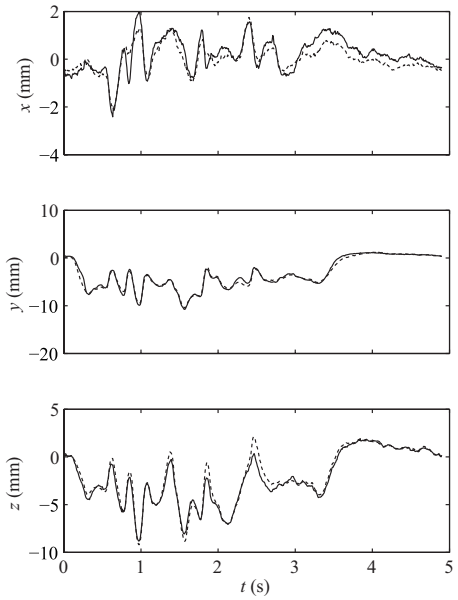


FIG. 9. Trajectory of marker 56 at the jaw for subject S1 and sentence 31, when using a basis of ten markers. Full line: trajectory computed from the model. Broken line: measured trajectory.

markers. The plots show that the model recovers the jaw trajectory with good accuracy. The maximum errors are 0.8, 2.0, and 0.9 mm for the x, y, and z directions, respectively.

Figure 10 shows the mean error of the computed trajectories when varying the number of independent markers, for both subjects in all the above sentences (the mean error is computed from the secondary markers only). Naturally, the error decreases when the number of markers is increased. The absolute error seems low even with few markers, but at the same time, the relative error seems high. The high relative error is a consequence of the small displacements of most markers.

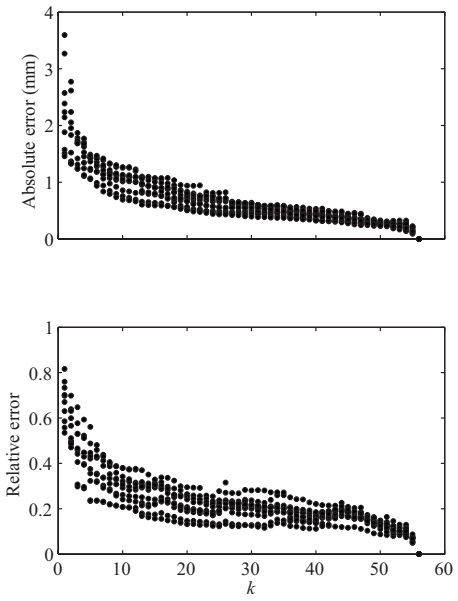


FIG. 10. Mean error of computed trajectories for subject S1 (sentences 31 to 40) vs number of markers in the selected basis. Top: absolute error. Bottom: relative error.

VI. CONCLUSION

The paper has shown that the QR factorization with column pivoting algorithm provides a convenient technique for facial motion analysis and animation. It identifies a subset of independent facial points (markers), which may be used to build an individualized linear model of facial kinematics. Each of the independent points defines a kinematics region, whose motion is determined by the point's motion. The total motion of the face is then expressed as the linear combination of the motion of the independent kinematic regions. The regions vary for different subjects, and to some degree, for different data sets. In the studied cases, however, the lower lip and both lip corners tend to be among the first (and therefore most independent) kinematic regions.

Note that the purpose of the technique is not just the reduction of the dimensionality of the data. For that purpose, the singular value decomposition (used in PCA and ICA techniques) is superior, since it permits the computation of the matrix of a given rank k that is closest to the data matrix A. Its disadvantage, for the present modeling objective, is that the computed k-rank matrix is defined in terms of a basis of k eigenvectors (the singular vectors), which do not belong, in general, to the set of column vectors of the original data. The proposed technique, on the other hand, uses a basis formed by column vectors of the data matrix, at the expense of achieving a suboptimal overall dimension reduction. However, as explained in Sec. I, the PCA and the QR factorization methods are defining DOF differently, and by extension their notion of redundant features of the data are quite different. Thus, the optimality of the dimension reduction needs to be considered in this light. The two analyses are not producing unique decompositions of the data and therefore the residual variances are quite different.

The model has an empirical nature, however, it reflects the underlying biomechanical structure of the face and may be used to infer aspects of that structure. The kinematic regions are the result of the interaction of the muscular driving forces and the the biophysical characteristics of skin tissue. Normally, when building a mathematical model of a given physiological system, one wants to separate out the plant characteristics from the control signals that are instantiated in the muscle activity. The present model, on the other hand, provides a lumped representation of the facial biomechanics.

In biological motion (point light) studies of human gait patterns, the motion is said to contain two sources of information (Troje, 2002): "motion-mediated structural information" and information that is strictly dynamical (e.g., patterns of accelerations and velocities). The human perceptual system is sensitive to both. The point light displays reveal structural information by revealing the rigid body segments and their dimensions as well as showing where the segment joints are. The face is a unique biomechanical system and the analysis of its soft tissue deformations is different from studying articulated motion like locomotion. However, motion patterns of the face can reveal its structural form. In this sense, the proposed technique is carrying out biomechanical analyses of the face. For each individual, it is letting the motions define what regions of the face move as an indepen-

dent unit, what the boundaries of the surface regions are, and where the spatial peak of motion is located. This can be seen as a lumped representation of the muscular actions and their influence on the facial tissue biophysics.

Many aspects of the technique require further improvement; for example, a criteria to determine the appropriate number of independent markers to be selected is needed. An important issue that must be also considered is that the this modeling approach is dependent on the data captured by the finite set of facial markers. Therefore, building a successful facial model would require researchers to cover a subject's face densely enough to capture all details of its kinematic behavior, or require researchers to place a smaller number of markers at optimal positions. Those and related issues are currently being considered as next research steps.

ACKNOWLEDGMENTS

This work has been supported by MCT/CNPq (Brazil), the National Institute of Deafness and Other Communications Disorders (Grant DC-05774), and NSERC (Canada). Some results were presented in partial form at the 7th International Seminar on Speech Production held in Ubatuba, Brazil, in December 2006, and are available in the online proceedings at http://www.cefala.org/issp2006/cdrom/main_index.html. We are grateful to Dr.A. R. Baigorri for useful discussions on statistical aspects of the algorithms and graduate student E. V. Fortes for his help with some of the numerical computations.

[1]That is the "thin" version of the factorization. In the full version $Q \in \mathbb{R}^{m\times m}$ and $R \in \mathbb{R}^{m\times n}$.

[2]The ϵ-rank of a matrix A is defined as $\min_{\|A-B\|\leq\epsilon}\text{rank}(B)$ and may be interpreted as the number of columns of A that are guaranteed to remain linearly independent for any perturbation to A with norm less or equal ϵ (Chan and Hansen, 1992).

Badin, P., Bailly, G., and Revéret, L. (**2002**). "Three-dimensional linear articulatory modeling of tongue, lips, and face, based on MRI and video images," J. Phonetics **30**, 533–553.

Beautemps, D., Badin, P., and Bailly, G. (**2001**). "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling," J. Acoust. Soc. Am. **109**, 2165–2180.

Bernstein, N. (**1967**). *The Co-ordination and Regulation of Movements* (Pergamon, Oxford).

Bischof, C. H., and Quintana-Ort, G. (**1998**). "Computing rank-revealing QR factorization of dense matrices," ACM Trans. Math. Softw. **24**, 226–253.

Björck, A. (**2004**). "The calculation of linear least squares problems," Acta Numerica **13**, 1–53.

Chan, T. F. (**1987**). "Rank revealing QR factorizations," Linear Algebr. Appl. **88/89**, 67–82.

Chan, T. F., and Hansen, P. C. (**1992**). "Some applications of the rank revealing QR factorization," SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **13**, 727–741.

Davis, H., and Silverman, S. R., eds. (**1970**). *Hearing and Deafness*, 3rd ed. (Holt, Rinehart and Winston, New York).

de Hoog, F. R., and Mattheij, R. M. M. (**2007**). "Subset selection for matrices," Linear Algebr. Appl. **422**, 349–359.

Ekman, P., Friesen, W. V., and Hager, J. C. (**2002**). *The Facial Action Coding System*, 2nd ed. (Research Nexus eBook, Salt Lake City).

Engwall, O., and Beskow, J. (**2003**). "Effects of corpus choice on statistical articulatory modeling," in Proceedings of the 6th International Seminar on Speech Production, edited by S. Palethorpe and M. Tabain, pp. 1–6.

Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. Learn. Mem. Cogn. **22**, 1166–1183.

Golub, G. H., and Loan, C. F. V. (**1996**). *Matrix Computations*, 3rd ed. (The Johns Hopkins University Press, Baltimore).

Kalra, P., Mangili, A., Thalmann, N. M., and Thalmann, D. (**1992**). "Simulation of facial muscle actions based on Rational Free Form Deformations," Comput. Graph. Forum **11**, 59–69.

Kanjilal, P., Dey, P. K., and Banerjee, D. N. (**1993**). "Reduced-size neural networks through singular value decomposition and subset selection," Electron. Lett. **29**, 1516–1518.

Knappmeyer, B., Thornton, I. M., and Blthoff, H. H. (**2003**). "The use of facial motion and facial form during the processing of identity," Vision Res. **43**, 1921–1936.

Kuratate, T., Munhall, K. G., Rubin, P. E., Vatikiotis-Bateson, E., and Yehia, H. (**1999**). "Audio-visual synthesis of talking faces from speech production correlates," in Proceedings of the 6th European Conference on Speech Communication and Technology (EuroSpeech99), Vol. **3**, pp. 1279–1282.

Kuratate, T., Yehia, H., and Vatikiotis-Bateson, E. (**1998**). "Kinematics-based synthesis of realistic talking faces," in *International Conference on Auditory-Visual Speech Processing (AVSP'98)*, edited by D. Burnham, J. Robert-Ribes, and E. Vatikiotis-Bateson (Causal Productions, Terrigal-Sydney, Australia), pp. 185–190.

Lawson, C. L., and Hanson, R. J. (**1987**). *Solving Least Squares Problems* (SIAM, Philadelphia).

Lee, Y., Terzopoulos, D., and Waters, K. (**1995**). "Realistic modeling for facial animation," Comput. Graph. **29**, 55–62.

Lorenzelli, F., Hansen, P. C., Chan, T. F., and Yao, K. (**1994**). "A systolic implementation of the Chan/Foster RRQR algorithm," IEEE Trans. Signal Process. **42**, 2205–2208.

Lucero, J. C., and Munhall, K. G. (**1999**). "A model of facial biomechanics for speech production," J. Acoust. Soc. Am. **106**, 2834–2842.

Lucero, J. C., Maciel, S. T. R., Johns, D. A., and Munhall, K. G. (**2005**). "Empirical modeling of human face kinematics during speech using motion clustering," J. Acoust. Soc. Am. **118**, 405–409.

Migliore, M. D. (**2006**). "On the role of the number of degrees of freedom of the field in MIMO channels," IEEE Trans. Antennas Propag. **54**, 620–628.

Müller, P., Kalberer, G. A., Proesmans, M., and Van Gool, L. (**2005**). "Realistic speech animation based on observed 3-D face dynamics," IEE Proc. Vision Image Signal Process. **152**, 491–500.

Munhall, K. G., and Vatikiotis-Bateson, E. (**1998**). "The moving face during speech communication," in *Hearing By Eye, Part 2: The Psychology of Speechreading and Audiovisual Speech*, edited by R. Campbell, B. Dodd, and D. Burnham (Taylor & Francis Psychology, London).

Parke, F., and Waters, K. (**1996**). *Computer Facial Animation* (AK Peters, Wellesley).

Pitermann, M., and Munhall, K. G. (**2001**). "An inverse dynamics approach to face animation," J. Acoust. Soc. Am. **110**, 1570–1580.

Prasad, S., and Chandna, B. (**1991**). "Direction-of-arrival estimation using rank revealing QR factorization," IEEE Trans. Signal Process. **39**, 1224–1229.

Ruckman, C. E., and Fuller, C. R. (**1995**). "Optimizing actuator locations in active noise control systems using subset selection," J. Sound Vib. **186**, 395–406.

Setnes, M., and Babuska, R. (**2001**). "Rule base reduction: some comments on the use of orthogonal transforms," IEEE Trans. Syst. Man Cybern., Part C Appl. Rev. **31**, 199–206.

Terzopoulos, D., and Waters, K. (**1990**). "Physically-based facial modeling, analysis, and animation," J. Visual. Comp. Animat. **1**, 73–80.

Troje, N. F. (**2002**). "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," J. Vision **2**, 371–387.

Vatikiotis-Bateson, E., and Yehia, H. (**1996**). "Physiological modeling of facial motion during speech," Trans. Tech. Com. Psycho. Physio. Acoust. **H-96-65**, 1–8.

Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories

W. F. L. Heeren[a] and M. E. H. Schouten
Utrecht Institute of Linguistics, OTS Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

(Received 21 May 2007; revised 25 April 2008; accepted 19 June 2008)

Listeners discriminate acoustic differences between phoneme categories at a higher level than similarly sized differences within phoneme categories. The question this paper aims to answer is *how* this pattern in perceptual sensitivity develops along an acoustic dimension that contrasts two non-native speech sounds: through acquired distinctiveness, through acquired similarity, or through a combination of the two. A pretest–training–post-test experiment was designed to study perceptual development directly, i.e., by including (i) a discrimination task to measure perceptual sensitivity, (ii) a transfer test to ensure language learning instead of stimulus learning, and (iii) a control group to exclude task repetition as an explanation of improvement. It is shown that the typical peak in perceptual sensitivity near a phoneme boundary that native listeners show is not found in relatively inexperienced language learners, despite their ability to classify a continuum in a nativelike way after short laboratory training. Experiment II indicates that a discrimination peak may be achieved by language learners, but only after much more language experience than short-term laboratory training can offer. Furthermore, reasons are given why classification improvement in the laboratory should not be taken as evidence for (i) increased discrimination of the newly learned phonemes and (ii) learning of phoneme representations.
[DOI: 10.1121/1.2967472]

PACS number(s): 43.71.Ft, 43.71.Hw, 43.71.An [PEI] Pages: 2291–2302

I. INTRODUCTION

Learning a language includes learning phoneme contrasts, i.e., learning the phonetic and acoustic features that define the differences between pairs of phonemes. This paper reports on experiments undertaken to answer the question of how phoneme contrasts are learned or, more precisely, how perceptual sensitivity develops along an acoustic dimension that contrasts two non-native speech sounds.

Listeners are highly sensitive to acoustic differences that distinguish phoneme contrasts in their native language. This ability was first described about 50 years ago in a study in which listeners classified and discriminated stimuli taken from an acoustic continuum between native phonemes, and were found to be better at discriminating stimuli belonging to different phoneme categories than stimuli belonging to the same category, with equal acoustic distances in both cases (Liberman *et al.*, 1957). This discontinuity in the perception of acoustically similar stimulus pairs has become known as categorical perception. Even though this concept means, strictly speaking, that listeners can only hear differences between speech sounds that are categorized differently, experimental results have never been in accordance with such a strict definition: Whether and to what extent speech sounds are perceived categorically are also influenced by, for example, the type of discrimination task (Gerrits and Schouten, 2004) and stimulus quality (Van Hessen and Schouten, 1999). Still, the tendency for listeners to discriminate between-category differences at a higher level than within-category differences has often been replicated [see, e.g., Liberman *et al.* (1961); Pisoni (1975); Strange and Dittman (1984); and Van Hessen and Schouten (1999)]. The question of how that effect arises is central to this paper.

Liberman *et al.* (1961) also wondered how the advantage of between-category discrimination over within-category discrimination develops: Is the discrimination peak near a native phoneme boundary innate or has it resulted from learning? To answer this question, the perception of a native phoneme continuum was compared to that of nonspeech control stimuli. The reasoning was that if discrimination of the nonspeech continuum shows the same peaks, the peak must be innate. However, if it does not, the peaks in the speech continuum must have been learned, and the level of discrimination in the nonspeech continuum must reflect the path of learning. Two learning hypotheses were forwarded [see, e.g., Gibson (1963)]. The first, *acquired distinctiveness* (AD), states that discrimination of unknown stimulus pairs is generally poor, but that the pair straddling the phoneme boundary becomes easier to discriminate as a result of learning. The second, *acquired similarity* (AS), states that discrimination is high before learning, but within-category sensitivity decreases with learning. If the nonspeech continuum shows chance-level discrimination, the peak in speech was learned through AD. If discrimination is high, however, the speech peak was obtained through AS. In accordance with the expectations, perception of the speech continuum showed a discrimination peak near the phoneme boundary. The non-

[a]Present address: Human Media Interaction group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands. Electronic mail: w.f.l.heeren@ewi.utwente.nl.

speech continuum, however, did not show such a peak—discrimination was low throughout. Hence, Liberman *et al.* (1961) concluded that the peak at the phoneme boundary is learned and that this is achieved through AD.

This study and other studies that made use of nonspeech control signals gave only indirect evidence for the way phoneme contrasts are learned [see, e.g., Miyawaki *et al.* (1975)]. The learning process was not studied, but the perception of native phoneme categories was compared to that of unknown nonspeech categories. Later on, other methods came to be used to study perceptual development of phoneme contrasts and cross-language speech perception. On the one hand, between-subjects designs were used in which groups of listeners with different language backgrounds or different amounts of language experience were compared. On the other hand, within-subjects designs were employed to investigate development in the course of listeners' language learning. The contribution this latter type of research has made to the study of perceptual development is the pursuit of direct evidence for development in perceptual sensitivity during the learning of phoneme contrasts. For this, within-subjects designs have to be preferred. Such designs typically consist of a baseline measurement, a training period, and post-training measurements that are compared to the baseline. We will discuss earlier within-subjects designs, their shortcomings, and the efforts the present research make to overcome them. Before that, however, results from between-subjects studies will be discussed, since these provide support for both learning hypotheses.

Many investigations into cross-language perception have compared listeners with different language backgrounds and varying degrees of language experience. Classical examples are Lisker and Abramson (1970), who compared perception along a dimension that was divided into three phoneme categories by Thai listeners, but into only two by English listeners, and Miyawaki *et al.* (1975), who chose a stimulus dimension that was phonemic to only one of the two listener groups. Studies that compared perception between groups differing in experience with a particular language showed that discrimination peaks near a non-native phoneme boundary were higher in advanced than in beginning learners, but not as high as in native listeners [see, e.g., MacKain *et al.* (1981)]. These between-subjects designs thus yield evidence in favor of learning through AD.

A study that gives evidence consistent with the AS hypothesis is Best *et al.* (2003), but Iverson *et al.* (2003) suggest a combination of AD and AS. AS has also been associated with perceptual development in infants (Pisoni, 1991), who seem to detect any contrast in the first six months of their lives but lose this ability when they are around 12 months old. Best *et al.* (2003) examined cross-language perception of !Xóõ click contrasts and found that adult American English listeners performed at a level comparable to native listeners. On one particular click contrast, the English listeners even outperformed the click language users; this implies that their perceptual sensitivity may only become nativelike by a decrease in their ability to discriminate the contrast, i.e., by AS. It is possible, however, that the similarity in the discrimination results was actually the outcome of different processing strategies, since the English listeners often reported hearing nonspeech sounds when clicks were presented, whereas native listeners heard speech. A cross-language perception study whose results hint at learning through a combination of within-category AS and between-category AD is Iverson *et al.* (2003). Listeners from different language backgrounds were tested on a /ra-la/ continuum varying along the *F*3-dimension. Within-category discrimination in Japanese listeners—for whom the contrast is nonphonemic—was higher than that of native listeners within both the /r/ and the /l/ category. American listeners had higher discrimination scores near the phoneme boundary. Japanese listeners could only obtain nativelike perception along the *F*3-dimension by within-category decreases and a between-category increase in perceptual sensitivity, i.e., by a combination of AS and AD. The actual learning process, however, cannot be shown without the use of within-subjects designs.

Early studies that used laboratory training to change the perception of non-native phoneme contrasts were not very successful [see Strange and Jenkins (1978)]. Later studies, however, have shown that such phoneme contrasts can be learned through relatively short laboratory training [see Strange (1995)]; their success was due to improved training paradigms [see, e.g., Jamieson and Morosan (1986) and Lively *et al.* (1993)]. Up to now, however, perceptual development of non-native phoneme contrasts has more often been studied in classification than in discrimination designs. Research has shown that adult learners generally improve their ability to classify non-native phoneme categories as a result of training [see, e.g., Jamieson and Morosan (1986); Logan *et al.* (1991); and Flege (1995)], even though learning of some contrasts was more successful than learning of others [see Strange and Jenkins (1978)]. Furthermore, categorization training has resulted in generalization to novel stimuli [see, e.g., Logan *et al.* (1991) and Strange and Dittman (1984)], long-term retention of trained stimuli [see, e.g., Flege (1995)] and transfer to production [see, e.g., Bradlow *et al.* (1997)]. When listeners became better at identifying a speech sound as belonging to a certain phoneme category, this was interpreted as an increased ability to perceive the difference between the categories or as an increase in the discrimination of between-category differences. We will show that this assumption may not always be justified.

Only a few training studies monitored changes in perceptual sensitivity along the acoustical dimension to be learned by including a discrimination task (Strange and Dittman, 1984; Jamieson and Morosan, 1986; Van Hessen, 1992; Tremblay *et al.*, 1997). Together, these studies indicate that training a non-native phoneme contrast results in increased perceptual sensitivity near the newly learned phoneme boundary. The discrimination peak, however, remains lower than in native listeners (Strange and Dittman, 1984).

A shortcoming in the design of earlier within-subjects studies, however, is the fact that only one training continuum was used. This raises the possibility that listeners learn only that particular stimulus set, instead of a language-relevant distinction. Lively *et al.* (1993) showed that variability in the training materials helps category learning and that listeners

who were trained on a single speaker did not show any transfer (generalization) of the categories learned to new instances of those phonemes. This would explain the lack of transfer found in Strange and Dittman (1984). Jamieson and Morosan (1986) and Van Hessen (1992, Chap. 7) did not address transfer at all: Learners were able to improve their discrimination of the phoneme boundary region in the stimulus set they were exposed to, but possibly not beyond that. Hence, in order to demonstrate language learning, and not learning of a particular stimulus continuum, testing for transfer is imperative. In the present study this problem is addressed by using a speaker for testing who does not belong to the set of speakers used for training.

A third important design issue is to make sure that changes found between pretest and post-test are actually due to language learning and are not the effect of task repetition or of a reduction on task uncertainty. This can be guaranteed by including a control group, which has not always been the case in previous work.

As has been argued above, the question of whether perceptual learning along an acoustic dimension occurs through AD, through AS, or through a combination of the two has never been addressed *directly* with respect to the learning of phoneme contrasts. There is some indirect evidence from speech perception studies, but it is only with respect to the development of visual categorization and the perception of auditory nonspeech stimuli that our hypotheses were actually investigated. When more than one category was learned, a category-relevant dimension acquired distinctiveness (Goldstone, 1994). Within learned categories, however, AS was found [see, e.g., Guenther *et al.* (1999)]. Moreover, these experiments revealed that changes in perceptual sensitivity were relatively fast: 45–60 min of training were sufficient to induce changes in perception. Combining this with the conclusion from later phoneme training studies that non-native phoneme contrasts can be learned relatively quickly [see, e.g., Lively *et al.* (1993)], we expected that it would be feasible to monitor the development of perceptual sensitivity along a dimension that contrasts newly learned phonemes in a within-subjects design. In this paper, the question of how the phoneme boundary is learnt was operationalized by addressing Dutch learners' development of the Finnish length contrast /t-tː/.

II. EXPERIMENT I

The first goal of this experiment was to evaluate the main research question of how a non-native phoneme contrast is learned. In an earlier experiment we had found that perceptual development as a result of phoneme training was minimal, although not inconsistent with learning of the phoneme boundary through between-category AD (Heeren, 2006, Chap. 2). This lack of a clear effect could have been caused by minimal development during the experiment or by an unsuitable discrimination task. The second goal of the present experiment was therefore to find an answer to the question of why the increase in perceptual sensitivity near the learned phoneme boundary had remained so small in the earlier study. These goals were pursued by running pretest–training–post-test experiments in which Dutch listeners learned the Finnish length contrast /t-tː/. The choice of a durational contrast in laboratory training was motivated by the fact that duration is regarded as a salient cue for non-native listeners as opposed to, for example, spectral information (Bohn, 1995), which should positively affect learning speed.

In quantity languages, such as Finnish, length is phonologically contrastive and is mainly cued by segment duration. Dutch is not a quantity language. It does have vowels that contrast in length, but these are also cued by spectral differences, such as "tak-taak" (/tak-tɑːk/), meaning "branch-task." Listeners normally use both spectral and durational information to identify these vowels [see Nooteboom and Cohen (1995)]. As for consonants, Dutch has no short-long contrasts. In Finnish, on the other hand, most consonants and all vowels occur either in singleton or in geminate form, signaling different word meanings such as "tuli-tuuli" (/tuli-tuːli/), meaning "fire-wind."

How do Dutch adults learn to perceive the Finnish /t-tː/ contrast? Through AD, AS, or a combination of the two? Earlier studies on the perception of *L*2 length contrasts have shown that listeners whose native language does not use quantity, such as the Dutch, can learn to perceive and produce such contrasts [see, e.g., McAllister *et al.* (2000) and Tajima *et al.* (2003)]. McAllister *et al.* (2000) studied the influence of language background on the acquisition of a Swedish quantity contrast by English, Spanish, and Estonian learners. They predicted that learning success depended on the role of quantity in the native languages and this was borne out. Speakers of Spanish, whose first language does not have length differences, had great difficulty perceiving and producing the Swedish vowel contrast. Estonian speakers, on the other hand, whose native language does have quantity, did very well; English speakers, whose native language employs some durational differences, performed in between. Taijima *et al.* (2003) compared the perception of Japanese length contrasts between native and non-native English listeners. They found that English listeners were less accurate in identifying stimuli than Japanese listeners, but that trained English listeners performed better than untrained ones. A similar pattern of results was found for advanced Russian second-language users of Finnish, as opposed to nonexperienced Russian listeners (Ylinen *et al.*, 2005). The trained listeners in Taijima *et al.* (2003), however, did not become as accurate as natives in the identification of stimuli after 15 240-trial training sessions spread over five days. Still, some transfer from the trained vowel contrasts to other length contrasts was found.

Brain imaging studies have shown that neural responses to quantity contrasts in *L*2 listeners differ from responses in native listeners [see, e.g., Gandour *et al.* (2002) and Nenonen *et al.* (2003)]. For example, Nenonen *et al.* (2003) compared the detection of preattentive changes in auditory stimuli between Finnish listeners and advanced second-language users of Finnish. They showed that processing of *speech* stimuli differing in duration varied between the listener groups, whereas this was not the case for *nonspeech* stimuli differing in duration. Moreover, in a cross-language discrimination task, Ylinen *et al.* (2005) found that Russian listeners who

had on average lived in Finland for two years, but who judged themselves relatively inexperienced users of the Finnish language, did not show a discrimination peak near the phoneme boundary of a vowel quantity contrast and showed lower discrimination performance than natives.

On the basis of the earlier research, we expect focused training to result in development of Dutch learners' perceptual sensitivity to the Finnish durational contrast according to AD at the phoneme boundary. As for the question of why, in our earlier study (Heeren, 2006, Chap. 2), the change in perceptual sensitivity remained so small after training with a non-native phoneme contrast, two answers were advanced: Either because the discrimination tasks used did not tap into phonemic processing or because perceptual sensitivity changed only minimally with learning. These possibilities were explored by using the two-interval forced-choice (2IFC) discrimination task [see, e.g., Macmillan *et al.* (1977)], which is known to reflect phonemic processing [see, e.g., Gerrits and Schouten (2004)], instead of the four-interval same-different (4IAX) discrimination task used earlier [see, e.g., Pisoni (1975)]. If the change of task reveals a discrimination peak after training, this will be evidence for the first explanation. However, if no discrimination peak is found after training, the second explanation will be supported.

A. Method

A pretest–training–post-test design was used to monitor development. Different phoneme continua were made by using speech from various speakers, i.e., five training speakers and one test speaker. The use of different voices is thought to force listeners to abstract from individual speaker differences during learning. On the other hand, listeners have more difficulty processing stimuli from several voices than stimuli from a single voice in their native language (Martin *et al.*, 1989), as well as in a non-native language (Bradlow and Pisoni, 1999).

For pretest and post-test, 2IFC discrimination and classification tasks were used. The discrimination test was administered with one-step and two-step stimulus distances. The different step sizes were included to prevent any floor or ceiling effects, in order to give both hypotheses an equal chance. If discrimination performance *improves* as a function of training, this may show up primarily in the one-step discrimination task. As the small differences in this task are generally difficult to discriminate, there will be enough room on the scale to register an increase in discrimination performance as a result of training, and a ceiling effect will be avoided. On the other hand, if discrimination performance *falls* as a result of training, this may most clearly show up in reduced percentages of correct responses to two-step differences, where the risk of a floor effect is smaller. A pilot study showed that the one-step 2IFC task reflected increased perceptual sensitivity near the phoneme boundary in native Finnish listeners; see Heeren (2006).

For the training phase, the classification paradigm was chosen, because it stresses the presence of the two categories that need to be learned. Feedback was provided on the correctness of the listeners' responses; they did not receive explicit feedback on what to attend to.

1. Materials

In order to produce phoneme continua from six different speakers, three steps were taken: (1) Phoneme contrasts spoken by different Finnish natives were recorded, (2) the phoneme contrasts were turned into continua, and (3) the resulting continua were checked by presenting them to native Finnish listeners.

In the recording phase, nonsense words were used as targets in order to avoid word frequency effects or lexical bias [see Ganong (1980)]. A pair of Finnish pseudowords *ata-atta* (/ɑtɑ-ɑtːɑ/) was recorded, spoken by eight native speakers of Finnish (three males). Their ages varied from 20 to 56 years. Speech from three speakers was recorded in a sound-treated booth at the Utrecht Institute of Linguistics OTS using a Sennheiser ME-64 microphone and a Grundig DAT-9009 recorder, with a sample frequency of 48 kHz. Recordings were downsampled to 22.05 kHz. Speech from the other five speakers was recorded in a silent room at a primary school they were visiting as part of an exchange program. These recordings were made directly onto a Toshiba Satellite 1400 153E laptop computer with an ALi Audio Wave sound card using a Trust PC microphone.

Before the recordings were manipulated into continua, the effects of word form, /ɑtɑ/ vs /ɑtːɑ/, on the duration of each of the segments and on the vowels' formant frequencies were examined. First, the durations of the consonant segments and the vowels first and second formants revealed that—as expected—the duration of /tː/ was longer than that of /t/ [$t(7)=-8.9$, $p<0.001$], (mean=330 ms versus mean =123 ms), but also that the duration of $/V_2/$ in *atta* was significantly shorter than that of $/V_2/$ in *ata* [$t(7)=4.0$, $p=0.005$] (mean=100 ms versus mean=140 ms). A similar effect was obtained by Lehtonen (1970, p. 111). The vowels' formants did not vary as a function of consonant length.

Stimulus continua were made by progressively shortening the closure duration of the stop consonant in each speaker's pronunciation of /ɑtːɑ/. The difference in vowel duration as a function of consonant length had first been neutralized to eliminate it as perceptual cue. This was accomplished either by cutting whole periods from the middle of a speaker's final vowel in /ɑtːɑ/ or by copying and adding one or more whole periods taken from a speaker's final vowel in /ɑtːɑ/ to that same vowel, in order to approximate the mean duration between the final vowels in /ɑtɑ/ and /ɑtːɑ/ as closely as possible. Next, stop closure duration of each /tː/ was linearly shortened to that of the corresponding /t/. The size of the durational steps differed per speaker. Eight /ɑtɑ-ɑtːɑ/ continua from different speakers were obtained, each consisting of seven stimuli.

In order to determine the location of the phoneme boundary in each of the phoneme continua, a classification experiment was run with native listeners of Finnish. Nine female speakers of Finnish living in The Netherlands participated. Their ages varied from 24 to 54 years, with a mean age of 38. All participants were born in Finland and had grown up there. Their length of stay in The Netherlands var-

ied from 6 months to 30 years. They all regularly used the Finnish language, e.g., during phone conversations or reading. Participants were either tested at their homes or in the experimenter's office under quiet circumstances. Each stimulus was presented ten times. One listener classified seven phoneme continua out of eight, since she had also participated in the recording session and did not classify her own speech. The other listeners classified the eight phoneme continua from all speakers.

Listeners' responses were expressed as a percentage of one response category per stimulus and should form an *S*-shaped curve as a function of stimulus step. In order to check this shape, the responses were transformed to z-scores, which should result in a straight line. The z-scores were submitted to a linear regression analysis per listener, with stimulus step as an independent variable. If a linear fit of a listener's data was not possible, those data were excluded from further analysis. Phoneme boundaries and boundary widths were derived from the regression fits. The boundary was fixed at $z=0$, the 50% point, for each listener. Boundary widths were determined by subtracting the stimulus value corresponding to the 25% point ($z=-0.674$) from that corresponding to the 75% point ($z=0.674$).

Since six speakers were needed for further training experiments, the phoneme continua from the two speakers that were most difficult for the native listeners, i.e., those with the largest relative boundary widths, were not used any further. The continuum from one speaker, a 24-year-old female (recorded in the sound-treated booth), was selected as test material. The others were assigned to be training speakers.

2. Participants

There were 28 participants (4 males) with a mean age of 23 (range 19–32). All were native speakers of Dutch and reported no hearing disorders when asked. Half of them participated in the full experimental procedure, and the other half formed the control group and only participated in pretest and post-test, with a time interval of approximately one week. This roughly corresponded to the time interval trained listeners had between pretest and post-test.

3. Procedure

All experiments were run in a quiet room at the Utrecht Institute of Linguistics OTS. A laptop computer was used both to present the stimuli in a random order and to register responses. Stimuli were presented over Beyerdynamic DT 770 closed-ear headphones at a comfortable listening level. Listeners were tested individually. The pretest was completed on the first day. On subsequent days, training sessions were run until a listener classified the new phoneme contrast correctly in at least 90% of the trials in two consecutive training tests. During each training session, one to four training tests of 480 trials each were run. On the final day, one day after the last training session, the post-test took place.

Listeners received written instructions. In pretest and post-test participants first classified the continuum by labeling each stimulus as containing a "short *t*" or a "long *t*." The Finnish written forms, *ata* and *atta*, were not shown, because Dutch listeners would have expected different vowels on the basis of the spelling. In Dutch, letter *a* is produced as /aː/ in an open syllable, as in *ata*, but as /ɑ/ before the coda consonant in a closed syllable, as in *atta* (Booij, 1995). As we showed earlier, in Finnish, vowel quality does not depend on whether the vowel occurs before a singleton or a geminate consonant.

The test was preceded by a short introduction consisting of two presentations of each of the continuum's end points. The test itself contained ten repetitions of each of the test speaker's seven stimuli, included one short break halfway, and lasted about 4 min. Stimulus order was randomized. Next, 2IFC discrimination with one-step and two-step pairs was administered. Two stimuli from the continuum were presented in a pair with an interstimulus interval (ISI) of 200 ms, and the participant was asked to determine the order in which the stimuli were presented: "shorter-longer" or "longer-shorter." The ISI was set at 200 ms, since it allows listeners to compare both acoustic traces and phoneme labels [see, e.g., Massaro (1974) and Pisoni and Tash (1974)]. If phoneme categories have not yet been learned, the listener can still compare traces, which makes the task less difficult. If phoneme categories are present, they will influence the listener's perception, as was found in our pilot study with native Finnish listeners. As an introduction, four examples of six-step stimulus pairs, consisting of the continuum's end points, were presented. Next, participants got practice with six one-step and five two-step pairs. After that, the actual test started, consisting of 20 repetitions per stimulus pair (10 in each order). Presentation order was randomized. The test included three breaks at regular intervals and lasted between 15 and 20 min.

Listeners received training until they reached a score of at least 90% correct responses in two consecutive tests. They were given written instructions and trial-by-trial feedback on the correctness of their responses. There were 350 trials per training test (5 speakers $\times$ 7 stimuli $\times$ 10 repetitions). Three breaks were given at regular intervals and during each break the percentage of correct responses so far was shown. Before a session started, the continuum's end points, /ɑtɑ/ and /ɑtːɑ/, from each of the training speakers were presented once as an introduction. The post-test was always administered one day after the last training session in order to test the effects of language learning instead of short-term auditory memory.

B. Analysis and results

Pretest and post-test classification responses, as well as training results, were represented as percentages of /tː/ answers to each of the duration steps per listener, and subsequently transformed to z-scores that were subjected to linear regression analyses. Next, the phoneme boundaries and boundary widths of the classification functions were derived. The 2IFC discrimination results were given in percentages of correct responses to each of the stimulus pairs in one-step and two-step discriminations, separately.

1. Training

To reach the training goal, listeners needed between 700 and 3150 training trials (mean=1625 trials), i.e., between 2

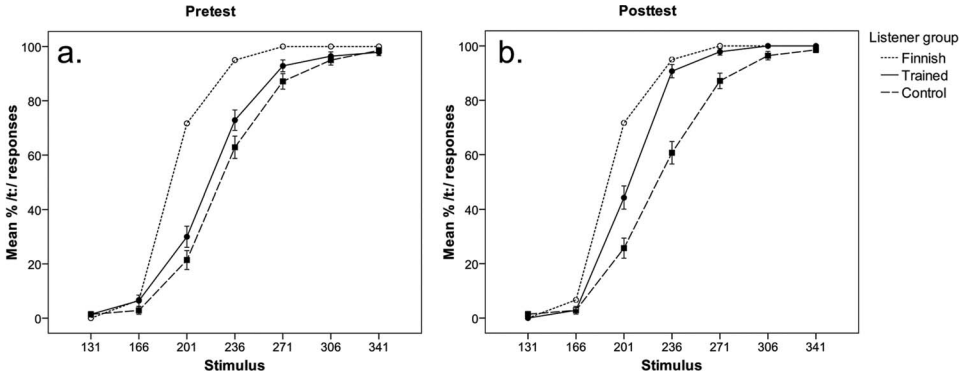


FIG. 1. Classification results of trained and control listeners in pretest (a) and post-test (b), with the pilot results from the Finnish listeners as a reference. Along the horizontal axis the stimuli from /t/-like to /t:/-like speech sounds are indicated, whereas the vertical axis shows the percentage of /t:/-responses. The error bars indicate one standard error of the mean.

and 9 tests. The average percentage of correct responses was already quite high in the pretest (88%). This is assumed to be due to the fact that Dutch listeners can classify the continuum into short and long stimuli without necessarily referring to phoneme categories. Between pretest and post-test, the mean percentage of correct responses increased to 91% [$t(13)=-3.2$, $p=0.008$]. For each training speaker, the location of the boundary and its width were compared between the first and the last training session in t-tests for paired samples. These analyses showed that the boundary location changed only for one speaker ($p=0.001$). Boundary widths became significantly smaller for three speakers ($p<0.02$) and marginally smaller for a fourth one ($p=0.063$). In sum, listeners became more consistent with training.

2. Classification of the /t-t:/ test continuum

Figure 1 shows the mean classification functions from pretest and post-test for both trained and control listeners. Along the horizontal axis, the duration steps along the continuum are given, and the vertical axis gives the percentages of /t:/ responses. Furthermore, the pilot results from the Finnish listeners are shown as a reference. Repeated-measures analyses were run on the arcsine transformed (Studebaker, 1985) proportions of /t:/ responses to assess the effect of training on classification behavior. Within-subjects variables were test (pretest versus post-test) and duration step (131–341 ms); training (trained versus control) was included as a between-subjects variable.

The mean percentages of /t:/ responses did not show a main effect of test ($p=0.111$), but a duration step × training interaction was present [$F(5.1,131.8)=5.0$, $p<0.001$]. Across testing moments, the listener groups' responses differed from each other. Moreover, a main effect of duration step was found [$F(5.1,131.8)=944.0$, $p<0.001$]. The between-subjects effect of training was also significant [$F(1,26)=12.7$, $p=0.001$]. This last effect indicates that trained listeners differed from control listeners. The exact difference was studied further by means of paired samples *t*-tests on each listener group. These analyses showed that trained listeners' perception of the 236 ms stimulus had changed significantly from pretest to post-test ($p=0.018$), i.e., they more consistently perceived it as a realization of /t:/, whereas no changes were found for the controls.

The mean phoneme boundary of the trained Dutch listeners shifted in the direction of the Finnish boundary as a result of training. The Dutch, however, did not seem to fully transfer the boundary learned from the training speakers to the test speaker, as a difference between the Dutch and the Finnish listeners' phoneme boundaries remained (215 versus 203 ms). The location of the boundary and its width, as determined from the regression fits of the *z*-transformed proportions of /t:/ responses, were subjected to repeated-measure analyses with within-subjects factor test (pretest versus post-test) and between-subjects factor training (trained versus control). The mean phoneme boundaries showed a main effect of training [$F(1,26)=7.2$, $p=0.013$]. On average, trained listeners put their phoneme boundary closer to that of the Finnish. Before training, the trained listeners' phoneme boundary did not differ from that of the controls ($p=0.168$). The trained group's mean phoneme boundary shifted from 223 to 215 ms, i.e., toward the Finnish phoneme boundary, while the control group's boundary remained at the same location (i.e., at 233 ms). In the post-test, trained listeners' phoneme boundaries differed significantly from those of the controls, as a *t*-test showed [$t(26)=-3.1$, $p=0.005$].

The repeated-measures analysis with boundary width as a dependent variable showed a test × training interaction [$F(1,26)=4.9$, $p=0.035$] and a main effect of test [$F(1,26)=5.2$, $p=0.031$]. The trained group's mean boundary width decreased from 39 ms in the pretest to 33 ms in the post-test, whereas the control group's width remained the same throughout both tests (38 ms). A *t*-test showed that, in the post-test, the trained group's boundary width was significantly smaller than that of the controls [$t(26)=-2.9$, $p=0.008$].

3. One-step 2IFC discrimination

Section II B 2 has shown that trained listeners' classification of a speech continuum that had not been included in their training improved from pretest to post-test. Moreover, for control listeners, this was not the case. In this section and in Sec. II B 4 it will be examined whether discrimination of the test materials also changed with learning. Remember that the trained listeners put their post-test boundary at 215 ms and that controls put it at 233 ms.

The arcsine transformed proportions of correct answers per stimulus pair were subjected to repeated-measures analyses of variance (ANOVAs) with test (pretest versus post-test) and duration pair (6) as within-subjects factors and training (trained versus control) as between-subjects factor.

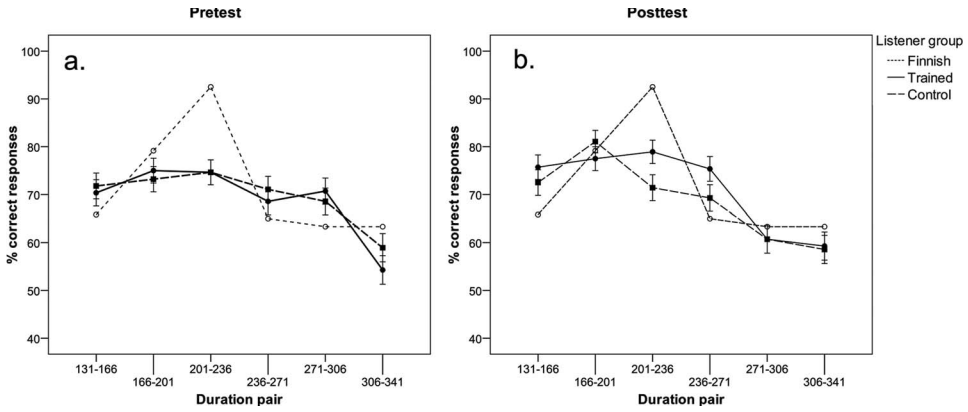


FIG. 2. One-step 2IFC discrimination results for both the trained and the control listeners in pretest (a) and post-test (b), with the Finnish results as a reference. Along the horizontal axis the stimulus pairs from /t/-like to /tː/-like speech sounds are indicated, whereas the vertical axis shows the percentages of correct discrimination responses. The error bars indicate one standard error of the mean.

The pretest and post-test results for both trained and control listeners are shown in Fig. 2. Along the horizontal axes the stimulus pairs are given, and along the vertical axes the percentages of correct responses are given. Furthermore, the Finnish norm, as established in the pilot test, is given as a reference. For one-step 2IFC discrimination, a test $\times$duration pair interaction [$F(5,130)=4.3$, $p=0.001$] and a main effect of duration pair were found [$F(4.0,102.3)=20.5$, $p<0.001$]. Not all stimulus pairs were equally well discriminated, and, as a function of testing moment, discrimination levels shifted. Furthermore, the trained listeners' mean percentage of correct discrimination increased somewhat from pretest to post-test (69%–71%), whereas for control listeners a minimal decrease was found (70%–69%). There was, however, no effect of training or a test $\times$training interaction, which means that trained listeners' perceptual sensitivity along the continuum did not show large changes compared with the controls' sensitivity.

It was expected that the trained listeners would develop heightened perceptual sensitivity, reflected by a discrimination peak around the location of their phoneme boundary, i.e., around stimulus pair 201–236. Therefore, within-category and between-category discrimination performance were examined further within and between listener groups in posthoc analyses. *t*-tests for paired samples showed that discrimination of pair 271–306 decreased ($p<0.01$) for both groups, and that discrimination of pair 166–201 increased for control listeners ($p=0.023$). The latter increase did not, however, co-occur with their perceived phoneme boundary at 233 ms.

Both listener groups had positioned their perceived phoneme boundary in such a way that increased sensitivity to pair 201–236 would be expected. *t*-tests showed that, in the post-test, trained listeners performed better on the between-category pair than control listeners did [$t(26)=2.1$, $p=0.044$], whereas there had been no difference in the pretest. Within-category pairs did not show differences between the listener groups.

4. Two-step 2IFC discrimination

The arcsine transformed proportions of correct answers per stimulus pair were subjected to repeated-measures ANOVAs with test (pretest versus post-test) and duration pair (5) as within-subjects factors, and training (trained versus control) as between-subjects factor. This revealed the main effects of test and of duration pair [$F(1,26)=12.6$, $p=0.002$ and $F(3.5,91.4)=39.1$, $p<0.001$]. Listeners' mean percentages of correct responses increased from pretest to post-test (trained: 84%–88%, controls: 84%–86%). The increase was significant for trained listeners ($p=0.006$) but not for untrained listeners ($p=0.076$). Paired-samples *t*-tests showed that trained listeners' sensitivity increased for within-category pair 131–201 [$t(13)=-2.7$, $p=0.018$] and between-category pair 201–271 [$t(13)=-3.0$, $p=0.010$]. Moreover, a three-way interaction of test$\times$duration pair$\times$training was present [$F(4,104)=2.9$, $p=0.027$], revealing that training had an effect on the listener groups' responses.

If within-category perceptual sensitivity had changed according to AS, i.e., fallen in regions that were relatively uninformative, mean percentages correct in the two-step discrimination experiment would have been lower, especially in trained listeners. This effect was not found. On the contrary, the percentages of correct responses increased.

C. Discussion

The present experiment was designed to study how perception develops in Dutch adults as they learn the Finnish quantity contrast /t-tː/. It was expected that Dutch listeners would learn to perceive the phoneme boundary in a new phoneme contrast through between-category AD. Additionally, if the changes in perceptual sensitivity obtained in our earlier work were as small as they were due to a discrimination task that did not sufficiently reflect phonemic processing, the use of 2IFC discrimination should have revealed a discrimination peak after training. However, since this did not happen, the conclusion can only be that our training did not induce a peak.

The classification results showed that trained listeners partially transferred the newly learned phoneme boundary from the training speakers to the test speaker by a boundary shift in the direction of the Finnish boundary. These findings are in line with other results on quantity learning [see, e.g., Tajima *et al.* (2003) and Ylinen *et al.* (2005)] and learning of non-native phoneme contrasts in general [see, e.g., Flege (1989); Jamieson and Morosan (1986); and Lively *et al.* (1993)]. Moreover, the trained listeners' boundary widths decreased, reflecting an increase in their classification consistency, which was not found for the control listeners.

Classification training has improved non-native listeners' classification of the test continuum. If perceptual sensitivity along that continuum had also developed as a result of training, this would have shown up in the 2IFC discrimination task. One-step discrimination results, however, hardly changed toward the native sensitivity pattern: The trained listeners did not show a peak in perceptual sensitivity near the newly learned phoneme boundary. Still, in the post-test, trained listeners gave more correct responses to the stimulus pair straddling the phoneme boundary than control listeners did. Both listener groups' sensitivity to duration pair 271–306, within the /t:/-category, decreased. Since this effect of within-category AS was observed in both listener groups, it may not have been due to learning. Moreover, in two-step discrimination, percentages of correct responses increased significantly for the trained listeners but not for the controls.

These results provide little direct evidence in favor of the development of the phoneme boundary through AD, since the changes in discrimination behavior were small, as had also been found in our earlier experiment. Indirectly, however, the evidence did support learning of the boundary through between-category AD, as trained listeners outperformed controls at the between-category stimulus pair in the post-test. Moreover, since several within-category stimulus pairs seemed to be discriminated better by the Dutch than by the Finnish, some within-category AS may also be indicated.

As a peak in discrimination behavior did not arise in the 2IFC discrimination task, heightened perceptual sensitivity was truly minimal. Apparently, Dutch adults' sensitivity along the /t-t:/ continuum hardly changes with phoneme training. Therefore, the conclusion is that the results of Heeren, (2006, Chap. 2) were caused by genuinely minimal learning. This contrasts with the relative ease with which identification/classification behavior changed toward a more nativelike pattern.

This finding raises the question of whether second-language learners can ever acquire a nativelike sensitivity pattern. It is possible that much more experience is necessary to reliably change perceptual sensitivity. In experiment I, learners participated in short-term phoneme training. Even though such training paradigms can successfully alter classification behavior, as has been repeatedly demonstrated [see, e.g., Jamieson and Morosan (1986) and Pisoni *et al.* (1982)], and can also cause changes in perceptual sensitivity along visual dimensions [see, e.g., Guenther *et al.* (1999)], they may be too short to induce nativelike perceptual sensitivity patterns in language learners [see also Strange and Dittman (1984)]. More experience with the second language may be required. Alternatively, it has been suggested that nativelike perception of a nonnative continuum cannot be achieved at all (Takagi, 2002).

III. EXPERIMENT II

Short-term perceptual training seems to be insufficient to induce nativelike discrimination behavior. The question experiment II examines is whether non-native language learners can achieve more nativelike perceptual sensitivity with much greater exposure. We will argue that advanced students' perceptual sensitivity along the /t-t:/ continuum increasingly resembles that of native Finnish listeners.

A. Method

Since experiment I indicated that development of perceptual sensitivity to nativelike levels may be a long process, a longitudinal design would have been a very time-consuming way of testing the question raised in the present experiment. We therefore chose to monitor development in a between-subjects design, using the same procedure as in the pretest of experiment I. This time, however, the perception of Dutch students of Finnish was compared to that of native Finnish listeners.

In The Netherlands, only a handful of students major in the Finnish language, and they do so at the University of Groningen. About half of the student population responded to the call for subjects for this perception experiment. Eight students of Finnish, who had Dutch as their native language, participated. Their mean age was 21 years, ranging from 18 to 26. There were two third-year students, three second-year students, and three first-year students. At the time of testing, the first-year students had received about five months of education in Finnish language and culture, which included courses on Finnish grammar, writing, and speaking. The second-year-students had followed courses on grammar, writing, and analyzing texts, and had furthermore received training in speaking and, most importantly, listening. Moreover, they had all spent some time in Finland. The third-year students had also visited Finland and they were of course the most advanced ones.

The listening tests were run in a quiet, but not sound-treated, testing room at the University of Groningen. The equipment was the same as that used in experiment I. The testing procedure was exactly the same as in the pretest of the previous experiment: Participants first completed the classification of the female test speaker's *ata-atta* continuum, followed by one-step and two-step 2IFC discriminations of the same continuum. For comparison, the data from the Finnish listeners obtained in experiment I's pilot were used. In addition, the students completed a questionnaire on their knowledge of and experience with Finnish and other languages. This was done after the listening tests. The entire procedure was completed within 45 min and participants received a 10-euro book-token for their efforts.

B. Results

1. Classification

Figure 3 shows the classification functions of the students compared to that of the Finnish. To determine the effect of language experience on the perception of the *ata-atta* continuum, the listeners' arcsine transformed responses to each of the stimuli were subjected to an ANOVA with stimulus duration (131–341 ms) and nativeness (Dutch versus Finnish) as between-subjects variables.

The analysis revealed a nativeness × stimulus duration interaction [$F(6,84)=5.3$, $p<0.001$]. Moreover, the main effects of nativeness [$F(6,84)=206.8$, $p<0.001$] and stimulus duration on the mean responses were found [$F(1,84)=26.3$,

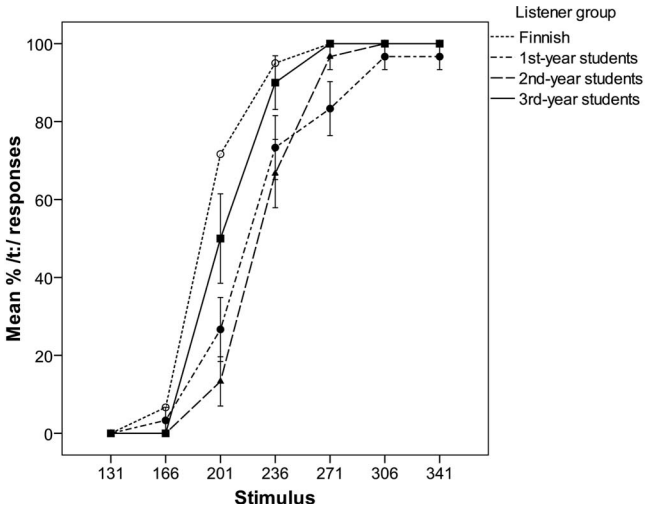


FIG. 3. Classification results from each of the student levels and the Finnish native listeners. Along the horizontal axis the stimuli from /t/-like to /t:/-like speech sounds are indicated, whereas the vertical axis shows the percentage of /t:/-responses. The error bars indicate one standard error of the mean.

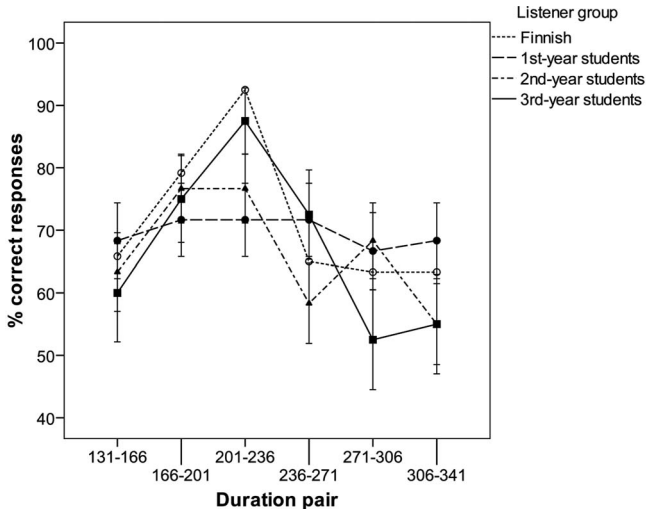


FIG. 4. One-step 2IFC discrimination results for the three levels of students (first, second, and third year) and the Finnish natives. Along the horizontal axis the stimulus pairs from /t/-like to /t:/-like speech sounds are indicated, whereas the vertical axis shows the percentages of correct discrimination responses. The error bars indicate one standard error of the mean.

$p<0.001$]. These findings indicate that differences between relatively advanced learners and natives still exist, but Fig. 3 shows that differences tend to decrease with increasing language experience.

To further support this interpretation, Table I gives the individual students' phoneme boundaries and boundary widths. The location of the phoneme boundary in the Finnish control listeners varied from 190 to 222 ms, with a mean of 203 ms. Only the third-year students put their phoneme boundary within the Finnish range (217 and 221 ms). Less-experienced listeners placed their boundary at longer closure durations, i.e., around 230 ms. Moreover, the learners' boundary widths tended to decrease with language experience: Both the second-year and the third-year students' boundary widths fell near that of the Finnish (at 35 ms).

Remember that the trained adult learners from experiment I put their mean post-test phoneme boundary at 215 ms, whereas their pretest boundary had been at 223 ms, and that they had a post-test boundary width of 33 ms. The phoneme boundary of the untrained listeners lay at 238 ms and did not change. It is remarkable that the degree of classification improvement from first-year to third-year students roughly coincides with the improvement attained by naive Dutch listeners after only a few hours of classification training. The type and amount of experience inducing these two boundary shifts, however, differ greatly. The naive learners had only been trained on the *ata-atta* word pair spoken by a few different speakers for about 1500 trials. The advanced students had encountered the Finnish singleton-geminate contrast not only in the /t-t:/ consonant pair but also in the eight vowels and ten other consonant pairs that Finnish employs in different contexts spoken by many more speakers. Therefore, the category representations underlying the post-test phoneme boundary of trained naive listeners were expected to differ from those of the advanced students of Finnish. This was supported by the results from 2IFC discrimination.

TABLE I. The phoneme boundary location and boundary width for each of the participating students of Finnish. The first column indicates the students' level in their studies of the Finnish language.

Year	Listener No.	Phoneme boundary (ms)	Boundary width (ms)
1	2	232	36
1	3	230	34
1	6	235	51
2	1	227	31
2	4	235	30
2	8	230	34
3	5	217	33
3	7	221	32

2. One-step and two-step 2IFC discrimination

The question this experiment pursued was whether language learners obtain nativelike perceptual sensitivity with more language experience. The results show that this seems to be the case. Figure 4 gives the results of one-step 2IFC discrimination, comparing perceptual sensitivity between the eight Dutch students of Finnish and the six Finnish natives. On average, the Finnish gave 72% correct responses, whereas the Dutch learners of Finnish gave between 66% and 70% correct responses.

To study the effect of language exposure on discrimination of one-step stimulus differences along the /t-t:/ continuum, an ANOVA with between-subjects factors nativeness (two levels) and duration pair (six levels) was run on the listeners' arcsine transformed responses. A main effect of duration pair was found [$F(5,72)=10.8$, $p<0.001$]. Posthoc analyses showed that discrimination of the between-category pair 201–236 was significantly higher than that of all other pairs, and that duration pair 166–201 was higher than pair 306–341. Moreover, the main effect of nativeness was marginally significant [$F(1,72)=3.9$, $p=0.054$]. We assume that this effect was mainly caused by the higher sensitivity of native listeners at the phoneme boundary (92.5%) than the students' average sensitivity (77.5%). As Fig. 4 shows, how-

ever, the more advanced the students are, the steeper the peak at the phoneme boundary is: Third-year students' discrimination at pair 201–236 was 87.5% correct, whereas discrimination scores of 71.7% and 76.7% were obtained for the first-year and second-year students, respectively.

In order to further investigate the existence of a discrimination peak near the phoneme boundary in advanced learners, we examined within-category and between-category discriminations on the one-step 2IFC task for the third-year students and the Finnish natives. As for the natives, the between-category pair was discriminated better than all other pairs ($p<0.01$) except pair 166–201, which is in line with expectations. The third-year students' discrimination performance showed that duration pair 201–236, the between-category pair, was significantly higher than within-category pairs 131–166, 271–306, and 306–341 ($p<0.05$). The advanced students showed behavior that was similar to the native Finns, except for the difference between the between-category pair and its immediate neighbor, pair 236–271. The trained listeners from experiment I had not shown a between-category peak.

The effect of language exposure on discrimination of two-step stimulus differences was addressed by an ANOVA with between-subjects factors nativeness (two levels) and duration pair (five levels). A main effect of duration pair was present [$F(4,60)=13.5$, $p<0.001$]. All listeners showed high performance levels, with group means varying from 86% to 89%, but could not be separated on the basis of their performance on this test.

C. Discussion

After the previous experiment's finding that short-term training with non-native phoneme contrasts produces minimal changes in perceptual sensitivity, the present experiment was designed to investigate whether language learners would, in the long run, be able to reach nativelike perceptual sensitivity. Despite its small scale, the results from this experiment indicate that the discrimination peak in advanced learners shows a tendency to become nativelike. With increasing language experience, one-step 2IFC discrimination revealed increasing sensitivity near the perceived phoneme boundary, which indicates that advanced language learners can display perceptual sensitivity that is comparable to sensitivity in native listeners.

To further investigate this issue, first the behavior of the two advanced students of Finnish in the one-step test on the stimulus pair straddling the phoneme boundary, pair 201–236, was compared to that of the learners in experiment I. The corresponding discrimination functions can be found in Figs. 2 and 4. The mean pretest discrimination performance of the learners at pair 201–236 was 75% correct, with the 95% confidence interval of the mean extending from 70% to 80%. Their post-test results gave a mean of 79% correct, with a 95% confidence interval of the mean of between 74% and 84%. In experiment II, the two advanced students' individual mean scores at the same stimulus pair were 85% and 90% correct, i.e., outside the confidence intervals of the means found for the learners in experiment I. Second, Pearson correlation coefficients between the mean discrimination scores in the one-step test for each of the groups (learners, advanced students, and Finnish natives) showed that there was a high correlation between the results of the advanced students and the natives, namely, 0.88 ($p=0.019$), but that it was lower between learners and natives, namely, 0.59 (not significant, pretest) or 0.65 (n.s., post-test). These measures support our interpretation of the results obtained in experiment II, but due to the small scale of that experiment we cannot rule out that these results are based on coincidence.

The conclusion from experiment I that perceptual development occurs through AD near the phoneme boundary was corroborated. Especially Fig. 4 seems to illustrate that, as students learn to use closure duration as a cue for distinguishing singleton-geminate contrasts, perceptual sensitivity near the phoneme boundary increases. Moreover, in a visual comparison of the first-year and the third-year students, within-category discrimination tended to be lower in the advanced learners (and also in natives), which parallels our observation in experiment I that within-category AS may also contribute to the development of a boundary peak. Even though the number of participants in this experiment was relatively small, the increasingly nativelike discrimination results co-occurred with increasingly nativelike classification responses.

We observed that the degree of classification improvement from first-year to third-year students roughly coincided with the improvement in classification attained by naive Dutch listeners after only a few hours of training. Section IV discusses how this finding can be explained, given the differences in the discrimination results between experiments I and II.

IV. GENERAL DISCUSSION

In this paper, we set out to study how the peak in perceptual sensitivity develops during the learning of speech sounds, through acquired distinctiveness near the boundary, through acquired similarity within categories, or through a combination of the two. The evidence collected over two experiments mainly supported the first hypothesis and also gave some evidence for the third. Acquired similarity alone was not found. Our experiments challenged claims from earlier research that perception of new speech sounds can develop relatively quickly.

A peak in perceptual sensitivity along a stimulus dimension may have two causes, according to Ades (1977). First, the peak may result from a discontinuity at the sensation level, i.e., in auditory processing. Second, the peak may result from phonetic categorization. The absence of a peak means that neither is the case, implying that the listener does not perceive a phonemic contrast along the dimension under investigation. In experiment I, learners revealed minor changes in perceptual sensitivity, presumably because their phoneme representations developed only minimally. Experiment II suggests that advanced learners develop a peak in perceptual sensitivity, which reflects the acquisition of phonemic representations for the non-native speech sounds. In the rest of this section, we will attempt to explain why the development of such representations needs much more experience than can be provided in short-term training.

The question that remains after experiment I is what the listeners in the short-term training sessions were attending to while performing (and improving in) the classification task. This task does not necessarily reflect the categorization of phoneme classes. A stimulus continuum can be classified into two (or more) categories on the basis of a property that varies between stimuli, as long as perceivers pick up that property. As for auditory stimuli, listeners can also classify plucks and bows (Rosen and Howell, 1981), rise times (Hary and Massaro, 1982), and timbres (Healy and Repp, 1982). Furthermore, in experiment I we saw that Dutch listeners were able to classify the Finnish /t-tː/ continuum into two categories rather sharply without any earlier exposure to Finnish (or any other quantity language). Classification responses can be made purely auditorily, i.e., without reference to underlying representations in long-term memory. A similar finding was reported by Xu *et al.* (2006), who compared the perception of native Mandarin and English listeners on a Chinese tone contrast and a nonspeech tone contrast. They found that both listener groups classified the continua into two categories, whereas only the Chinese listeners also showed a clear discrimination peak near the category boundaries. Therefore, classification results obtained from non-natives or language learners, no matter how nativelike, may insufficiently reflect the presence/absence of phoneme representations, since—as we saw in experiment I—nativelike classification is not necessarily accompanied by nativelike discrimination.

In native listeners, classification and discrimination results show a close relationship [see, e.g., Liberman *et al.* (1957)]. In language learners who have received short-term training, classification and discrimination results are only weakly related. This discrepancy is probably caused by the fact that the two listener groups use different resources to make their auditory judgments. In the psychophysical model of speech perception discussed in Macmillan *et al.* (1988), listener judgments are assumed to be influenced by three types of variance: sensory variance, context variance, and trace variance. Sensory variance is caused by the mapping of a stimulus into an auditory trace. Context variance is associated with the labeling of a trace and is influenced by the size of the stimulus range in an experiment. Trace variance is influenced by the time the trace must remain available before a decision can be made. Sensory variance is always present, but the other two types, context and trace variance, are task dependent.

According to Macmillan *et al.* (1988), both context and trace variance may affect discrimination responses. Their data showed that listeners did not use trace coding for consonant perception in the 2IFC discrimination task: Responses depended on context variance only, i.e., on labeling. In their view, labels are perceptual anchors located at the ends of the experimental stimulus range. Van Hessen and Schouten (1992, pp. 1857–1858), however, subdivided context variance into two types of labels: Apart from *temporary* context labels depending on the stimulus range, they also assumed *permanent* phoneme labels. Under that assumption, the Dutch learners could not use the Finnish phoneme labels in the 2IFC discrimination tests reported in experiment I. The learners did, however, make use of context labels, or "perceptual anchors," that depended on the stimulus range. Due to their temporary nature, context labels fade more rapidly than permanent phoneme labels. That, in turn, complicates the task for non-native listeners: They assign relatively noisy context labels to incoming stimuli and base their order judgment in 2IFC discrimination on a comparison of these temporary labels. Native listeners, on the other hand, use permanent phoneme labels that are much more robust.

In a classification design, listener behavior is also determined by context variance. As long as non-native listeners are unable to make use of phoneme labels in the second language, their identification decision is based on the perceptual anchors they have established on the basis of the stimulus range. Early in the process of learning a second language, classification of non-native phonemes is therefore similar to classification of nonspeech continua. As was mentioned earlier, listeners can do this quite consistently. Short-term training probably helped to establish semipermanent perceptual anchors that explain an improvement in classification behavior. A similar account of our findings is given by a more recent model of categorical perception, the multistore model (Xu *et al.*, 2006), in which a separate short-term categorical memory component is introduced. That component can run in parallel with the processing of sensory traces and has the specific task of filtering out those features from the input signal that are important for categorization. Classification improvement is, in this model, also explained by the establishment of more permanent representations—in so-called long-term categorical memory—that reduce processing load.

In comparison with classification, the 2IFC discrimination task is probably hindered more by the absence of phoneme representations. It is more difficult for a perceiver to compare two fading perceptual traces in time than it is to identify a single trace. This difference may underlie the finding that responses in these tasks are unrelated in language learners. With ample training or language exposure, however, temporary perceptual anchors may develop into permanent phoneme labels in long-term memory. That development is expected to eventually reduce the perceptual differences between native and non-native listeners.

The fact that classification results may not reflect the presence of underlying phoneme categories poses a serious problem for studies that address learning of non-native phoneme categories by means of only a classification task [see, e.g., Flege (1989) and Logan *et al.* (1991)]. Experiment I has shown that there is good reason to doubt fast advances in the laboratory, at least with respect to the development of second-language phonological representations. These need more time and plenty of experience to become nativelike, which is in agreement with Strange and Dittman's (1984, p. 131) view that "the modification of perception of some phonetic contrasts in adults is *slow* and *effortful*".

ACKNOWLEDGMENTS

The authors are grateful to Laurie Stowe from the University of Groningen for providing laboratory facilities and to two anonymous reviewers for their helpful comments.

Ades, A. E. (**1977**). "Vowels, consonants, speech, and nonspeech," Psychol. Rev. **84**, 524–530.
Best, C. T., Triall, A., Carter, A., Harrison, K. D., and Faber, A. (**2003**). "!Xóõ click perception by English, Isizulu, and Sesotho listeners," in Proceedings of the 15th ICPhS, Barcelona, Spain (unpublished), pp. 853–856.
Bohn, O.-S. (**1995**). "Cross-language speech perception in adults: First language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 275–300.
Booij, G. (**1995**). *The Phonology of Dutch* (Clarendon, Oxford).
Bradlow, A. R., and Pisoni, D. B. (**1999**). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," J. Acoust. Soc. Am. **106**, 2074–2085.
Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.
Flege, J. E. (**1989**). "Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training," J. Acoust. Soc. Am. **86**, 1684–1697.
Flege, J. E. (**1995**). "Two procedures for training a novel second language phonetic contrast," Appl. Psycholinguist. **16**, 425–442.
Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Iong, Y., and Lurito, J. (**2002**). "Neural circuitry underlying perception of duration depends on language experience," Brain Lang. **83**, 268–290.
Ganong, W. F. (**1980**). "Phonetic categorization in auditory word perception," J. Exp. Psychol. [Hum Learn] **6**, 110–125.
Gerrits, E. and Schouten, M. E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**, 363–376.
Gibson, E. J. (**1963**). "Perceptual learning," Annu. Rev. Psychol. **14**, 29–56.
Goldstone, R. (**1994**). "Influences of categorization on perceptual discrimination," J. Exp. Psychol. Gen. **123**, 178–200.
Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (**1999**). "Effects of categorization and discrimination training on auditory perceptual space," J. Acoust. Soc. Am. **106**, 2900–2912.
Hary, J. M., and Massaro, D. W. (**1982**). "Categorical results do not imply categorical perception," Percept. Psychophys. **32**, 409–418.
Healy, A. F., and Repp, B. H. (**1982**). "Context independence and phonetic mediation in categorical perception," J. Exp. Psychol. [Hum Learn] **8**, 68–80.
Heeren, W. F. L. (**2006**). "Perceptual development of phoneme contrasts in adults and children," Ph.D. thesis, Utrecht University, Utrecht, The Netherlands.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.
Jamieson, D. G., and Morosan, D. E. (**1986**). "Training non-native speech contrasts in adults: Acquisition of the English contrast by francophones," Percept. Psychophys. **40**, 205–215.
Lehtonen, J. (**1970**). "Aspects of quantity in Standard Finnish," Ph.D. thesis, University of Jyväskylä, Jyväskylä.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368.
Liberman, A., Harris, K. S., Kinney, J. A., and Lane, H. (**1961**). "The discrimination of relative onset-time of the components of certain speech and nonspeech patterns," J. Exp. Psychol. **61**, 379–388.
Lisker, L., and Abramson, A. S. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," in Proceedings of the Sixth ICPhS, Prague (unpublished), pp. 563–567.
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.
Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.
MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholinguist. **2**, 369–390.
Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). "The psychophysics of categorical perception," Psychol. Rev. **84**, 452–471.
Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (**1988**). "Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua," J. Acoust. Soc. Am. **84**, 1262–1280.
Martin, C. S., Mullennix, J. W., Pisoni, D. B., and Summers, W. V. (**1989**). "Effects of talker variability on recall of spoken word lists," J. Exp. Psychol. Learn. Mem. Cogn. **15**, 676–684.
Massaro, D. W. (**1974**). "Perceptual units in speech recognition," J. Exp. Psychol. **102**, 199–208.
McAllister, R., Flege, J. E., and Piske, T. (**2000**). "Aspects of the acquisition of Swedish quantity by native speakers of English, Spanish and Estonian," in Proceedings of the FONETIK 2000 (unpublished), pp. 109–112.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.
Nenonen, S., Shestakova, A., Huotilainen, M., and Näätänen, R. (**2003**). "Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing," Brain Res. Cognit. Brain Res. **16**, 492–495.
Nooteboom, S. G., and Cohen, A. (**1995**). *Spreken en Verstaan (Speaking and Hearing)* (Van Gorcum, Assen/Maastricht).
Pisoni, D. B. (**1975**). "Auditory short-term memory and vowel perception," Mem. Cognit. **3**, 7–18.
Pisoni, D. B., and Tash, J. (**1974**). "Reaction times to comparisons within and across phonetic categories," Percept. Psychophys. **15**, 285–290.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," J. Exp. Psychol. [Hum Learn] **8**, 297–314.
Pisoni, D. B. (**1991**). "Modes of processing speech and nonspeech signals," in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Lawrence Erlbaum, Hillsdale, NJ), pp. 225–238.
Rosen, S. M., and Howell, P. (**1981**). "Plucks and bows are not categorically perceived," Percept. Psychophys. **30**, 156–168.
Strange, W., and Jenkins, J. J. (**1978**). "Role of linguistic experience in the perception of speech," in *Perception and Experience*, edited by R. D. Walk and H. L. Pick (Plenum, New York), pp. 125–169.
Strange, W., and Dittman, S. (**1984**). "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English," Percept. Psychophys. **36**, 131–145.
Strange, W. (**1995**). "Cross-language studies of speech perception. A historical review," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 3–45.
Studebaker, G. A. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Tajima, K., Kato, H., Rothwell, A., and Munhall, K. G. (**2003**). "Perception of phonemic length contrasts in Japanese by native and non-native listeners," in Proceedings of the 15th ICPhS, Barcelona, Spain (unpublished), pp. 1585–1588.
Takagi, N. (**2002**). "The limits of training Japanese listeners to identify English /r/ and /l/: eight case studies," J. Acoust. Soc. Am. **111**, 2887–2896.
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.
Van Hessen, A. (**1992**). "Discrimination of familiar and unfamiliar speech sounds," Ph.D. thesis, Utrecht University, Utrecht, The Netherlands.
Van Hessen, A., and Schouten, M. E. H. (**1992**). "Modeling phoneme perception. II: A model of stop consonant discrimination," J. Acoust. Soc. Am. **92**, 1856–1868.
Van Hessen, A. J., and Schouten, M. E. H. (**1999**). "Categorical perception as a function of stimulus quality," Phonetica **56**, 56–72.
Xu, Y., Gandour, J. T., and Francis, A. L. (**2006**). "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am. **120**, 1063–1074.
Ylinen, S., Shestakova, A., Alku, P., and Huotilainen, M. (**2005**). "The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish," Lang. Speech **48**, 313–338.

Speech perception of noise with binary gains

DeLiang Wang[a)]
Department of Computer Science & Engineering, and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

Ulrik Kjems, Michael S. Pedersen, and Jesper B. Boldt
Oticon A/S, Kongebakken 9, DK-2765 Smørum, Denmark

Thomas Lunner
Oticon Research Centre Eriksholm, Kongevejen 243, DK-3070 Snekkersten, Denmark and Department of Clinical and Experimental Medicine, and Technical Audiology, Linköping University, S-58183 Linköping, Sweden

(Received 9 December 2007; revised 7 July 2008; accepted 8 July 2008)

For a given mixture of speech and noise, an ideal binary time-frequency mask is constructed by comparing speech energy and noise energy within local time-frequency units. It is observed that listeners achieve nearly perfect speech recognition from gated noise with binary gains prescribed by the ideal binary mask. Only 16 filter channels and a frame rate of 100 Hz are sufficient for high intelligibility. The results show that, despite a dramatic reduction of speech information, a pattern of binary gains provides an adequate basis for speech perception.
 [DOI: 10.1121/1.2967865]

PACS number(s): 43.71.Gv, 43.66.Dc [KWG] Pages: 2303–2307

I. INTRODUCTION

Human speech recognition shows remarkable robustness in a variety of listening conditions, including competing talkers, environmental sounds, and background noise. Understanding how speech is recognized under these conditions is important not only for auditory perception but also for automatic speech recognition, where robustness to acoustic interference remains elusive (Lippmann, 1997; Allen, 2005). Related research in computational auditory scene analysis (CASA) and blind source separation makes use of a binary time–frequency ($T-F$) masking technique (Roman *et al.*, 2003; Hu and Wang, 2004; Yilmaz and Rickard, 2004). Time–frequency masking operates on a $T-F$ representation or decomposition of the input into a two-dimensional matrix of $T-F$ units. Such a representation can be readily generated by passing the input signal through a filterbank and then time windowing the response of each filter. Then binary masking as a means of separation amounts to identifying a binary mask where 1 indicates that the acoustic energy in the corresponding $T-F$ unit is retained and 0 indicates that the energy is removed. In other words, binary masking applies a pattern of binary gains to the mixture signal. It should be noted that the term "masking" here means weighting the mixture, which is different from the same term used in psychoacoustics where it means blocking the target sound by acoustic interference.

Among $T-F$ masks, the so-called ideal binary mask (IBM) has been suggested to be a goal of CASA (Wang, 2005). The IBM is a matrix where 1 indicates that the signal-to-noise ratio (SNR) within the corresponding $T-F$ unit exceeds a threshold LC (local SNR criterion) and 0 otherwise. The mask is "ideal" because its construction requires that speech and noise be available before they are mixed, and the mask possesses certain optimality in terms of overall SNR gain when LC is set to 0 dB (Li and Wang, 2008).

Recent studies in speech perception show that applying IBM to noisy speech leads to large speech intelligibility improvements (Brungart *et al.*, 2006; Anzalone *et al.*, 2006; Li and Loizou, 2008). In particular, Brungart *et al.* (2006) and Li and Loizou (2008) have shown that, with fixed levels of input SNR (between −10 and 0 dB), a range of LC values produces nearly 100% correct scores. The large intelligibility gain has been attributed to ideal segregation (or detection) that directs the listener's attention to the $T-F$ regions of noisy speech where the target speech is relatively strong. This explanation emphasizes the importance of the target signal contained in the $T-F$ units labeled 1 for intelligibility. How important is the binary pattern of an ideal mask itself? This investigation was designed to isolate the intelligibility contribution of an IBM by removing the target speech signal from all $T-F$ units.

Specifically, with linear filters, including gammatone filters (Patterson *et al.*, 1988; Wang and Brown, 2006), increasing or decreasing the SNR of a mixture while changing LC by the same amount does not alter the IBM. On the other hand, although co-reducing input SNR and LC does not change the IBM, the masked mixture becomes progressively noisier or contains less target signal. Taking this manipulation to the limit, i.e., setting both mixture SNR and LC to $-\infty$ dB, leads to an output that contains only noise with no target speech at all. This particular output corresponds to turning on or off the filtered noise according to a pattern prescribed by the IBM. Our study evaluates speech intelligibility of noise gated by the IBM obtained in this way.

[a)]Author to whom correspondence should be addressed. Electronic mail: dwang@cse.ohio-state.edu

II. METHOD

A. Stimuli

Our tests use sentences from the Dantale II data set as target speech and a speech-shaped noise as interference (Wagener *et al.*, 2003). The speech material in the Dantale II corpus consists of sentences recorded by a female Danish speaker. Each sentence has five words with a fixed grammar (name, verb, numeral, adjective and object), e.g., "Michael had five new plants" (English translation). Each position in a sentence takes a randomly chosen word from ten equally meaningful words. The speech-shaped noise included in the Dantale II corpus is produced by adding repeated utterances of each of the 250 test sentences in the corpus (see Wagener *et al.*, 2003). Both speech and noise materials were digitized at 20 kHz sampling frequency.

A speech utterance and the noise are first processed by a gammatone filterbank with varying numbers of filter channels. With 32 filters equally spaced on the equivalent rectangular bandwidth (ERB) rate scale with center frequencies distributed in the range of 2–33 ERBs (or 55–7743 Hz), the frequency response of the filterbank is nearly flat. In addition to a 32-channel gammatone filterbank, we also tested 16-, 8-, and 4-channel filterbanks. Each of the filterbanks spans the same frequency range with filters equally spaced on the ERB-rate scale, and in all cases each filter has the bandwidth of 1 ERB. With reduced channels, the frequency response of a filterbank is no longer flat and information in certain frequency bands is lost in comparison to the 32-channel filterbank case. A filter response is then windowed into time frames using 20 ms rectangular windows and a frame shift of 10 ms. This 100 Hz frame rate is commonly used in speech processing (Rabiner and Juang, 1993). The resulting $T-F$ representation has been called a cochleagram (Wang and Brown, 2006). The IBM is constructed from the cochleagrams of the target speech and the speech-shaped noise with both the mixture SNR (calculated during the period of a sentence) and LC set to 0 dB. The IBM is then applied to the noise cochleagram alone in a synthesis step to generate a waveform stimulus [see Wang and Brown (2006) for details of cochleagram analysis and synthesis]. Figure 1 illustrates the signal processing scheme using a Dantale II sentence. Take, for example, the 8-channel filterbank case. Figure 1(G) shows the IBM for this case, which is produced by comparing the 8-channel cochleagram of the Dantale II sentence and the 8-channel cochleagram of the speech-shaped noise. Applying the binary mask in Fig. 1(G) to gate the noise results in a waveform signal, which is represented in the cochleagram format in Fig. 1(H). Note that Fig. 1 represents the waveform signals from different channel numbers using the same 32-channel cochleagram representation in order to facilitate comparison. In other words, all the cochleagrams in Fig. 1 serve the purpose of signal representation and do not indicate the size of the filterbank used in IBM construction.

B. Subjects

Twelve normal-hearing, native Danish-speaking listeners participated in the experiment. All listeners had normal hearing, i.e., their hearing thresholds did not exceed 20 dB HL, and their ages ranged from 26 to 51 with the average age of 36.

C. Procedure

In each condition of the experiment, two lists, each with ten sentences, were randomly selected from the Dantale II corpus for IBM construction. Speech-shaped noise gated by the IBM was then presented to a listener. The subjects were instructed to repeat as many words as they could after listening to each stimulus corresponding to one sentence, and no feedback was provided to them regarding whether their responses were correct or not. A stimulus was presented only once. Subjects were given a training session by listening to two lists of clean (or unprocessed) sentences, which were not included in the subsequent test. Each subject test had four conditions corresponding to the filterbanks with 4, 8, 16, and 32 channels. The four test conditions plus training took less than 30 min. The presentation order of the four conditions was randomized and balanced among the 12 listeners.

Speech and noise were both set to the sound pressure level of 70 dB initially. To account for level differences caused by the use of different-sized filterbanks, stimuli were scaled by factors of two, four, and eight, for the 16-channel, the 8-channel, and the 4-channel filterbank, respectively. This level calibration resulted in stimuli with approximately the same loudness. On each trial, a stimulus was generated by the built-in sound card (SoundMAX) in a control computer (IBM ThinkCenter S50) and then presented diotically to a listener through headphones (Sennheiser HD 280 Pro) in a sound treated hearing test room.

III. RESULTS

Figure 2 shows the word recognition performance for all four conditions. The mean speech intelligibility scores for the four conditions are: 7.75%, 54.25%, 92.92%, and 97.08%, with increasing number of filter channels. The results show that nearly perfect speech recognition is obtained with 32 channels, and a high recognition rate is obtained with 16 channels. The subjects recognized more than half of the words when the number of channels was set to 8, but were unable to perform the recognition task when the number of channels was 4. A repeated measures analysis of variance (ANOVA) was conducted and the effect of number of channels was significant, $F(3,33)=179.05$, $p<0.00001$. The Tukey honest significant difference (HSD) procedure revealed that all pairwise differences among the means were significant, $p<0.001$, except for the difference between 16 and 32 channels, which was not significant. Both ANOVA and post hoc Tukey HSD tests were conducted on the rationalized arcsine-transformed percentage scores (Studebaker, 1985).

The performance variability across different listeners was small for the 32-channel and the 16-channel cases, suggesting that the acoustic information was sufficient for them to perform the recognition task. On the other hand, the individual variability for the 8-channel case was significantly larger than the 16-channel case, $F(1,11)=5.50$, $p<0.01$, sug-

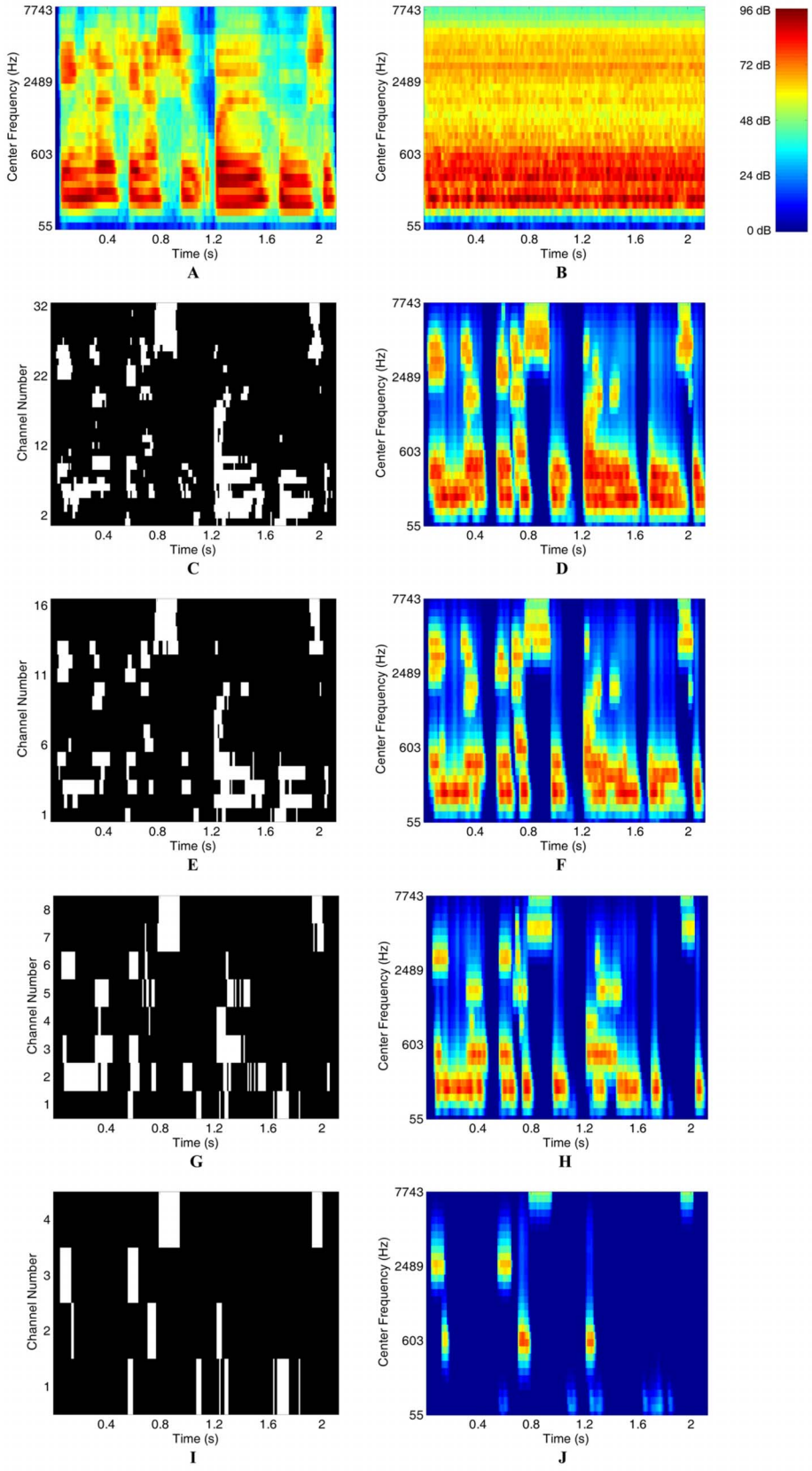


FIG. 1. (Color online) Illustration of gated noise by binary gains. (A) 32-channel cochleagram of a Dantale II sentence. (B) 32-channel cochleagram of speech-shaped noise. (C) IBM with 32 channels, where 1 is indicated by white and 0 by black. (D) 32-channel cochleagram of gated noise by the IBM in (C). (E) IBM with 16 channels. (F) 32-channel cochleagram of gated noise by the IBM in (E). (G) IBM with 8 channels. (H) 32-channel cochleagram of gated noise by the IBM in (G). (I) IBM with 4 channels. (J) 32-channel cochleagram of gated noise by the IBM in (I).

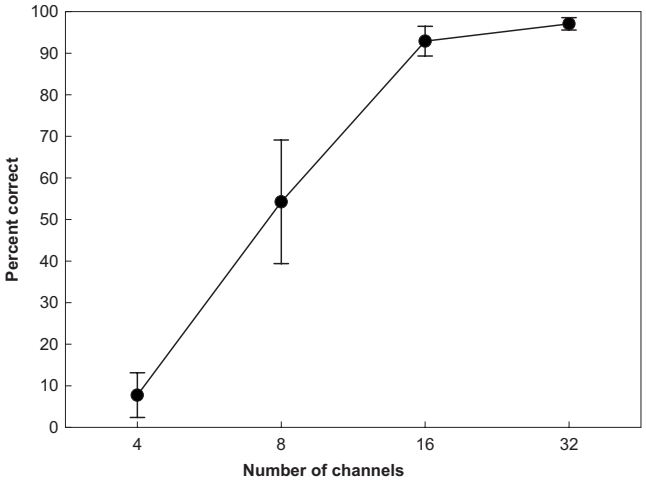


FIG. 2. Word intelligibility scores for 12 normal-hearing listeners with respect to different filterbank sizes. Dots denote the mean scores and vertical bars indicate 95% confidence intervals.

gesting that factors such as the ability and tendency to guess, concentration, and prior experience with corrupted speech, come into play.

The results in Fig. 2 clearly demonstrate that very high recognition can be obtained by turning on and off 16 bands of noise at a rate of 100 Hz following a specific pattern. The speech signal plays the sole role of determining the IBM. Such a stimulus contains little speech-specific information. The spectral shape of speech is drastically reduced to a binary variation, and so is the temporal envelope. The harmonic structure of voiced speech is absent, and the temporal fine structure (the carrier signal underlying the temporal envelope) of the stimulus reflects that of noise, not speech. Despite this dramatic reduction of speech information, listeners are capable of speech perception.

So what cues enable listeners to perceive speech from IBM-gated noise? The binary pattern encodes a general outline of spectrotemporal energy variations of speech relative to noise. Binary-gated noise crudely reflects the formant structure of speech; as shown in Fig. 1, IBM-gated noise appears to "carve out" regions of noise energy that roughly match the spectrotemporal peaks of speech. Our results indicate that such a pattern of energy variations is sufficient for recognition purposes.

IV. DISCUSSION AND CONCLUSION

Our study bears resemblance to the well-known study by Shannon *et al.* (1995) demonstrating that only a few bands of noise modulated by the corresponding speech envelopes suffice for speech intelligibility [for a much earlier study using more bands see Dudley (1939)]. There are, however, several differences between our binary-gated noise and the vocoded noise of Shannon *et al.* Perhaps the most important and obvious difference is that, within a frequency channel, noise modulation uses a binary envelope in our study and a full envelope in vocoded noise. Second, the IBM is derived by a comparison between target speech and speech-shaped noise, while temporal envelopes in vocoded noise are obtained from target speech alone. We note that many speech separation algorithms compute a binary mask by implicitly or explicitly exploiting local SNR (Divenyi, 2005; Wang and Brown, 2006), making the ideal mask amenable to computational estimation. Third, the bandwidths of noise bands in Shannon *et al.* change as the number of the bands varies in order to cover the entire frequency range of interest; in IBM-gated noise, the bandwidth of each frequency channel is fixed to 1 ERB regardless of the number of filtbank channels. It is also worth mentioning that recognizing vocoded noise takes hours of training, while no training on binary-gated noise was given in our experiment.

Like vocoded noise, the type of noise used in binary gating likely has an effect on speech intelligibility. The speech-shaped noise used in this study is a steady noise with a long-term spectrum matching that of the utterances in the Dantale II corpus, and may be particularly effective for IBM gating, although our informal listening indicates that other types of steady noise, such as pink noise, can also produce intelligible speech. Our experiment was conducted using Danish utterances. Byrne *et al.* (1994) reported that the long-term average speech spectra of a group of languages, including Danish and English, are quite similar, suggesting that, though there are likely some language effects, the main observations of our experiment may hold for English and other languages. Also, the IBM used in this study is constructed when input SNR and LC are set to be equal ($-\infty$ dB). Fixing one of them while varying the other produces different IBMs. For example, when input SNR is set to 0 dB, increasing LC results in ideal masks with fewer and fewer 1's, whereas decreasing LC leads to more and more 1's. Is equating input SNR and LC most effective for intelligibility of IBM-gated noise? Further investigation is required to address the issues regarding noise type, language, and input SNR and LC levels.

That a pattern of binary gains is apparently sufficient for human speech recognition, like previous work on vocoded noise, raises intriguing questions on the nature of human speech recognition. What speech information is truly indispensable for intelligibility? Could the IBM itself be what is being recognized? Almost perfect speech recognition from such drastically reduced speech information has broad implications for CASA, automatic speech recognition, hearing prosthesis, and coding and compression in speech communication.

ACKNOWLEDGMENTS

The authors thank Associate Editor Ken Grant, and two anonymous reviewers for their constructive criticisms and suggestions. The work was performed while D.L.W. was a visiting scholar at Oticon A/S. The authors thank M. Schlaikjer, L. Bramsløw, and M. Hartvig for their assistance in the experiments, and Y. Li for his assistance in figure preparation. D.L.W. was supported in part by AFOSR Grant No. F49620-04-01-0027 and NSF Grant No. IIS-0534707.

Allen, J. B. (**2005**). *Articulation and Intelligibility* (Morgan & Claypool, San Rafael, CA).

Anzalone, M. C., Calandruccio, L., Doherty, K. A., and Carney, L. H. (**2006**). "Determination of the potential benefit of time-frequency gain

manipulation," Ear Hear. **27**, 480–492.
Brungart, D., Chang, P. S., Simpson, B. D., and Wang, D. L. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.
Byrne, D., Dillion, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.
Divenyi, P., ed. (**2005**). *Speech Separation by Humans and Machines* (Kluwer Academic, Norwell, MA).
Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.
Hu, G., and Wang, D. L. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.
Li, N., and Loizou, P. C. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673–1682.
Li, Y., and Wang, D. L. (**2008**). "On the optimality of ideal binary time-frequency masks," in *Proceedings of IEEE ICASSP*, pp. 3501–3504.
Lippmann, R. P. (**1997**). "Speech recognition by machines and humans," Speech Commun. **22**, 1–16.
Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (**1988**). "SVOS final report, part B: Implementing a gammatone filterbank," Rep. No. 2341, MRC Applied Psychology Unit.
Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
Roman, N., Wang, D. L., and Brown, G. J. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Wagener, K., Josvassen, J. L., and Ardenkjær, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.
Wang, D. L. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.
Wang, D. L., and Brown, G. J., eds. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley, Hoboken, NJ).
Yilmaz, O., and Rickard, S. (**2004**). "Blind separation of speech mixtures via time-frequency masking," IEEE Trans. Signal Process. **52**, 1830–1847.

Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility

Alexander Kain, Akiko Amano-Kusumoto, and John-Paul Hosom
Center for Spoken Language Understanding (CSLU) at the OGI School of Science and Engineering, Oregon Health and Science University (OHSU), 20000 NW Walker Road, Beaverton, Oregon 97006

(Received 1 June 2007; revised 1 July 2008; accepted 7 July 2008)

Speakers naturally adopt a special "clear" (CLR) speaking style in order to be better understood by listeners who are moderately impaired in their ability to understand speech due to a hearing impairment, the presence of background noise, or both. In contrast, speech intended for nonimpaired listeners in quiet environments is referred to as "conversational" (CNV). Studies have shown that the intelligibility of CLR speech is usually higher than that of CNV speech in adverse circumstances. It is not known which individual acoustic features or combinations of features cause the higher intelligibility of CLR speech. The objective of this study is to determine the contribution of some acoustic features to intelligibility for a single speaker. The proposed method creates "hybrid" (HYB) speech stimuli that selectively combine acoustic features of one sentence spoken in the CNV and CLR styles. The intelligibility of these stimuli is then measured in perceptual tests, using 96 phonetically balanced sentences. Results for one speaker show significant sentence-level intelligibility improvements over CNV speech when replacing certain combinations of short-term spectra, phoneme identities, and phoneme durations of CNV speech with those from CLR speech, but no improvements for combinations involving fundamental frequency, energy, or nonspeech events (pauses). © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967844]

PACS number(s): 43.71.Gv, 43.72.Ja, 43.71.Es [DOS] Pages: 2308–2319

I. INTRODUCTION

Approximately 28×10^6 people in the United States have some degree of hearing loss, with 40%–45% of the population over 65, and about 83% of those over 70, classified as hearing impaired (Cruickshanks *et al.*, 1998). Elderly listeners often have an especially difficult time understanding speech in noise or under distracting conditions (Salthouse, 1996). Wearing a hearing aid is one of the most often-used strategies that can partially compensate for a hearing impairment. The primary benefit of hearing aids is to restore hearing loss resulting from reduced sensitivity, by amplifying signal energy in one or more frequency bands, with optional dynamic compression and expansion (Dillon, 2001). No attempts are made to perform prosodic or fine-grained spectral modifications, even though it is known that increased speech intelligibility can be obtained by processes distinct from simply regulating the energy of the speech signal. For example, speakers naturally adopt a special speaking style when aiming to be understood by listeners who are moderately impaired in their ability to understand speech, due to hearing loss, the presence of background noise, or both. This style has been termed "clear" (e.g., Picheny *et al.*, 1985). In contrast, speech intended for a normal-hearing listener in a quiet environment is commonly referred to as "conversational." These two styles of speaking will be referred to as CLR and CNV speech, respectively. The intelligibility of CLR speech is higher than that of CNV speech, as measured in listeners of different age groups, with normal and impaired hearing abilities, using different types of speech materials, and in environments with different types of background noise (Picheny *et al.*, 1985; Payton *et al.*, 1994; Schum, 1996; Helfer, 1998; Bradlow and Bent, 2002; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002; Bradlow *et al.*, 2003; Ferguson, 2004; Liu *et al.*, 2004).

Previous research has examined acoustic differences between CNV and CLR speech. The following *prosodic* features have been noted to distinguish CLR speech from CNV speech: (1) The fundamental frequency (F0) was typically increased in range and mean (Picheny *et al.*, 1986; Bradlow *et al.*, 2003; Krause and Braida, 2004); (2) the consonant-vowel energy ratio (CVR) was increased (energies of the consonants were greater in CLR speech), particularly for stops and affricates (Bradlow *et al.*, 2003); however, other researchers found an increased CVR for affricates only (Krause and Braida, 2004); (3) phoneme durations were prolonged, especially in the tense vowels /i/, /u/, /ɑ/, and /ɔ/ (Picheny *et al.*, 1986; Ferguson and Kewley-Port, 2002); (4) pauses were longer in duration and occurred more frequently (Picheny *et al.*, 1986; Bradlow *et al.*, 2003); and (5) the speaking rate was significantly decreased from 160–200 words/min to 90–100 words/min, which was at least partially the result of longer vowels in CLR speech (Picheny *et al.*, 1986; Krause and Braida, 2004).

The following *spectral* features have been noted to distinguish CLR speech from CNV speech: (1) Vowel formant frequencies showed expanded vowel spaces for lax vowels, (Picheny *et al.*, 1986), (2) long-term spectra had increased energies at higher frequencies (1000–3150 Hz) (Krause and Braida, 2004), (3) alveolar flaps occurred less often and consonant stops (e.g., /t/) tended to be released with following aspiration (Picheny *et al.*, 1986; Krause and Braida, 2004; Bradlow *et al.*, 2003); and (4) four out of five speakers

 0001-4966/2008/124(4)/2308/12/$23.00

exhibited increased modulation indices for low modulation frequencies up to 3–4 Hz (Krause and Braida, 2004).

Different speakers appear to employ different strategies to produce CLR speech, resulting in significantly different vowel intelligibility (Ferguson and Kewley-Port, 2002; Ferguson, 2004). In the same study, the authors found that not all speakers can improve their intelligibility beyond their CNV speech level. The least intelligible speakers produced CLR speech with (1) shorter durations, (2) the least differentiated vowel spaces as measured by the first two formants, (3) minimal cues for consonantal contrasts, and (4) the most varied amplitude of stressed vowels (Bond and Moore, 1994). Speakers with larger vowel spaces were generally more intelligible than speakers with reduced vowel spaces (Bradlow *et al.*, 1996). The relative intelligibility of an individual speaker has been shown to be consistent across listener age groups (7–8 year olds, 11–12 year olds, and adults with mean age of 29.9 years) (Hazan and Markham, 2004).

Current hearing aid systems focus on amplifying the speech signal in one or more frequency bands. However, this approach does not address important problems that may be encountered by users, especially elderly listeners (Pichora-Fuller *et al.*, 1995). For example, elderly listeners have more difficulty than younger listeners understanding rapid speech due to decreased auditory processing capabilities (e.g., Gordon-Salant and Fitzgibons, 2001) or reduced working memory capacity (Wingfield *et al.*, 2005). Motivated by these findings, researchers have developed signal-processing algorithms to increase the intelligibility of speech independent of amplification. Modifications to the speech waveform include decreasing the rate of speech by inserting pauses (Liu and Zeng, 2006), modifying phoneme durations (Nejime and Moore, 1998; Gordon-Salant, 1986; Uchanski *et al.*, 1996), and enhancing the consonant-to-vowel energy ratios (Gordon-Salant, 1986, 1987; Hazan and Simpson, 1998). Of these, only one study showed a statistically significant intelligibility increase of 4.2% at the sentence level, by amplifying the energy of specific consonants (Hazan and Simpson, 1998). Other studies did not report statistically significant improvements at the sentence level; this relative lack of success was sometimes conjectured to be caused by signal-processing artifacts (e.g., Uchanski *et al.*, 1996), although the manipulation of isolated features may also have been a contributing factor, as interactions among features are common. Ultimately, the causal relationship between sets of acoustic features and speech intelligibility, using sentence-level speech materials, is not known.

The long-term goal of this research effort is to develop a model that quantifies the cause-and-effect relationship between acoustic features of speech and speech intelligibility. Such a model may have applications in novel signal processing algorithms for hearing aids and other assistive listening devices that transform CNV speech into a closer approximation of CLR speech, for postprocessing speech output from general-purpose communication devices (e.g., phones and video playback devices), and for objective measures of speech intelligibility.

This article reports on initial experiments that measured the degree of contribution (DOC) of six high-level acoustic features to intelligibility. It is not the intent of this work to provide a complete list of all relevant feature combinations with their respective degrees of contribution, but to demonstrate that it is possible to determine the significance of specific feature combinations to intelligibility. Similar to previous approaches, in which phoneme durations from CLR speech were applied to CNV speech (Uchanski *et al.*, 1996), or in which the temporal envelope from CLR speech was applied to CNV speech using "chimerized" speech (Liu and Zeng, 2006), the present work applies certain CLR features to CNV speech. This has been accomplished by using a "hybridization" algorithm that (1) extracts CNV and CLR features from the same sentences spoken in both CNV and CLR styles, then (2) constitutes a "hybrid" (HYB) feature set from a particular subset of CLR features and from the complementary subset of CNV features, and finally (3) synthesizes HYB sentences from the HYB features. Testing the intelligibility of the HYB speech allows us to identify not only which acoustic features contribute to intelligibility but also their DOC. The DOC is subject to listener variation; because of the relatively small number of listeners used in the following experiments, DOC values are not precise metrics but are provided to give a sense of how much impact a feature has. It was hypothesized that certain subsets of acoustic features of CLR speech contribute significantly more to speech intelligibility than others.

II. SPEECH CORPUS

A. Text material and recording

Two types of text material were used in this study, referred to as material A and material B.

Material A. This material consisted of 70 sentences from the set of IEEE Harvard Psychoacoustic Sentences (Rothauser *et al.*, 1969), which are phonetically balanced, syntactically, and semantically normal sentences, with each sentence containing five keywords (e.g., His *shirt* was *clean* but *one button* was *gone*).

Material B. This material consisted of 70 syntactically correct, but semantically anomalous sentences (e.g., They *slide far across* the *tiny clock*), created by randomizing and exchanging words and grammar structures from material A (Vaughan *et al.*, 2006). Using identical words and sentence lengths in both materials allowed for direct comparisons between experimental results.

It was expected that material A would be easier to understand because the availability of semantic context is an important factor contributing to speech intelligibility (Gordon-Salant and Fitzgibbons, 1997), but these sentences are still not as predictable as everyday speech (Rothauser *et al.*, 1969).

One male, a native speaker of American English with no professional training in public speaking, was recruited as a speaker (J.P.H.). First, he recorded the 140 sentences of materials A and B spoken in the CNV speaking style, followed by the same sentences spoken in the CLR speaking style. When recording CNV speech, he was instructed to speak in

TABLE I. Example configurations governing the hybridization algorithm. Each configuration determines a speech style (CNV or CLR) as source for six acoustic features of speech.

Name	Energy	F0	Duration	Spectrum	Phoneme	Non speech
CNV	CNV	CNV	CNV	CNV	CNV	CNV
HYB-D	CNV	CNV	CLR	CNV	CNV	CNV
HYB-EFN	CLR	CLR	CNV	CNV	CNV	CLR
CLR	CLR	CLR	CLR	CLR	CLR	CLR

the way that he uses to communicate in his daily life. When recording CLR speech, he was instructed to speak clearly as he would when communicating with hearing-impaired listeners.

The recording was carried out in a sound-treated booth (Whisperroom MDL4260) located inside a control room. Recordings were made using a head-mounted close-talking microphone (AKG HSC200), positioned approximately 5 cm and off axis from the speaker's mouth. The speaker recorded the materials at his own pace by operating a computer program. A technician listened to each sentence and the speaker was asked to record a sentence again when pronunciation or style were not satisfactory. The speech signals were captured and stored digitally at a sampling rate of 22.05 kHz with 16 bit resolution.

Sentences were partitioned into two groups, based on an informal characterization of the degree of difference between renditions of the two styles. For each material, the 48 sentences with apparently larger intelligibility differences (partition LD) were used for intelligibility testing (see Sec. IV D), while the remaining 22 sentences with apparently smaller differences (partition SD) were used for setting noise levels for those tests (see Sec. IV B).

B. Additional annotation

Time-aligned phoneme label and "pitch mark" (also known as glottal closure instant) annotation was added to the corpus of speech waveform recordings. Labels included nonspeech events, almost always occurring in the form of pauses, but also in the form of breath noise, lip smack, and tongue click. This additional annotation was required for processing the speech waveforms as part of the hybridization algorithm (see Sec. III). Initial estimates of phoneme identities and boundaries in each waveform were obtained using an existing forced-alignment system (Hosom, 2002). To create an initial estimate of the pitch marks in each waveform, a standard software package was used (Boersma, 1993). Then, a trained labeler checked and adjusted pitch marks as well as phoneme identities and boundaries manually. The entire corpus, including waveforms, phoneme labels, and pitch marks, is available through the CSLU corpus distribution mechanism (Center for Spoken Language Understanding, 2007).

C. Corpus characteristics

The phonetic and acoustic characteristics of the CNV and CLR sentences were analyzed using partition LD.

1. Phonetic characteristics

Phoneme statistics confirmed typical differences between the two speaking styles; for example, CLR speech had more pauses and stop releases, while CNV speech had a larger number of reduced vowels. A phonetic alignment (see Sec. III B 1) of partition LD of materials A and B (9090 labels in 96 sentences) resulted in 109 labels from CLR speech being considered as insertions into CNV speech (among them 38 pauses and 29 unvoiced plosive releases), while 20 labels from CNV speech were considered as insertions into CLR speech (8 of which were pauses). This corresponds to approximately 0.9 and 0.5 insertion/deletion operations per sentence, for phonemes and nonspeech events, respectively. Phoneme substitutions (e.g., /u/ versus /ʉ/) occurred 150 times.

2. Acoustic characteristics

An acoustic analysis showed that, for CNV speech, the mean and standard deviation (SD) of F0 were 105.0 and 15.4 Hz, respectively, whereas CLR speech had mean and SD values of 106.2 and 17.6 Hz, respectively. Using a frame-by-frame root-mean-squared energy measure, it was found that the energy of CLR speech (excluding nonspeech sounds) was 0.9 dB above that of CNV speech (1.1 dB for vowels only), and the SD of CLR speech energy was approximately 30% larger than the SD of CNV speech. The average CVRs were measured and determined to be −7.2 dB in CNV speech and −7.0 dB in CLR speech for plosives, and −6.9 dB in CNV speech and −7.1 dB in CLR speech for affricates. Spectrally, the vowels of CLR speech had an approximately constant energy increase of 2 dB in frequencies above 1800 Hz. Finally, the average duration of interword pauses was 33 ms in CNV speech and 62 ms in CLR speech.

III. HYBRIDIZATION ALGORITHM

The purpose of the hybridization algorithm was to obtain a speech waveform that combines acoustic features of the CLR and CNV speech waveforms. Table I shows example feature configurations and corresponding stimulus conditions (see Sec. IV D). Configurations involve the following six high-level acoustic features: energy trajectory (E), F0 trajectory (F), phoneme durations (D), short-term spectra (S), phoneme sequence (P), and presence of nonspeech sounds such as pause, breath noise, lip smack, and tongue click (N). Configurations are named according to the acoustic features that are taken from CLR speech; for example, HYB-EFN represents a HYB speech waveform whose fea-

tures consist of energy, F0, and nonspeech events from CLR speech (the ordering of features is immaterial), while all other features are taken from CNV speech. Experiments 1–3 (see Sec. IV) evaluated the intelligibility of these HYB speech stimuli to determine the DOC of particular sets of acoustic features used from CLR speech, as compared to the baseline CNV speech.

A. Normalizing levels

Because increased loudness can contribute to speech intelligibility, it was important to minimize sentence-level loudness differences by normalizing all sentences (both CLR and CNV styles, and both text materials) in the corpus, using the following procedure: Speech signal waveforms were first filtered with an A-weighted filter (International Electrotechnical Commission, 2002), and levels were calculated by averaging frame-by-frame (using frame durations on the order of 10 ms) root-mean-squared (rmsA) sample values of nonpausal (as determined by the phoneme alignment described in Sec. II B) portions of the speech. Each waveform was then multiplied by an appropriate gain factor so that the resulting waveforms all had the same average rmsA value, while at the same time ensuring good resolution with sufficient headroom (e.g., setting the global sample peak to 80% of the maximum absolute sample value). This allowed for possible energy increases during hybridization.

B. Algorithm implementation

The hybridization algorithm was implemented by first aligning the CLR phonetic sequence with the CNV phonetic sequence (step 1) and then "parallelizing" the two original waveforms in terms of phonetic content (step 2). Next, a speech analysis was carried out on the parallelized waveforms, consisting of first computing auxiliary marks (step 3) and then extracting acoustic features (step 4). Next, configuration-specific features of CLR speech were combined with complementary features of CNV speech to form HYB features (step 5) and finally synthesized to create the HYB speech waveform (step 6). Each step is now described in more detail.

1. Alignment of phoneme sequences

The phoneme sequences of CNV and CLR speech are occasionally different because, even when identical sentences are used, a speaker may pronounce the material differently, depending on the speaking style (see Sec. II C 1). However, the feature combination (step 5) required that the input speech signals have a one-to-one mapping between phoneme sequences; therefore, phoneme-sequence pairs of each sentence produced in the CNV and CLR styles needed to be aligned.

To accomplish the alignment, a phoneme feature table was created, specifying voicing (1–5: from fully voiced to voiceless, e.g., 1: vowels, 3: voiced fricatives), manner (0–10: from highest to lowest sonority, e.g., 3: glides, 6: fricatives), place (1–8: from front to back, e.g., 1: bilabial, 8: velar), and height (1–10: from lowest to highest tongue height, e.g., 1:/a/, 4:/u/) features), with one four-dimensional vector for each phoneme (e.g., /i/: [1,2,4,4,], /k/: [4,7,8,7]). Each phonetic symbol in both label sequences was assigned its associated feature vector, resulting in two feature matrices. Then, dynamic time warping (Rabiner and Juang, 1993) was used to find an optimal alignment path between the two matrices that resulted in the minimum Euclidean distance between the corresponding phonetic features, while observing sensible local constraints. As a result, each phoneme in one speaking style was associated with one phoneme in the other speaking style, either by a perfect match or a best-fit match. In those cases where a one-to-one mapping was not possible, the phoneme was considered to be an insertion (or a deletion, depending on the point of view). The alignment path was then automatically converted to a list of operations (insertion/deletion for one-to-many mappings, no change for perfect or best matches) and stored. This operation list was sometimes changed manually based on phonetic knowledge.

TABLE II. An example of the phoneme alignment operations and corresponding parallelization for a HYB-P configuration. The first two columns contain the CNV and the CLR speech phoneme sequence, after alignment, with the third column indicating the corresponding operation. The last column contains the hybrid phoneme sequence obtained when setting *Phoneme*=CLR and *Nonspeech*=CNV, necessitating an insertion of the CLR plosive release /$d^{(\urcorner)}$/ into the CNV speech, and a deletion of the pause /(.)/ from the CLR speech.

CNV	CLR	Operation	HYB
b	b	⋯	b
ɪ	i	⋯	ɪ;i
s	s	⋯	s
aɪ	aɪ	⋯	aɪ
$d^{(\urcorner)}$	$d^{(\urcorner)}$	⋯	$d^{(\urcorner)}$
⋯	$d^{(\ulcorner)}$	ins	$d^{(\ulcorner)}$
⋯	(.)	del	⋯

2. Parallelization of waveforms

The final HYB phoneme sequence was dependent on the values of the phoneme (P) and nonspeech (N) configuration settings (see Table I). Therefore, the original waveforms were modified to implement phoneme deletions or insertions, which may occur in the CNV speech, the CLR speech, or both.

For phoneme insertions, the relevant portion was extracted from the corresponding alternative style and inserted without any time-domain cross-fade. (This technique was later modified, see Sec. IV E.) Phoneme deletions were carried out by removing the relevant portion from the waveform and concatenating the two remaining sections together. Because an inserted waveform segment resulted in identical waveforms in that region for both styles, hybridization was not possible in those regions.

Table II shows the result of an example phoneme alignment and the resulting parallelization, given a HYB-P configuration, which means that phonemes were taken from CLR speech, but that nonspeech events were taken from CNV speech. In this example, the CNV speech phoneme /ɪ/ was aligned with the CLR speech phoneme /i/, which was considered a substitution (noted in the HYB phoneme se-

quence as /ɪ; i/); hybridization can proceed normally here. The waveform associated with the /$d^{(\ulcorner)}$/ phoneme label from the CLR speech was copied and inserted into the CNV speech. (Note that in this work the symbol /$d^{(\urcorner)}$/ denotes /d/ closure and /$d^{(\ulcorner)}$/ denotes the released part of the plosive.) As a result, because both speaking styles contained identical waveforms for this phoneme, and thus all acoustic features were identical, hybridization could not take place in the /$d^{(\ulcorner)}$/ phoneme. Finally, the waveform associated with the (interword) pause /(.)/ was deleted from the CLR speech, in order to satisfy the requirement of following the nonspeech pattern of the CNV speech.

3. Computing auxiliary marks

The hybridization algorithm processed speech by performing computations on short successive segments of speech, known as frame-by-frame processing. The extent of each frame was defined by the location of consecutive "time marks." In voiced regions, these time marks were defined by pitch marks (the instants of glottal closure); in unvoiced regions, time marks were defined by "auxiliary marks," which needed to be created to keep the frame size appropriately short. Whenever the distance between two pitch marks was more than 16 ms (or, equivalently, when F0 was lower than 62.5 Hz), the region was considered unvoiced. Auxiliary marks were placed inside unvoiced regions at intervals of approximately 10 ms. Care was taken that auxiliary marks and pitch marks were not closer than 10 ms to each other, thus avoiding very short frames that could lead to problems during the feature extraction step.

4. Feature extraction

Acoustic features from CNV and CLR speech were extracted by performing a frame-by-frame pitch-synchronous analysis. Consecutive overlapping analysis frames were defined by three consecutive time marks spanning two periods of speech. The short-term spectrum (S) was represented by storing the analysis frame's waveform directly. To obtain an energy trajectory (E), a Hanning window was applied to each analysis frame, and then the corresponding rmsA values were calculated (see Sec. III A). F0 trajectory values (F) were obtained by inverting the differences between pitch marks. These features were considered to represent the speech waveform at the time of their respective analysis frame centers. Finally, speech durations (D) were specified at the phoneme level and could be directly derived from the labels in the speech corpus.

5. Feature combination

In preparation for the final synthesis step, HYB features were formed from a particular subset of CLR features and from the complementary subset of CNV features, as specified by the desired configuration (see Table I). For example, the configuration HYB-DSP specified that the CLR speech was taken as the source for phoneme durations, spectral information, and phoneme sequence, and that the complementary features, namely, energy trajectory, F0 trajectory, and nonspeech sequences, were taken from CNV speech.

6. Speech synthesis

The HYB features created in the previous step were used to synthesize a HYB speech waveform using a pitch-synchronous, overlap-add, residual-excited, linear predictive coefficient (LPC) synthesis method (similar to Taylor *et al.*, 1998). The method controlled energy trajectories, F0 trajectories, and phoneme durations using short frames of speech, which were constructed from three consecutive time marks (spanning the waveform of two speech periods).

F0 modification of a frame of speech was implemented by first pre-emphasizing the speech and then calculating 24th-order LPC coefficients (using the autocorrelation method) and corresponding LPC residuals. The length of the residual was changed to match the new speech period, by adding zeros to both ends symmetrically when decreasing F0 or shortening the residual for increasing F0. Finally, the original LPC filter was excited with the modified residual. Duration modification was implemented by repetition (for time expansion) or deletion (for time compression) of individual frames. Frames were selected by linear sampling throughout the duration of the phoneme. Energy modification was carried out by first calculating gain factors for each frame, given by the specified energy divided by the existing energy of that frame. The gain factor trajectory was first filtered by a tenth-order median filter and then smoothed using a zero-phase low-pass filter (five-point symmetric, normalized Hanning window). The intent of these filters was to preserve existing frame-to-frame energy fluctuations, while implementing the desired longer-term energy changes. Finally, the waveforms of each frame were scaled by their associated gain factor.

For the overlap-add operation, each frame was windowed and the second half of one frame was added to the first half of the next frame. The window was an asymmetric trapezoidal given by

$$w(t|l,r)=\begin{cases}0 & \text{if } t=1\cdots\dfrac{l}{2}\\ \dfrac{t-l/2}{l/2+1} & \text{if } t=\left(\dfrac{l}{2}+1\right)\cdots l\\ 1, & \text{if } t=(l+1)\cdots\left(l+\dfrac{r}{2}\right)\\ \dfrac{t-l-r/2}{r/2+1} & \text{if } t=\left(l+\dfrac{r}{2}+1\right)\cdots(l+r),\end{cases} \tag{1}$$

where $t=1\cdots l+r$ represents time and l and r are the lengths of the left and right speech periods of a frame. The advantage of using this trapezoidal window is that it avoids LPC filter startup artifacts by discarding initial sample values, and that it continues to "ring" beyond the original excitation, desirable features when using shortened or lengthened (zero-padded) residuals.

The hybridization algorithm can be seen as a LPC resynthesis of either CNV or CLR speech, while optionally modifying energy, F0, and phoneme durations, as well as inserting or deleting phonemes. For example, the condition HYB-SPN is equivalent to a resynthesis of CLR speech with energy, F0,

and phoneme durations modified to match the corresponding acoustic features extracted from the CNV speech. Conversely, this configuration can be thought of as a resynthesis of CNV speech where frame waveforms have been replaced with the corresponding and appropriately selected and modified frame waveforms from CLR speech, in addition to inserting or deleting waveform segments as required by the CLR speech labels during the parallelization process.

The hybridization algorithm shares some characteristics with voice transformation (VT) algorithms. During training, VT algorithms estimate parameters of a mapping function that predicts target features from time-aligned source features, extracted from a target and a source speaker, respectively. During transformation, a new source speaker utterance is analyzed, its features transformed by the trained mapping function, and a new speech signal is synthesized with characteristics of the target speaker. Applications include speaker modification (Kain, 2001), improving the intelligibility of dysarthric speech (Kain *et al.*, 2007), and speech morphing (Abe, 1996). However, the current work is distinct from VT algorithms in that its goal is the study of feature selection and the accompanying changes in intelligibility. Moreover, as the CLR speech features are known, there is no training function and no mapping in the current work.

IV. PERCEPTUAL EXPERIMENTS

Three experiments were conducted to test intelligibility differences as a function of speech material and hybrid configurations. The conditions of the second and third experiments depended on the results from the previous experiment, and so each experiment employed a unique set of listeners and should be considered in sequence. Intelligibility results between experiments are not directly comparable because the set of listeners varied between experiments. Because of this variability between listener groups, we restrict our comparison of results, other than noting significance, to within-experiment results.

Each perceptual test was carried out on a fanless computer with an external sound card (M-Audio USB Duo). To ensure that experimental stimuli were administered at controlled fixed sound pressure levels while minimizing ambient sounds, stimuli were presented through headphones in a quiet room. The speech signal was calibrated acoustically at 65 dBA, as measured in a custom-made coupler with a sound-level meter (Brüel and Kjær Type 2238) attached to a condenser microphone (Brüel and Kjær Type 4133). The signal level was averaged over 10 s of continuous speech, excluding pauses. Throughout the study, whenever noise was added to speech, the energy of the speech signal was kept constant, while the energy of the noise was varied. The noise consisted of near-field microphone recordings of 12 speakers that were combined digitally, called "babble" noise (Bilger and Nuetzel, 1984).

A. Listeners

Men and women aged 18–39 with self-reported normal hearing participated in the experiments described in this section. The first language of all listeners was North-American English. Every perceptual test was carried out using unique listeners.

B. Obtaining SNR50% values

The term SNR50% (Chang *et al.*, 2006) refers to that signal-to-noise ratio (SNR) at which listeners can correctly identify key words in the presence of noise 50% of the time. Experiments were carried out with the noise level set to each listener's individual SNR50% level, to help normalize for differences in hearing performance between listeners.

The approach to estimate SNR50% levels was based on the "up-down" method (Levitt, 1971). The details of the procedure were as follows: Initially, the SNR was set to −3 dB. The first sentence was repeated at increasing SNR levels until the listener could obtain the correct response. Four keywords were required to be recalled correctly for a sentence to be considered correct. After the first correct response, a different sentence was presented each time. The noise level was increased (SNR decreased) when the listener's response was correct, and the noise level was decreased (SNR increased) when the response was incorrect. A correct response followed by an incorrect response, or vice versa, was counted as one reversal. The procedure was continued until eight reversals took place. Finally, the SNR50% value was computed by averaging SNR levels from reversals 3 to 8.

C. Speech corpus verification

In a first preliminary test, the intention was to verify that all (unmodified) CNV sentences (in both SD and LD partitions) were, in fact, intelligible in the absence of background noise. Four listeners listened to all sentences from materials A and B (140 sentences total). A test administrator measured key word identification, scoring a sentence as correct when five out of five words were identified correctly. The resulting sentence-level intelligibility rates were 98.57% and 92.50% for materials A and B, respectively. This confirmed that the CNV sentences were generally intelligible. It was assumed that the intelligibility of CLR sentences would be even higher.

In a second preliminary test, speaking styles and material types were compared by determining SNR50% values, using the procedure described in Sec. IV B. Eight listeners first listened to CNV sentences of materials A and B from the SD partition, and then listened to CNV and CLR sentences of both material types from the LD partition in a Latin Square design. The size of the SD partition did not allow testing of two styles; therefore, only the CNV style was evaluated, as in the initial SNR50% tests of the main intelligibility experiments. SNR50% values for each of these six conditions were determined separately. An administrator measured key word identification, scoring a sentence as correct when four out of five words were identified correctly. Results are shown in the first two rows of Table III. The data of the LD partition were submitted to a 2×2 (speaking styles CLR and CNV, speech materials A and B) analysis of variance. There were significant ($\alpha=0.05$) main effects of speaking style ($F(1,28)=7.32$, $p=0.012$) and sentence mate-

TABLE III. Listeners' average SNR50% levels in dB, with SDs in parentheses, for materials A and B in partitions SD and LD. Results are from the speech corpus verification (V) test, as well as from the three intelligibility experiments (E).

Test	A-SD	B-SD	A-LD	B-LD
V. CNV	−0.71(1.18)	−0.39(1.68)	−2.52(0.84)	−1.53(1.50)
V. CLR	n/a	n/a	−4.26(1.25)	−2.54(1.94)
E. 1 CNV	0.58(0.88)	1.81(1.78)	n/a	n/a
E. 2 CNV	−0.24(1.11)	−0.07(1.39)	n/a	n/a
E. 3 CNV	0.22(1.44)	0.30(1.74)	n/a	n/a

rial ($F(1,28)=7.1$, $p=0.013$). Results between the partitions cannot be compared because the SD partition was always tested first. No significant interactions between speaking style and sentence material were found ($p=0.475$). This test shows that listeners could more easily identify words spoken in the CLR style. Moreover, identification was easier for material A, as expected. These results are consistent with the results from previous studies (Picheny *et al.*, 1986) and confirmed that the speech corpus reflects the inherent differences between CLR and CNV speech.

D. Intelligibility experiment 1

1. Stimuli and procedures

The search for acoustic features that are relevant to speech intelligibility can be performed in many ways. Two approaches are "top-down" (starting with the entire space of candidate acoustic features and splitting into smaller sets) and "bottom-up" (starting with individual candidate acoustic features and recombining into larger sets). In pilot studies (Kusumoto *et al.*, 2007, experiment 1), the former approach was used to test a prosodic (HYB-EFDN) versus spectral (HYB-SP) hybridization configuration, without success. One hypothesis as to the failure of those hybrids was that spectral features and duration features cannot be separated without severely impacting intelligibility, due to coarticulation. Therefore, the purpose of experiment 1 was to examine the intelligibility of a HYB speech condition for which spectral and duration features were grouped together, namely, HYB-DSP (see Table I), which replaced phoneme duration, spectrum, and phoneme sequence from CNV speech with those features from CLR speech. A second condition, HYB-EFN, replaced energy, F0, and nonspeech features of CNV speech with those from CLR speech. To calculate the DOC, the intelligibility of the CNV and CLR speech was also measured.

Forty-eight sentences from the LD partitions of materials A and B were used, respectively (96 sentences total), processed in one of the four conditions (12 sentences per condition, per material). The experiment was carried out in a 2×4 Latin Square design (two speech material types and four conditions). Each sentence was heard only once by a listener. The order of the sentences was randomized, and kept the same for all of the subjects. Sentences were counterbalanced, i.e., for conditions *A*, *B*, *C*, and *D*, and a block of consecutive listeners *L*, *L*+1, *L*+2, *L*+3, each sentence was heard in condition *A* by listener *L*, in condition *B* by listener *L*+1, in condition *C* by listener *L*+2, and in condition *D* by listener *L*+3. For each listener, a test administrator first measured the listener's SNR50% for the CNV sentences of the SD partitions of materials A and then B. The resulting values were then used during the immediately following intelligibility test by adjusting the energy of the noise waveform to the measured SNR50% level for that material. Material A was presented in a first session, and Material B in a second session, with a short break between the two sessions. Sentences were considered correct if listeners recalled four out of five key words (both for the initial SNR50% test and the main intelligibility test).

Eight listeners aged 23–39 participated in the experiment. Listeners were informed that they were going to hear syntactically correct sentences of two types, semantically correct and semantically anomalous. They were given one written example sentence of each type. They were instructed to repeat each sentence aloud to the best of their ability and to guess when unsure or when they could not assign meaning to a sentence.

2. Results and discussion

Listeners' average SNR50% levels in the CNV condition for materials A and B were 0.58 and 1.81 dB, respectively (see also Table III). Figure 1(a) shows listeners' raw average intelligibility scores and SDs. In order to determine statistical significance, the raw scores were subjected to the arcsine transformation (Anscombe, 1948) given by the equation

$$x = \arcsin\sqrt{\frac{r+3/8}{n+3/4}}, \tag{2}$$

where r represents the number of sentences a listener recalled correctly, and n represents the number of sentences presented. This transformation was used because the raw intelligibility score differences in percentages are not comparable in a probabilistic sense, especially at the lower and upper ends of the scale (Studebaker, 1985). The intelligibility of CLR speech was significantly ($\alpha=0.05$) increased compared to the intelligibility of CNV speech, using planned, pairwise, two-tailed *t*-tests ($p<0.001$ and $p=0.036$ for materials A and B, respectively). The hybridized condition HYB-DSP yielded a significant increase in intelligibility over the baseline CNV speech ($p=0.003$ and $p<0.001$ for materials A and B, respectively). On the other hand, HYB-EFN did not show a significant difference over CNV levels ($p=0.229$ and $p=0.311$ for materials A and B, respectively).

The DOC was defined as

$$\mathrm{DOC} = \frac{I_{\mathrm{HYB}} - I_{\mathrm{CNV}}}{I_{\mathrm{CLR}} - I_{\mathrm{CNV}}}, \tag{3}$$

where I represents intelligibility levels in percent and the subscript refers to a specific condition. Thus, the DOC of the HYB-DSP condition was 92% and 146% for materials A and B, respectively. The DOC is, of course, subject to listener variation, and DOC values are not precise metrics but are provided to give a sense of how much impact a feature has.

These results indicated that, for this speaker, the combination of phoneme durations, short-term spectrum, and pho-

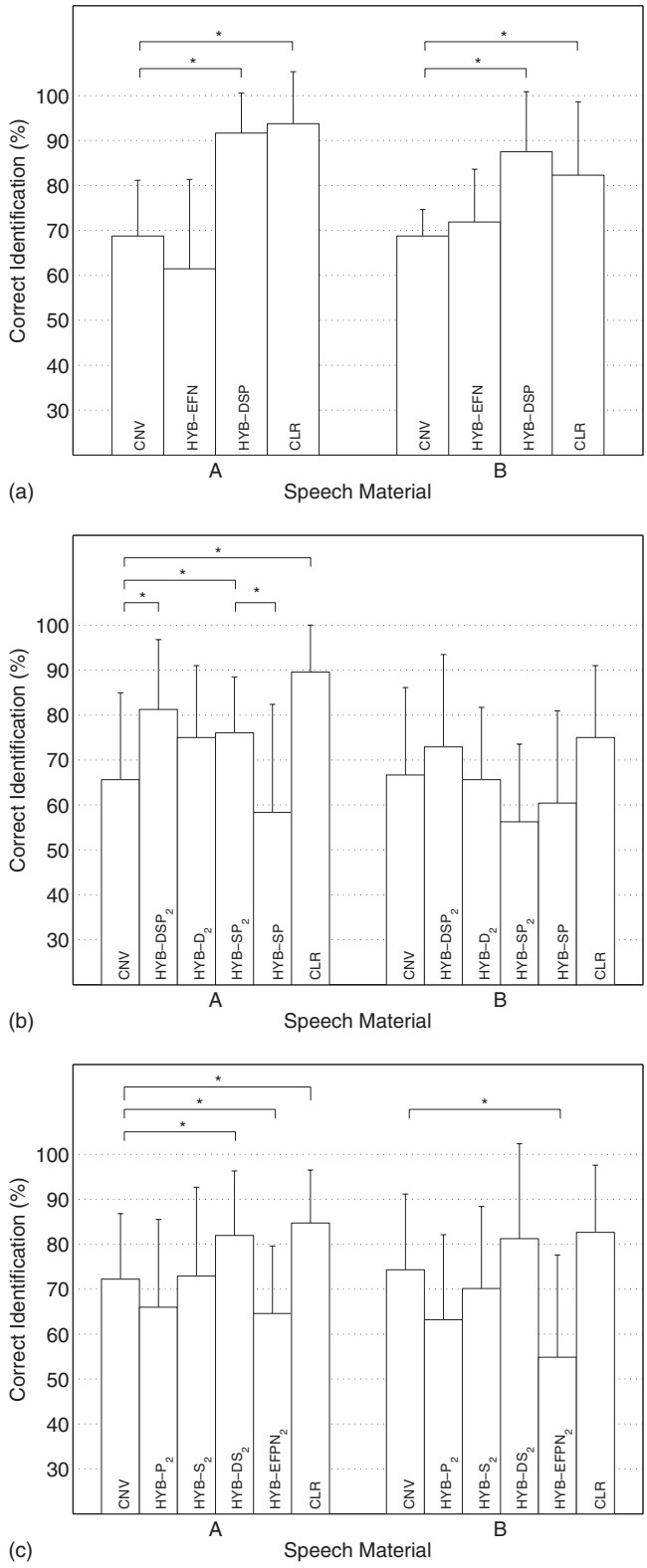


FIG. 1. Raw intelligibility scores for CNV, CLR, and HYB conditions. The error bars indicate the extent of the SDs. The results marked with an asterisk are significantly different (α=0.05).

neme sequence from CLR speech carried most of the information contributing to intelligibility. The combination of prosodic features F0, energy, and nonspeech events (e.g., pauses) did not appear to contribute to intelligibility. However, it was not possible to determine whether duration and spectrum must both be present in the HYB feature set, because the two features were not tested independently; this was further addressed in experiment 2.

E. Improving the algorithm implementation

To further improve speech modification quality, sources of signal processing artifacts were identified and addressed through changes to the hybridization algorithm, leading to a newer implementation. Implementation 2 addressed the following four possible sources of signal-processing artifacts that were hypothesized to impact speech intelligibility and quality.

(1) *Preserving glottalization*. Normally, auxiliary marks were added to the collection of time marks whenever the distance between pitch marks was larger than a minimum threshold (see Sec. III B 3). However, voiced regions may at times have even lower fundamental frequencies than the minimum threshold if the speech sound is glottalized, a phenomenon that occurs when the vocal folds vibrate irregularly or with very low frequency. Therefore, phoneme identity was used to force voiced phonemes to be excluded from having auxiliary marks placed within their region, allowing for very low fundamental frequencies. Specifically, vowels, nasals, flaps, and approximants were excluded from having auxiliary marks added. As a result, the hybridization algorithm could preserve glottalization that was present in an original waveform.

(2) *Preventing artifacts during bursts*. During synthesis, modifications to duration or F0 (when combining spectral information from one style with duration and/or F0 information from another style) may require the duplication of frames. Duplicating frames that contain bursts can cause artifacts because short, single, impulselike acoustic events may sound unnatural when perceived more than once in short succession. Therefore, frames of affricates, plosives, and flaps were prevented from being candidates for duplication; instead, neighboring nonplosive frames were used for achieving the desired duration or F0. Effectively, no duration modification was performed for the phonemes listed above. While this makes the hybridization somewhat less precise, it was hypothesized that the improved perceptual quality would lead to improved intelligibility.

(3) *Preventing artifacts during voicing transitions*. Duplicating frames in unvoiced-to-voiced transitions or in voiced-to-unvoiced transitions may cause artifacts, because these transitions contain unique events that, when duplicated, may be perceived as unnatural. More precisely, these transitory frames are special in terms of their energy, period length, and formant frequency dynamics. Therefore, duplication in frames near unvoiced-to-voiced transitions or voiced-to-unvoiced transitions was prevented; instead, neighboring frames from more steady-state regions were used for processing, similar to the previously described approach to preventing artifacts during bursts.

(4) *Smoothing phoneme insertions and deletions*. Phoneme insertions and deletions, as required by the waveform parallelization stage (see Sec. III B 2), may cause signal discontinuities at phoneme boundaries, which may result in audible clicks, possibly reducing intelligibility. Therefore, during phoneme insertion and deletion operations, all required waveforms were faded in and out smoothly (cross-faded), using linearly weighted windows with durations of approximately 1 ms, centered at the concatenation points.

In summary, implementation 2 allows the occurrence of very low F0 due to glottalization in voiced sounds, prevents frame duplication near voicing transitions or during plosive speech regions such as bursts, and smoothly fades in and out of phoneme insertions and deletions.

F. Intelligibility experiment 2

1. Stimuli and procedures

As observed in experiment 1, the combination of phoneme duration, spectrum, and phoneme sequence of CLR speech significantly improved intelligibility for this speaker, relative to CNV speech. To further test the hypothesis that phoneme duration cannot be separated from spectral features while maintaining high intelligibility (see Sec. IV D), this second experiment used phoneme duration only; spectral features and phoneme sequence; and duration, spectral, and phoneme sequence features from CLR speech. To test the hypothesis that artifacts introduced by the hybridization algorithm can cause decreased intelligibility, the intelligibility of HYB speech produced by implementation 2 was compared with HYB speech that was produced by implementation 1.

Six stimulus conditions were tested in experiment 2. The conditions were CNV speech, CLR speech, and four HYB speech conditions, namely, HYB-DSP_2, HYB-D_2, HYB-SP_2, and HYB-SP, where the subscript "2" indicates hybridization algorithm implementation 2, and the lack of a subscript indicates implementation 1. The HYB-DSP_2 condition examined the combined effects of phoneme duration, spectral features, and phoneme sequences from CLR speech. The HYB-D_2 condition examined the effect of using only phoneme durations from CLR speech. The HYB-SP_2 condition examined the combination of spectral features and phoneme sequences from CLR speech. Finally, the HYB-SP condition was the same in all respects as condition HYB-SP_2, except that it was generated using implementation 1.

As in the previous experiment, a total of 96 sentences from the LD partitions of materials A and B were tested, but this time processed in each of the six conditions described above. The experiment was carried out in a 2×6 Latin Square design (two speech material types and six conditions). Procedures, including SNR50% testing, were identical to experiment 1 (see Sec. IV D 1), except that the number of conditions was six (eight sentences per condition, per material). Twelve listeners aged 19–39 participated.

2. Results and discussion

Listeners' average SNR50% levels in the CNV condition for materials A and B were −0.24 and −0.07 dB, respectively (see Table III). Figure 1(b) shows average intelligibility levels and SDs. Results were transformed using the arcsine transformation [Eq. (2)] and analyzed for statistical significance using a *t*-test, in the same manner as experiment 1. For material A, the intelligibility of CLR speech was significantly higher than the intelligibility of CNV speech ($p<0.001$). However, in material B, a significant difference between CNV and CLR speech was not shown ($p=0.277$). Therefore, the lack of a significant increase between CNV and any hybrid condition for material B was an expected result, as the CLR speech intelligibility level represents the maximum expected performance of HYB speech intelligibility, with the exception of possible listener variation. Planned tests that did not have significant results were not further analyzed. It was speculated that the lack of a significant intelligibility increase of CLR speech may be due to the increased comprehension difficulty of this text material, test-retest reliability, the current group of listeners, or increased variance in response to semantically anomalous sentences.

For material A, a pairwise comparison between HYB-DSP_2 and CNV ($p=0.010$), and between HYB-SP_2 and CNV ($p=0.045$) showed a significant improvement in intelligibility. A comparison between HYB-D_2 and CNV ($p=0.139$) did not show a significant difference. A comparison between HYB-SP_2 and HYB-SP ($p=0.007$) indicated a significant difference in intelligibility favoring implementation 2. The DOCs of the HYB-DSP_2 and HYB-SP_2 conditions were 72% and 56%, respectively.

It appears that, for this speaker, HYB speech that combines spectral, duration, and phonetic sequence features from CLR speech had greater intelligibility than HYB speech from either spectral and phonetic sequence or merely duration features; however, significance testing was not planned for these comparisons. The combination of spectral, duration, and phoneme sequence features, as well as the combination of spectral and phoneme sequence features, yielded significant improvements in intelligibility over CNV speech. Moreover, a significant difference in intelligibility between the two implementations of the hybridization algorithm was found.

G. Intelligibility experiment 3

1. Stimuli and procedures

To determine whether phoneme insertions and deletions between CNV and CLR were significant for intelligibility, the following six stimulus conditions were tested: CNV speech, CLR speech, and the four HYB speech conditions HYB-P_2, HYB-S_2, HYB-DS_2, and HYB-$EFPN_2$. The HYB-P_2 condition examined the effects of using CLR phoneme sequences. The HYB-S_2 condition examined the effect of using only spectral features from CLR speech, similar to the output of a spectral VT system (e.g., Kain and Macon, 1998) with perfect mapping. The HYB-DS_2 condition examined the combined effects of phoneme durations and spectral features from CLR speech. Finally, the HYB-$EFPN_2$ condi-

tion examined the combined effects of energy, F0, phoneme identity, and nonspeech events from CLR speech; this is equivalent to starting out with CLR speech and replacing spectral features and phoneme durations with those from CNV. Experimental conditions were identical to experiment 2, with 18 listeners aged 18–35 participating.

2. Results and discussion

Listeners' average SNR50% levels in the CNV condition for materials A and B were 0.22 and 0.30 dB, respectively (see Table III). Figure 1(c) shows average intelligibility levels and SDs. Results were transformed using the arcsine transformation [Eq. (2)] and analyzed for statistical significance using a *t*-test, in the same manner as previous experiments. For material A, the intelligibility of CLR speech was significantly higher than the intelligibility of CNV speech ($p<0.003$). Similar to experiment 2, material B did not show a significant difference between CNV and CLR speech ($p=0.175$). For material A, comparisons between CNV and hybrid conditions HYB-P_2 and HYB-S_2 showed no significant differences. A comparison between CNV and HYB-DS_2 showed a significant increase in intelligibility ($p=0.050$); the DOC was 78%. In contrast, a comparison between CNV and HYB-$EFPN_2$ showed a significant decrease in intelligibility ($p=0.044$) for both materials.

These results suggest that, for this speaker, the higher intelligibility of CLR speech for material A is not due to either the CLR phoneme sequence or CLR spectrum alone, or due to the combined effects of energy, F0, phoneme sequence, and nonspeech events including pauses. However, the combined effects of spectral features and phoneme durations yielded a significant intelligibility increase compared to CNV speech.

V. CONCLUSION AND FUTURE WORK

This article presented experimental results of sentence-level improvements to the intelligibility of CNV speech, by using hybrid stimuli which combine certain features of a CNV speech sentence with complementary features of the same CLR speech sentence. The condition HYB-DSP, for example, is equivalent to a LPC resynthesis of CLR speech with F0 and energy trajectories modified to match F0 and energy trajectories of CNV speech; additionally, nonspeech parts (pauses) were removed, according to the labels of the CNV speech of that same sentence. Listening tests estimated the significance and DOC of acoustic features, using normal-hearing listeners aged 19–41 and 12-speaker babble noise.

There are two main conclusions from this work. First, it is possible to create HYB speech that combines aspects of CNV and CLR speech to create a signal with greater intelligibility than the original CNV speech. Second, the results indicate that the two main sources of increased intelligibility of CLR speech for this speaker and material A are in the spectrum and duration, and not in the pausing patterns, F0, energy, or phoneme sequence. In particular, for material A, experiment 1 showed that the feature combination DSP was significant for improved intelligibility, while the combination EFN was not. Experiment 2 then showed that while the combination DSP was significant, and SP alone provided a significant contribution, duration alone was not considered significant. Experiment 3 then showed that DS alone was significant for improved intelligibility, but that neither spectrum nor phoneme sequence were significant by themselves. (Future experiments will be required to better assess the interactions between spectrum, duration, and phoneme sequence.) These results present the first known study in which sentence-level intelligibility of CNV speech has been improved by application of a subset of CLR speech features.

We speculate that the spectrum is important for intelligibility because of the general differences in the size of the vowel space between CLR and CNV speech (Picheny *et al.*, 1986). In addition, spectral tilt (included in our definition of spectrum) may also play a role in intelligibility. Duration, and in particular the combination of duration and spectrum, may be important to intelligibility because of formant dynamics. If formants are stretched or compressed in a way that is unnatural, the unusual formant dynamics may negatively impact intelligibility. Also, relative duration may affect phoneme characteristics such as the voicing of stops or the tense/lax quality of a vowel, and so control of duration may be important when matched with acoustic cues such as periodicity or formant locations.

Differences in intelligibility scores were more pronounced for material A than for material B. It was conjectured that the difference in intelligibility levels between materials may be caused by easier comprehension of material A, characterized by CNV and CLR psychometric functions with steeper slopes, leading to larger differences at or close to SNR50%. A possible increase in the variability of listener responses to semantically anomalous sentences may also be a contributing factor. The lack of significant differences between CNV and CLR for material B in experiments 2 and 3 prevented comparisons between CNV and HYB conditions in those experiments.

It is important to note that these results are for one speaker only and cannot be generalized to the larger population, since speakers use different strategies for speaking clearly (Ferguson and Kewley-Port, 2002; Ferguson, 2004; Krause and Braida, 2004). In addition, results may be specific to the particular corpus used.

The hybridization algorithm is equivalent to modifying CNV speech with an "oracle" mapping function, thus simulating maximum performance levels of an automatic modification system. An automatic modification system with a trained mapping function has potential to improve speech intelligibility for hearing-impaired listeners, such as individuals with hearing loss or in noisy environments. Speech modification of duration may be undesirable in live environments where speakers are both audible and visible, because of the resulting asynchrony between lip movements and auditory events. However, duration modification may be useful in live audio-only settings such as telephone-based speech. In such applications, long-term asynchrony might be compensated for by either additional feedback requesting the speaker to pause when necessary, or automatically modifying the length of pauses to compensate for changed speech durations. Moreover, recorded or bufferable content (such as

hard-disk-based viewing of TV shows) could be enhanced by modifying the video portion of the signal as well, preserving perceptually relevant synchronisms.

Future work will use the current methods for determining the importance of features such as formant dynamics, spectral tilt, and relative duration, using several speakers and several listener types (different age groups and hearing performances). The size of the units studied may also be varied. For example, spectral changes were applied at the phoneme level in the current study. Larger spectral units [e.g., syllable-level features, used in voice transformation (Rao *et al.*, 2007)] or smaller units (e.g., phoneme-transition regions) may also yield information of interest. This work may lead to objective measures of speech intelligibility and automatic speech intelligibility enhancement systems. The automatic mapping of CNV speech features to more closely resemble CLR speech features will not be a research focus until it is better known which specific acoustic features are relevant for the improved speech intelligibility of CLR speech.

ACKNOWLEDGMENTS

The authors thank Jan P. H. van Santen of OHSU, Marjorie R. Leek and Michelle Molis of the National Center for Rehabilitative Auditory Research at the Portland Veterans Affairs Medical Center for insightful discussions and advice, as well as the reviewers for their helpful comments and suggestions.

Abe, M. (**1996**). "Speech morphing by gradually changing spectrum parameter and fundamental frequency," in Proceedings of ICSLP, Philadelphia, PA, Vol. **4**, pp. 2235–2238.

Anscombe, F. J. (**1948**). "The transformation of poisson, binomial and negative-binomial data," Biometrika **35**, 246–254.

Bilger, R. C., and Nuetzel, J. M. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phonetic Sci. **17**, 97–110.

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. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Krause, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Bradlow, A. R., Torretta, B. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Center for Spoken Language Understanding (**2007**). "Clear Speech JPH," http://www.cslu.ogi.edu/corpora (last viewed April, 2008).

Chang, J. E., Bai, J. Y., and Zeng, F.-G. (**2006**). "Unintelligible Low-Frequency Sound Enhances Simulated Cochlear-Implant Speech Recognition in Noise," IEEE Trans. Biomed. Eng. **53**, 2598–2601.

Cruickshanks, K., Wiley, T., Tweed, B., Klein, B., Klein, R., Mares-Perlman, J., and Nondahl, D. (**1998**). "The prevalence of hearing loss in older adults," Am. J. Epidemiol. **148**, 879–885.

Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York).

Ferguson, S. H. (**2004**). "Talker differences in clear and conversatoinal speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Gordon-Salant, S. (**1986**). "Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing," J. Acoust. Soc. Am. **82**, 1599–1607.

Gordon-Salant, S. (**1987**). "Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 1199–1202.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., and Fitzgibbons, P. J. (**2001**). "Sources of age-related recognition difficulty for time-compressed speech," J. Speech Lang. Hear. Res. **44**, 709–719.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Am. Acad. Audiol **116**, 3108–3118.

Hazan, V., and Simpson, A. (**1998**). "The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise," Speech Commun. **24**, 211–226.

Helfer, K. S. (**1998**). "Auditory and auditory-visual recognition of clear and conversational speech by older adults," J. Am. Acad. Audiol **9**, 234–242.

Hosom, J. (**2002**). "Automatic phoneme alignment based on acoustic-phonetic modeling," in Proceedings of ICSLP, Boulder, CO, Vol. **1**, pp. 357–360.

International Electrotechnical Commission (**2002**). "Electroacoustics-sound level meters—Part 1: Specifications," Paper No. 61672.

Kain, A. (**2001**). "High resolution voice transformation," Ph.D. thesis, Oregon Graduate Institute, Portland, OR.

Kain, A., Hosom, J.-P., Niu, X., van Santen, J., Fried-Oken, M., and Staehely, J. (**2007**). "Improving the intelligibility of dysarthric speech," Speech Commun. **49**, 743–759.

Kain, A., and Macon, M. (**1998**). "Spectral voice conversion for text-to-speech synthesis," in Proceedings of ICASSP, Vol. **1**, pp. 285–288.

Krause, J. C., and Braida, L. D. (**2002**). "Investigation alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **15**, 362–378.

Kusumoto, A., Kain, A., Hosom, J.-P., and van Santen, J. (**2007**). "Hybridizing conversational and clear speech," in Proceedings of Interspeech.

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

Liu, S., Rio, E. D., Bradlow, A. R., and Zeng, F. G. (**2004**). "Clear speech perception in acoustic and electric hearing," J. Acoust. Soc. Am. **116**, 2374–2383.

Liu, S., and Zeng, F. G. (**2006**). "Temporal properties in clear speech perception," J. Acoust. Soc. Am. **120**, 424–432.

Nejime, Y., and Moore, B. C. J. (**1998**). "Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss," J. Acoust. Soc. Am. **103**, 572–576.

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.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

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.

Rabiner, L., and Juang, B. H. (**1993**). *Fundamental of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).

Rao, K. S., Laskar, R. H., and Koolagudi, S. G. (**2007**). "Voice transformation by mapping the features at syllable level," in Pattern Recognition and Machine Intelligence, Heidelberg, Germany, Vol. **4815**, pp. 479–486.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silberger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE Recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Salthouse, T. A. (**1996**). "The processing-speech theory of adult age differences in cognition," Psychol. Rev. **103**, 403–428.

Schum, D. J. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol **7**, 212–218.

Studebaker, G. A. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Taylor, P., Black, A., and Caley, R. (**1998**). "The architecture of the festival speech synthesis system," in Proceedings of the Third International Workshop on Speech Synthesis, Sydney, Australia.
Uchanski, R. M., Choi, S. 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.
Vaughan, N., Storzbach, D., and Furukawa, I. (**2006**). "Sequencing and nonsequencing working memory in understanding of rapid speech by older listeners," J. Am. Acad. Audiol **17**, 506–518.
Wingfield, A., Tun, P. A., and McCoy, S. L. (**2005**). "Hearing loss in older adulthood. What it is and how it is interacts with cognitive performance," Current directions in Psychological Science **14**, 144–148.

Neuronal periodicity detection as a basis for the perception of consonance: A mathematical model of tonal fusion

Martin Ebeling[a)]
Peter-Cornelius-Conservatory of Music, Mönchengladbach D-55122 Mainz, Germany

(Received 7 December 2007; revised 5 July 2008; accepted 13 July 2008)

A mathematical model is presented here to explain the sensation of consonance and dissonance on the basis of neuronal coding and the properties of a neuronal periodicity detection mechanism. This mathematical model makes use of physiological data from a neuronal model of periodicity analysis in the midbrain, whose operation can be described mathematically by autocorrelation functions with regard to time windows. Musical intervals produce regular firing patterns in the auditory nerve that depend on the vibration ratio of the two tones. The mathematical model makes it possible to define a measure for the degree of these regularities for each vibration ratio. It turns out that this measure value is in line with the degree of tonal fusion as described by Stumpf [Tonpsychologie (Psychology of Tones) (Knuf, Hilversum), reprinted 1965]. This finding makes it probable that tonal fusion is a consequence of certain properties of the neuronal periodicity detection mechanism. Together with strong roughness resulting from interval tones with fundamentals close together or close to the octave, this neuronal mechanism may be regarded as the basis of consonance and dissonance.
 [DOI: 10.1121/1.2968688]

PACS number(s): 43.75.Cd, 43.60.Ek, 43.64.Bt, 43.66.Ba [DD] Pages: 2320–2329

I. INTRODUCTION

A. The consonance theories of von Helmholtz and Stumpf

In the second half of the 19th century, two important theories of consonance were established. Von Helmholtz (1877) proposed a theory based on roughness, an unpleasant sensation produced by rapid sound fluctuations. He argued that roughness between partials of two tones forming a musical interval unconsciously leads to the sensation of dissonance. However, if two partials coincide, they cannot cause roughness. Thus, if many partials coincide, only a slight roughness is evoked, and, conversely, if many partials do not coincide, the interval sounds rough. Von Helmholtz (1877) believed that roughness explains Phythagoras' rule: In the case of simple vibration ratios, many partials coincide and the interval is consonant, but in the case of complex vibration ratios, only a few partials coincide, and the interval sounds dissonant.

In contrast to von Helmholtz's psychoacoustical approach, the German psychologist Stumpf (1930) tried to give a psychologically motivated definition of consonance. He had observed that consonant intervals show a tendency to cohere into a single sound image, which he called an entity ("Einheit," Stumpf, 1890). He termed this phenomenon tonal fusion ("Tonverschmelzung," Stumpf, 1890). Consonant intervals show a stronger tendency to fuse than less consonant or dissonant intervals. Stumpf (1930), however, could not give a psychoacoustical or physiological explanation for this phenomenon, so roughness-based concepts of dissonance and consonance are still currently widely accepted in many models (e.g., Terhardt, 1976, 1977, 1984; Plomp and Levelt, 1965; and Zwicker and Fastl, 1999; see Burns and Ward, 1982). Unfortunately, these models fail to explain consonance and dissonance in musical intervals without any roughness, such as intervals of pure tones or of primaries with only few partials. This gap is sometimes bridged by a learning hypothesis ("Lernhypothese," Terhardt, 1976, 1977; see Burns and Ward, 1982) which states that in the case of pure tones, the perception of consonance and dissonance and of the specific interval quality has been learned beforehand from musical intervals with complex tones. Von Helmholtz (1877) argued that in the case of pure component intervals, roughness is evoked by ear overtones. However, Plomp (1965) demonstrated that "for usual listening levels, the ear's distortion is sufficiently low to avoid audible combination tones" (Plomp, 1965, p. 1123). This contradicts Helmholtz's hypothesis and within all theories based on this hypothesis.

B. Neuronal code and pitch

At the end of the 19th century, scientists had little knowledge about auditory processing in the brain. However, since then, many insights into the neuronal code and processing of auditory signals have been gained. For example, neurophysiologists can demonstrate that pitch is coded in the interspike intervals (ISIs) of neuronal firing patterns in the auditory nerve and brain stem (Rhode, 1995). The inner ear provides a frequency analysis mechanism which transforms incoming sound into a neuronal code (Zhang *et al.*, 2001). Due to the mechanics of the basilar membrane and the frequency selectivity of the hair cells, sound induced pressure waves traveling through the *cochlea* are converted into neural impulses, representing the different resolvable frequency components of the sound (Goldstein and Scrulovic, 1977).

[a)]Electronic mail: mar.ebeling@arcor.de

In the case of a single pure tone, the traveling pressure wave maximally activates hair cells at a certain place on the basilar membrane and causes them to react with a periodic firing pattern. According to the volley principle of Wever (1949), this results in a running spike train in the auditory nerve with a period corresponding to the reciprocal of the frequency of the tone. The period is equal to the distance of two adjacent spikes. As the pitch becomes higher, the distance becomes smaller. One may say that neuronally, pitch corresponds to the period of the tone. The described mechanism shows that the fibers respond to the instantaneous phase of the motion of the basilar membrane. At higher frequencies, above about 4 Hz, this synchrony disappears. (Zwicker and Fastl, 1999). However, those high frequencies are beyond the range of musically relevant pitches.

Langner (2005) assumed "… since all frequency components (harmonics) of a harmonic sound are multiples of its frequency, the period of the fundamental is also encoded in the cochlea in amplitude modulations resulting from superposition of frequency components above the third harmonic. As a consequence, the period of the fundamental is coded temporally in spike intervals in the auditory nerve and analysed by neurons in the auditory brain stem (cochlear nucleus: CN) and midbrain (inferior colliculus: IC)."

The time between neural spikes, the so-called ISI, can be measured either between successive discharges ("first-order ISI") or between both successive and nonsuccessive spikes ("all-order ISI"). Counting all ISIs in a discharge pattern leads to histograms that show the ISI distributions recorded from the auditory nerve. Cariani and Delgutte (1996) have shown that ISI-histograms ("autocorrelograms") computed from all-order ISIs show high peaks for periods corresponding to the pitch. This demonstrated that "the most frequent all-order ISI corresponded to the pitch heard" (Cariani and Delgutte, 1996, p. 1698). This rule of pitch estimation holds for lower frequencies. In cats, Cedolin and Delgutte (2005) found an upper boundary of about 1300 Hz for pitch estimation from pooled ISI distributions and proposed a complementing rate-place profile for frequencies higher than 400 Hz.

As pitch perception is closely related to the perception of musical intervals, it is obvious that the neuronal firing patterns evoked by consonant and dissonant intervals should be studied to understand the phenomenon of consonance. Thus, Tramo *et al.* (2001) analyzed the neural responses to harmonic intervals. They used stimuli of isolated harmonic intervals formed by complex tones. Each of the two complex tones in the musical intervals (minor second, perfect fourth, triton, and perfect fifth) contained the first six harmonics with equal amplitude and equal phase.

They observed that for consonant intervals, the fine timing of auditory nerve fiber responses contains strong representations of harmonically related pitches implied by the interval (e.g., Rameau's fundamental bass) in addition to the pitches of notes actually present in the interval. Moreover, all or most of the partials can be resolved by finely tuned neurons throughout the auditory system. This finding can be compared with the results of psychoacoustic experiments testing the audibility of partials in complex tones. Moore and Glasberg (1986) measured the frequency of difference limens (DLFs) for the individual components in a complex tone. "For a complex tone containing the first 12 harmonics at equal amplitude, the DLFs were small (between 0.2 and 1.0%) for harmonic number up to four, but the DLFs increased rather abruptly around the fifth to seventh harmonic" (Moore and Glasberg, 1986, p. 283).

By contrast, for dissonant intervals, Tramo *et al.* (2001) observed auditory nerve fiber activity that does not contain strong representations of constituent notes or related bass notes. Many partials of the two complex tones are too close together to be resolved. "Consequently, they interfere with one another, cause coarse fluctuations in the firing of peripheral and central auditory neurons, and give rise to the perception of roughness and dissonance" (Tramo *et al.*, 2001, p. 92).

II. BACKGROUND

A. Autocorrelation functions and hearing theories

Tramo *et al.* (2001) determined the ISI distributions embedded in the responses of axons throughout the auditory nerve during stimulation with musical intervals. Comparing these ISI histograms with the computed autocorrelation functions (the primaries consisting of six equally strong harmonics), they found the same periodicity patterns in both the autocorrelation and the ISI distributions.

From a logical point of view, measuring and counting the all-order ISI are an analysis in the time domain equivalent to the computation of an autocorrelation function. The autocorrelation function shows peaks for all periods of a signal. As periods are distances in time, the autocorrelation function has to be regarded as an analysis in the time domain. From the investigations of Cariani and Delgutte (1996), it seemed probable that a neuronal autocorrelation mechanism for the detection of the periods of running spike trains in the auditory system provides the sensation of pitch. It must be pointed out that the autocorrelation function is as powerful a means for sound analysis as the Fourier transform. The (famous) theorem of Wiener–Khnitchine (Wiener, 1930; Hartmann, 2000) states that the autocorrelation function is the Fourier transform of the power spectrum (energy spectral density) (see also Papoulis, 1962, p. 246). As a consequence, the autocorrelation analysis is equivalent to a Fourier analysis of a signal. The Fourier analysis is used for spectral analysis in the frequency domain; the autocorrelation analysis is a periodicity analysis in the time domain. The power spectrum shows all frequencies inherent in the signal but no phase; the autocorrelation function shows all periods inherent in the signal including all subharmonic periods but also no phase information. This is an important feature of the autocorrelation process: Runtime differences between different spike trains in the auditory system are nullified, thus facilitating the highest possible coincidence rate between two correlated spike trains.

Neuronal spike patterns in the auditory system can be represented mathematically by pulse sequences. All information about their periodicities and ISIs can be provided by forming their autocorrelation functions (Papoulis, 1962, p.

249). Thus, the "existence of a central processor capable of analyzing these interval patterns could provide a unified explanation for many different aspects of pitch perception" (Cariani and Delgutte, 1996, p. 1698). Since Licklider (1951), a lot of auditory theories operating in the time domain have presumed an autocorrelation mechanism or a related model to detect the periodicity of the stimuli (for an overview, see Hartmann, 2000; de Cheveigné, 2005). These models have been tested psychoacoustically or with computer simulations (e.g., Meddis and Hewitt, 1991; Patterson *et al.*, 1995), using different stimuli. Psychoacoustical tests are interpreted as evidences for (e.g., Yost *et al.*, 1996) and sometimes against (e.g., Kaernbach and Demany, 1998) an autocorrelation mechanism in the auditory system. Few models are based on physiological data using properties of neuronal circuits in the auditory pathway (Langner, 1983; Hewitt and Meddis, 1994; Voutsas *et al.*, 2005). The present paper refers to properties of Langner's model of periodicity detection in the *inferior colliculus* (IC) (Langner, 2007) The mathematical model in this paper emulates the underlying logical structure of Langner's neuronal periodicity detection model for pitch perception and applies it to the perception of musical intervals. So, Langner's model will be briefly presented in Sec. II B.

B. Langner's neuronal correlator model

Langner (1983) measured the responses of neurons in the *cochlear nucleus* (CN) and IC to amplitude modulated signals and proposed a model that performs a correlation between signal fine-structure and modulation envelope.

As Langner (1983) explained, its processing elements are a trigger, an oscillator, a reducer, and a coincidence neuron, which are supposed to have their counterparts in well-described on-type, chopper neurons, and pauser neurons in the CN and disk cells in the IC. The trigger unit synchronizes the responses of oscillator and reducer cycle to the modulation. While the oscillator responds with short bursts of regular intrinsic oscillations to each modulation period, the reducer generates intervals precisely related to the signal fine structure. By integrating synchronized activity of many nerve fibers, the reducer is able to code frequencies to the upper limit of phase coupling. The coincidence unit is activated by simultaneous inputs from oscillator and reducer and responds best when signal fine structure and envelope are correlated and the envelope period matches the reducer delay. Thereby it responds best to a periodically modulated sound (BMF) and is simultaneously representing a certain frequency and a certain pitch. (Langner, 2005, p. 51)

Three different periods are crucial for coincidence detection: (a) τ_m the period of the envelope; (b) τ_c the period of a carrier frequency: in other words, the fine-structure of the sound; (c) τ_o the period of intrinsic oscillation.

Langner (2007) assumed that the detection of the envelope period yields to the sensation of pitch, whereas the timbre of the sound corresponds to the fine structure of the sound, represented by τ_c.

The intrinsic oscillation provides a time slot of coincidence. The distribution of intervals shows the highest peaks for periods of τ_o=0.8 ms, τ_o=1.2 ms, τ_o=1.6 ms, τ_o=2.0 ms, and τ_o=2.4 ms or generally τ_o=0.8 ms $+k\cdot$0.4 ms (Langner and Schreiner, 1988, p. 1813). Mathematically, these three periods correlate if there are small integers n and m, so that the "periodicity equation" is valid (see Langner and Schreiner, 1988, p. 1818):

$$m \cdot \tau_m + n \cdot \tau_c + \tau_o = 0. \tag{1}$$

The intrinsic oscillation with period τ_o may be interpreted as a time window for the coincidence detection as intrinsic oscillation raises the coincidence neuron onto an excitation level closely under the threshold. Therefore, one may hypothesize that the shortest possible oscillation interval of 0.8 ms is also the size of the shortest possible coincidence window. At the beginning of stimulation, each coincidence neuron shows a response characteristic of a comb filter which makes the whole bank of coincidence circuits act as an autocorrelator for the modulation frequencies. However, after about 30 ms, inhibition triggered by the onset neuron converts the coincidence neuron to a bandpass filter (Voutsas *et al.*, 2005) so that an unambiguous pitch discrimination is granted.

III. MATHEMATICAL MODEL OF GENERALIZED COINCIDENCE

A. Correlation functions

Applying autocorrelation functions makes it necessary to classify functions (signals) according to their average power that is defined by

$$\overline{f^2}(t) = \lim_{T\to 2T} \frac{1}{2T} \int_{-T}^{T} |f(t)|^2 dt. \tag{2}$$

The proposed model of generalized coincidence makes use only of functions with finite energy, which means that $\overline{f^2}(t)=0$. Nevertheless, it can easily be extended to finite power functions, which have the property that $0<\overline{f^2}(t)<\infty$ (Papoulis, 1962, p. 240).

In the case of functions with finite energy, the correlation functions of two functions $f_1(t)$ and $f_2(t)$ are defined by the following.

(a) Autocorrelation function [see Papoulis, 1962, p. 241 (12–7)]:

$$\rho_i(\tau) = \int_{-\infty}^{\infty} f_i(t) f_i(t+\tau) dt. \tag{3}$$

(b) Cross-correlation functions [see Papoulis, 1962, p. 244 (12–20)]:

$$\rho_{12}(\tau) = \int_{-\infty}^{\infty} f_1(t) f_2(t+\tau) dt, \tag{4}$$

$$\rho_{21}(\tau) = \int_{-\infty}^{\infty} f_2(t) f_1(t+\tau) dt. \tag{5}$$

Substituting $t'=t+\tau \Leftrightarrow t=t'-\tau$ in Eqs. (3) and (4) shows that

$$\rho_i(\tau) = \rho_i(-\tau), \tag{6}$$

$$\rho_{12}(\tau) = \rho_{21}(-\tau). \tag{7}$$

From Eq. (7), it follows that the sum of both cross-correlation functions is even or symmetric:

$$(\rho_{12} + \rho_{21})(\tau) = \rho_{12}(\tau) + \rho_{21}(\tau) = \rho_{21}(-\tau) + \rho_{12}(-\tau)$$
$$= (\rho_{21} + \rho_{12})(-\tau). \tag{8}$$

Let $F_i(\omega)$ be the Fourier transforms of $f_i(t)$. With $\bar{F}_i(\omega)$ as the conjugate complex of $F_i(\omega)$, definitions (3)–(5) are equivalent to

$$\rho_i(\tau) \leftrightarrow F_i(\omega)\bar{F}_i(\omega) =: A^2(\omega), \tag{9}$$

$$\rho_{12}(\tau) \leftrightarrow \bar{F}_1(\omega)F_2(\omega) =: E_{12}(\omega), \tag{10}$$

$$\rho_{21}(\tau) \leftrightarrow \bar{F}_2(\omega)F_1(\omega) =: E_{21}(\omega), \tag{11}$$

with the arrows indicating Fourier transform [see Papoulis, 1962, p. 242 (12–9) and p. 244 (12–17)]. A finite energy function is real if and only if (see Papoulis, 1962, p. 11)

$$\bar{F}_i(\omega) = F_i(-\omega). \tag{12}$$

Obviously, $A^2(\omega)$ is real. As Eq. (12) holds for $E_{12}(\omega)$ and $E_{21}(\omega)$, both cross-correlation functions are also real functions.

Sometimes it is more convenient to calculate the correlation functions from these formulas.

Let $S(t)=f_1(t)+f_2(t)$ be the sum of two functions $f_1(t)$ and $f_2(t)$. Using definition (3) immediately leads to the sum formula of autocorrelation functions:

$$\rho_S(\tau) = \int_{-\infty}^{\infty} S(t)S(t+\tau)dt$$
$$= \int_{-\infty}^{\infty} [f_1(t) + f_2(t)][f_1(t+\tau) + f_2(t+\tau)]dt$$
$$= \int_{-\infty}^{\infty} f_1(t)f_1(t+\tau)dt + \int_{-\infty}^{\infty} f_2(t)f_2(t+\tau)dt$$
$$+ \int_{-\infty}^{\infty} f_1(t)f_2(t+\tau)dt + \int_{-\infty}^{\infty} f_2(t)f_1(t+\tau)dt$$
$$= \rho_1(\tau) + \rho_2(\tau) + \rho_{12}(\tau) + \rho_{21}(\tau), \tag{13}$$

B. Sequence representation of a tone

In the auditory system, pitch is represented by periodic pulse trains which can be mathematically written as a sequence of equally spaced pulses (M positive integer or ∞):

$$x_\mu(t) = \sum_{m=-M}^{M} I_\mu(t - mT). \tag{14}$$

The constant T is the period of the pulse train, and it is the reciprocal of the frequency corresponding to the perceived pitch. The function $I_\mu(t)$ describes the pulse form. A neuronal pulse is built from several neuronal discharges randomly distributed around time mT. Therefore, we demand $I_\mu(t)$ to be a density function, which means

(i) $I_\mu(t) \geq 0$ for every t, (15)

$$\int_{-\infty}^{\infty} I_\mu(t)dt = 1. \tag{16}$$

Furthermore, the spread of the single discharges is determined by the dimensionless parameter μ describing the "width" of the pulse $I_\mu(t)$. The pulse width is crucial for the model. Later, it is chosen to meet the probability of coincidence of two different pulse sequences, processed in an autocorrelator such as the one described by Langner (1983) (Sec. II B). Taking μ as a real number, $I_\mu(t)$ becomes a family of functions with the generalized limit $\delta(t)$ (Papoulis, 1962, p. 277). Thus, a third property of $I_\mu(t)$ follows:

(iii) $\lim_{\mu\to 0} I_\mu(t) = \delta(t)$. (17)

The limit

$$\lim_{\mu\to 0} x_\mu(t) = \sum_{m=-M}^{M} \delta(t - mT) \tag{18}$$

is the idealized case of all neuronal discharges occurring exactly at time mT. Examples for $I_\mu(t)$ may be the rectangular pulse with the width μ, the Gaussian pulse with μ to determine the variance or the half-wave rectified cosine pulse with 2μ to determine the period of the cosine (see Papoulis, 1962, p. 279 and Hartmann, 2000, p.156; compare Fig. 1).

If two pulses fulfill properties (i)–(iii), their cross-correlation functions also fulfill properties (i)–(iii). Considering definition (4), properties (i) and (ii) become obvious for the cross-correlation function. Property (iii) can be proved by applying the definition of the generalized limit: If

$$\chi_{\mu\nu}(\tau) = \int_{-\infty}^{\infty} I_\mu(t)J_\nu(t+\tau)dt \tag{19}$$

is the cross-correlation function of two pulses $I_\mu(t)$ and $J_\nu(t)$, it can be shown that

$$\lim_{\nu\to 0} \lim_{\mu\to 0} \int_{-\infty}^{\infty} \chi_{\mu\nu}(\tau)\phi(\tau)d\tau = \phi(0), \tag{20}$$

for every continuous test function $\phi(t)$. By definition of the generalized limit, this is equivalent to property (iii) (Papoulis, 1962, p. 277).

As by definition the autocorrelation function is a special case of a cross-correlation function, properties (i)–(iii) also hold for the autocorrelation functions of $I_\mu(t)$ and $J_\nu(t)$.

C. Sequence representation of a musical interval

In the mathematical model presented here, the sum of two simultaneously running pulse trains is the mathematical description of the spike train representation of a musical interval in the auditory system. Let $I_\mu(t)$ and $J_\nu(t)$ be the two families of pulse functions with properties (i)–(iii) as above. The two tones of the interval shall be represented by the two sequences, as follows:

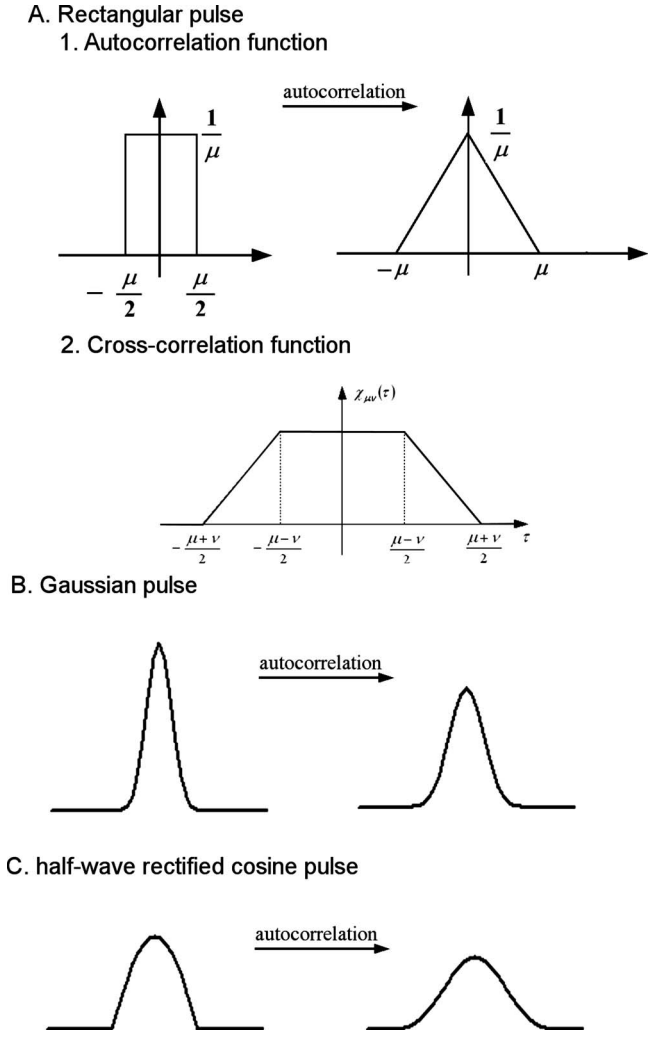


FIG. 1. The correlation functions of tree pulse forms. (A1) The autocorrelation function of a rectangular pulse is a triangular pulse. (A2) shows the cross-correlation function of two rectangular pulses with the different widths μ and ν. (B) The autocorrelation function of a Gaussian pulse is a Gaussian pulse with double the width. (C) shows the autocorrelation function of a half-wave rectified cosine pulse.

$$x_\mu(t) = \sum_{m=-M}^{M} I_\mu(t - mT_1), \tag{21}$$

$$x_\nu(t) = \sum_{n=-N}^{N} J_\nu(t - nT_2). \tag{22}$$

Their sum

$$S(t) = x_\mu(t) + x_\nu(t) \tag{23}$$

is the mathematical representative of the musical interval of this tones. This simple mathematical definition of a musical interval presumes that two neuronal pulse sequences add up without any disturbances and implies complete linearity of the auditory system. Actually, two simultaneously presented tones interfere in the auditory system, which generally does not behave like a linear time invariant (LTI) system. Carefully considering some restrictions, the processing of the auditory system may be approximated by LTI systems (for discussion, see Sec. IV).

Furthermore, let $\alpha_\mu(\tau)$ be the autocorrelation function of the pulse $I_\mu(t)$, $\alpha_\nu(\tau)$ be the autocorrelation function of the pulse $J_\nu(t)$, and $\chi_{\mu\nu}(\tau)$ and $\chi_{\nu\mu}(\tau)$ be the cross-correlation functions of the pulses $I_\mu(t)$ and $J_\nu(t)$. From Eq. (7), it follows

$$\chi_{\mu\nu}(-\tau) = \chi_{\nu\mu}(\tau). \tag{24}$$

If the pulses of both pulse sequences are of the same form and width, then it is valid that $I_\mu(t) = J_\nu(t)$. With Eqs. (3)–(5), it therefore follows that the cross-correlation functions are equal to the autocorrelation function:

$$I_\mu(t) = J_\nu(t) \Rightarrow \chi_{\mu\nu}(\tau) = \chi_{\nu\mu}(\tau) = \alpha_\mu(\tau) = \alpha_\nu(\tau). \tag{25}$$

From the definitions (3)–(5) and the sum formula (13) (the linearity of integration, respectively), it follows that the autocorrelation function of $x_\mu(t)$ is the sequence

$$\rho_\mu(\tau) = \sum_{m=-2M}^{2M} (2M + 1 - |m|)\alpha_\mu(\tau - mT_1). \tag{26}$$

The autocorrelation function of $x_\nu(t)$ is the sequence

$$\rho_\nu(\tau) = \sum_{n=-2N}^{2N} (2N + 1 - |n|)\alpha_\nu(\tau - nT_2). \tag{27}$$

The cross-correlation functions of $x_\mu(t)$ and $x_\nu(t)$ are

$$\rho_{\mu\nu}(\tau) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} \chi_{\mu\nu}(\tau - nT_2 + mT_1), \tag{28}$$

$$\rho_{\nu\mu}(\tau) = \sum_{n=-N}^{N} \sum_{m=-M}^{M} \chi_{\nu\mu}(\tau - mT_1 + nT_2). \tag{29}$$

As the parameters m and n are as well positive and negative and as $\chi_{\mu\nu}(-\tau) = \chi_{\nu\mu}(\tau)$ from Eq. (24), both cross-correlation functions are equal:

$$\rho_{\mu\nu}(\tau) = \rho_{\nu\mu}(\tau). \tag{30}$$

The pulse sequences $x_\mu(t)$ and $x_\nu(t)$ are real functions. Their correlation functions are also real functions [see Eqs. (9)–(12)]. The imaginary part of a signal describes its phase composition. Thus, as $\rho_\mu(\tau)$, $\rho_\nu(\tau)$, $\rho_{\mu\nu}(\tau)$, and $\rho_{\nu\mu}(\tau)$ are real, it follows that they have no phase shifts, or as one may say, they are all in phase. We must keep in mind that the correlation functions of pulse sequences are pulse sequences themselves. As the correlation functions are in phase, it is granted that at least one pulse of each sequence coincides with at least one pulse of all other sequences, namely, for $\tau = 0$. The occurrence of further coincidences depends exclusively on the vibration ratio s. Thus, to calculate the degree of coincidence, the variable s must be introduced.

D. Autocorrelation function of an interval

As above, let $x_\mu(t)$ and $x_\nu(t)$ be the mathematical descriptions of the two pulse sequences that are assumed to be the neuronal representations of a musical interval with the vibration ratio s. If ν_1 and ν_2 are the frequencies of the two tones constituting the interval, and T_1 and T_2 are the corresponding periods, the vibration ratio of the interval is

$$s = \frac{\nu_2}{\nu_1} = \frac{T_1}{T_2}. \tag{31}$$

As the autocorrelation function also depends on the vibration ratio, s is introduced as a second variable. Applying Eq. (13) to $S(t)=x_\mu(t)+x_\nu(t)$ with regard to Eq. (31), it follows from Eqs. (26)–(30) that

$$\begin{aligned}\rho_S(\tau,s) &= \rho_\mu(\tau) + \rho_\nu(\tau) + \rho_{\mu\nu}(\tau) + \rho_{\nu\mu}(\tau)\\ &= \rho_\mu(\tau) + \rho_\nu(\tau) + 2\rho_{\mu\nu}(\tau)\\ &= \sum_{m=-2M}^{2M} (2M+1-|m|)\alpha_\mu(\tau - mT_1)\\ &\quad + \sum_{n=-2N}^{2N} (2N+1-|n|)\alpha_\nu(\tau - ns^{-1}T_1)\\ &\quad + 2\sum_{m=-M}^{M}\sum_{n=-N}^{N} \chi_{\mu\nu}(\tau - (ns^{-1}-m)T_1)\end{aligned} \tag{32}$$

is the autocorrelation function of the musical interval with the vibration ratio s. If the pulses of both pulse sequences are of the same form and width, then together with Eq. (25) the autocorrelation function of Eq. (32) becomes

$$\begin{aligned}\rho_S(\tau,s) &= \sum_{m=-2M}^{2M} (2M+1-|m|)\alpha_\mu(\tau - mT_1)\\ &\quad + \sum_{n=-2N}^{2N} (2N+1-|n|)\alpha_\mu(\tau - ns^{-1}T_1)\\ &\quad + 2\sum_{m=-M}^{M}\sum_{n=-N}^{N} \alpha_\mu(\tau - (ns^{-1}-m)T_1).\end{aligned} \tag{33}$$

E. Definition of the generalized coincidence function

As only positive periods up to a certain length $D>0$ are of interest, the generalized coincidence function (GCF) is defined as the integral

$$K(s) := \int_0^D \rho_S^2(\tau,s)d\tau. \tag{34}$$

For each vibration ratio s, $K(s)$ is a measure value of overall coincidence between the two tones of the musical interval with regard to pulse forms and pulse widths. It should be mentioned that only a finite number of pulses are considered, because $D<\infty$.

IV. APPLICATIONS OF THE MODEL TO DIFFERENT PULSE FORMS

A. The general coincidence functions calculated from rectangular pulses

1. Autocorrelation function of the rectangular pulse

To apply the model, the degree of coincidence shall be calculated for all vibration ratios within an octave, $1<s<2$. A pulse function to fulfill properties (i)–(iii) is the rectangular pulse. The parameter μ determines the width of the pulse.

$$I_\mu(t) := \begin{cases} \frac{1}{\mu} & \text{if } |t| < \frac{\mu}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{35}$$

Its autocorrelation function is the triangular pulse (see Fig. 1; see also Papoulis, 1962, p. 243).

$$\alpha_\mu(\tau) = \Delta_\mu(\tau) := \begin{cases} \frac{1}{\mu}\left(1 - \frac{|\tau|}{\mu}\right) & \text{if } |\tau| < \mu \\ 0 & \text{otherwise} \end{cases}. \tag{36}$$

2. Cross-correlation function of the rectangular pulse

The cross-correlation functions of $I_\mu(t)$ and $I_\nu(t)$ can be found from the Fourier transforms of the rectangular pulses $I_\mu(t)$ and $I_\nu(t)$ that are given by (see Papoulis, 1962, p. 21)

$$I_\mu(t) \leftrightarrow \frac{1}{\mu}\frac{2\sin\left(\omega\frac{\mu}{2}\right)}{\omega} =: F_\mu(\omega), \tag{37a}$$

$$I_\nu(t) \leftrightarrow \frac{1}{\nu}\frac{2\sin\left(\omega\frac{\nu}{2}\right)}{\omega} =: F_\nu(\omega). \tag{37b}$$

As $\sin(ax)\sin(bx)=\frac{1}{2}(\cos((a-b)x)-\cos((a+b)x))$, it follows that

$$F_\mu(-\omega)F_\nu(\omega) = \frac{2}{\mu\nu\omega^2}\left(\cos\left(\frac{\mu-\nu}{2}\omega\right) - \cos\left(\frac{\mu+\nu}{2}\omega\right)\right). \tag{38}$$

Note that this product is even. Its Fourier transform is the cross-correlation function [see Eq. (10)]:

$$\begin{aligned}\chi_{\mu\nu}(\tau) &= \frac{1}{2\pi\mu\nu}\int_{-\infty}^{\infty}\frac{1}{\omega^2}\cos\left(\left(\frac{\mu-\nu}{2}-\tau\right)\omega\right)d\omega\\ &\quad + \frac{1}{2\pi\mu\nu}\int_{-\infty}^{\infty}\frac{1}{\omega^2}\cos\left(\left(\frac{\mu-\nu}{2}+\tau\right)\omega\right)d\omega\\ &\quad - \frac{1}{2\pi\mu\nu}\int_{-\infty}^{\infty}\frac{1}{\omega^2}\cos\left(\left(\frac{\mu-\nu}{2}-\tau\right)\omega\right)d\omega\\ &\quad - \frac{1}{2\pi\mu\nu}\int_{-\infty}^{\infty}\frac{1}{\omega^2}\cos\left(\left(\frac{\mu-\nu}{2}+\tau\right)\omega\right)d\omega.\end{aligned} \tag{39}$$

The integrals can be solved using the equation

$$\int_{-\infty}^{\infty}\frac{\cos(a\omega)}{\omega^2}d\omega = -\pi|a|, \tag{40}$$

which can be proved by the calculus of residue. As a result, the cross-correlation function of the two pulses $I_\mu(t)$ and $I_\nu(t)$ is the function [see Fig. 1 A2]

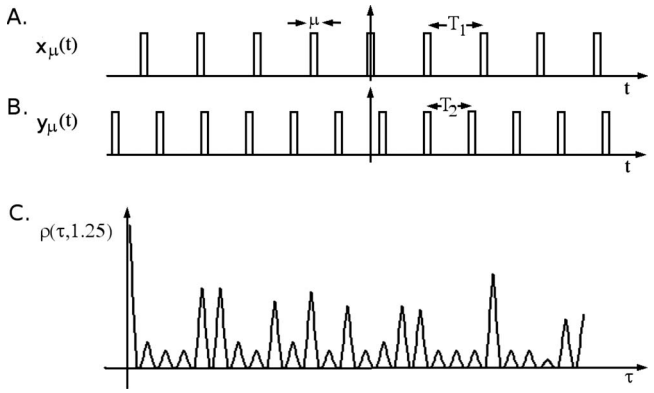


FIG. 2. (A) $x_\mu(t)$ is a sequence of rectangular pulses with a pulse width of μ. It has the period T_1. (B) $y_\mu(t)$ is a sequence of rectangular pulses also with a pulse width of μ. It has the period T_2 and is delayed to $x_\mu(t)$. (C) The ratio $s=(T_1/T_2)=(5/4)=1.25$ is the vibration ration of the major third. The graph shows the autocorrelation function $\rho(\tau,1.25)$ of the sum $S(t)=x_\mu(t)+y_\mu(t)$ of both pulse sequences neuronally representing a major third.

$$\chi_{\mu\nu}(\tau) := \begin{cases} \dfrac{1}{\mu\nu}\left(\dfrac{\mu+\nu}{2}-|\tau|\right) & \text{if } \dfrac{\mu-\nu}{2}<|\tau|\le\dfrac{\mu+\nu}{2} \\ \dfrac{1}{\mu} & \text{if } |\tau|\le\dfrac{\mu-\nu}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{41}$$

3. Autocorrelation function of a musical interval represented by rectangular sequences

Equations (21) and (22) can be applied to rectangular pulses as defined in Eq. (35). The parameter μ now describes a time window. Thus, the neuronal representation of the two tones are now mathematically described as sequences of rectangular pulses [see Figs. 2(a) and 2(b)] The autocorrelation function of the musical interval, represented by $S(t)=x_\mu(t)+x_\nu(t)$, is a sequence of partly coinciding triangular pulses [see Fig. 2(c)]:

$$\begin{aligned}\rho(\tau,s) = &\sum_{m=-2M}^{2M}(2M+1-|m|)\Delta_\mu(\tau-mT_1) \\ &+\sum_{n=-2N}^{2N}(2N+1-|n|)\Delta_\nu(\tau-ns^{-1}T_1) \\ &+2\sum_{m=-M}^{M}\sum_{n=-N}^{N}\chi_{\mu\nu}(\tau-(ns^{-1}-m)T_1).\end{aligned} \tag{42}$$

4. Calculation of the generalized coincidence function for musical intervals represented by rectangular sequences

To compute the GCF, the constants are set as follows.

- Audible frequencies range from about 20 to about 20 000 Hz, which corresponds to periods with lengths from 0 to 50 ms; thus we set $D:=50$.

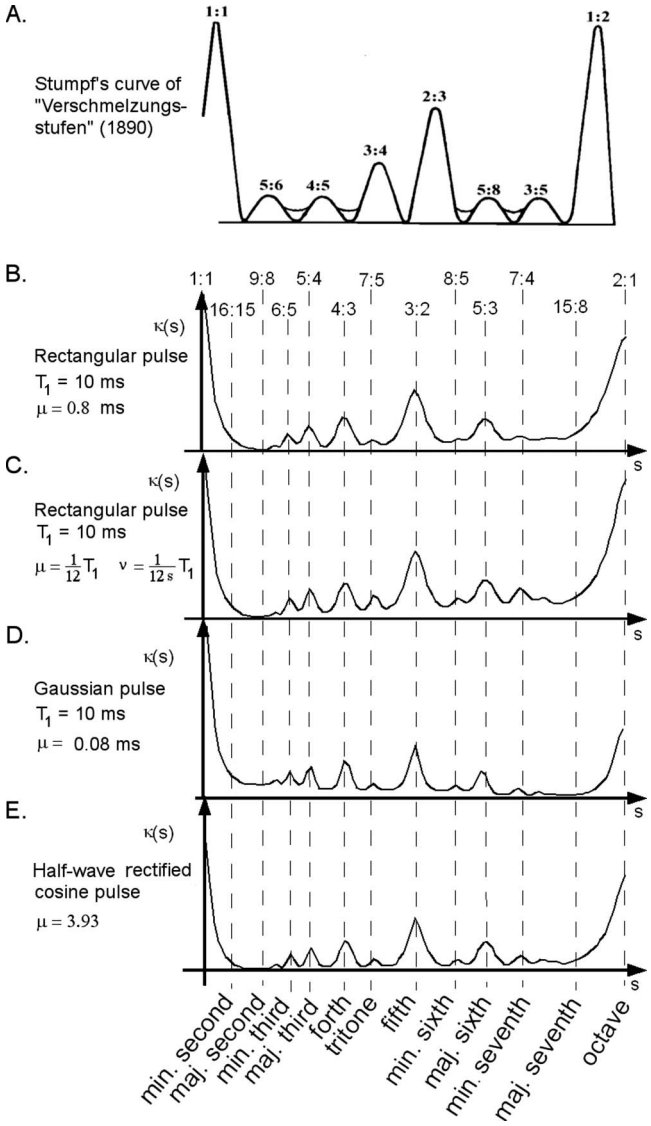


FIG. 3. (A) Stumpf's curve showing the system of the levels of tonal fusion ("System der Verschmelzungsstufen in einer Curve," 1890, 1965). It shows the degree of tonal fusion for all intervals within the range of an octave from the prime (1:1) to the octave (1:2) and the reciprocals of the vibration ratios of the musically important intervals. [(B)–(E)] The computed GCFs $\kappa(s)$ for different pulse forms: at the bottom, the 13 intervals within an octave are indicated and on the top the respective vibration ratios s are plotted.

- As at the most $M=D/T_1$ and $N=D/s^{-1}T_1$ pulses of $x_1(t)$ and $x_2(t)$, respectively, fit into the time window of D, we set $M=\text{floor}(D/T_1)$ and $N=\text{floor}(sD/T_1)$.
- In the first instance, let $\mu=\nu=0.8$ ms and $T_1=10$ ms so that the lowest tone has a frequency of 100 Hz. The relative pulse width can be described by the ratio

$$r := \frac{\mu}{T_1}. \tag{43}$$

In this case, the relative pulse width is $r=0.08$. The computer calculation leads to the GCF of Fig. 3(b) calculated from Eq. (33) with $\alpha_\mu(\tau)=\Delta_\mu(\tau)$ [see Eq. (25)].

- Again, let be $T_1=10$ ms and set $\mu=\frac{1}{12}T_1=0.898$ ms and $\nu=\frac{1}{12}T_2$. This case of a frequency dependent upper pulse width leads to the GCF of Fig. 3(c).

The graphs show maxima for musically important intervals within an octave. As can clearly be seen, the simpler the vibration ratio of a musical interval, the higher the corresponding peak in the GCF.

B. Two examples of the general coincidence functions calculated from other pulse forms

1. Generalized coincidence function calculated on the basis of the Gaussian pulse

The Gaussian pulse

$$I_\mu(t) = \frac{1}{\sqrt{\mu\pi}} e^{-1/\mu t^2} \tag{44}$$

is a distribution fulfilling properties (i)–(iii). The parameter μ is dimensionless and determines the width of the pulse. The autocorrelation function of the Gaussian pulse can be calculated from Eq. (9). Its Fourier transform is [see Papoulis, 1962, p. 25 (2–68)]

$$F(\omega) = e^{-\mu/4\omega^2}. \tag{45}$$

Thus, according to Eqs. (9) and (12) the Fourier transform of the product $F(\omega)F(-\omega)$ is the autocorrelation function of the Gaussian pulse [see Papoulis, 1962, p. 25 (2–68)]:

$$\alpha_\mu(\tau) = \frac{1}{\sqrt{2\mu\nu}} e^{-1/2\mu\tau^2} \leftrightarrow F(\omega)F(-\omega) = e^{-\mu/2\omega^2}. \tag{46}$$

It can be shown that $\alpha_\mu(\tau)$ is a distribution fulfilling properties (i)–(iii). Considering only the case that the pulses are of the same variance (width) for both pulse sequences, the autocorrelation function $\rho_S(\tau,s)$ of the sum of both sequences can be obtained from Eq. (33) with $\alpha_\mu(\tau)$ as in Eq. (46). Setting $\mu=0.08$ gives a Gaussian pulse that fits well into a rectangular pulse with a width of 0.8 ms, the Gaussian pulse with $\mu=0.102$ has the same center of gravity as the rectangular pulse with a width of 0.8. Again, set $T=10$ ms.The calculated GCF for $\rho_S(\tau,s)$ in the case of Gaussian pulses with $\mu=0.08$ is shown in Fig. 3(d).

2. Generalized coincidence function calculated on the basis of a half-wave rectified cosine pulse

The half-wave rectified cosine pulse, given by

$$I_\mu(t) := \begin{cases} \frac{1}{2\mu}\cos\left(\frac{1}{\mu}t\right) & \text{if } |t| < \frac{\pi}{2}\mu \\ 0 \quad \text{else} \end{cases}, \tag{47}$$

is a distribution fulfilling properties (i)–(iii). In this case, the parameter μ^{-1} can be interpreted as a frequency and determines the width of the cosine pulse. From definition (3), its autocorrelation function calculates to

$$\alpha_\mu(\tau) = \begin{cases} \frac{1}{8\mu^2}\left[(\pi\mu - |\tau|)\cos\left(\frac{|\tau|}{\mu}\right) + \mu\sin\left(\frac{|\tau|}{\mu}\right)\right] & \text{if } |\tau| < \pi\mu \\ 0 \quad \text{if } |\tau| \geq \pi\mu \end{cases}. \tag{48}$$

Again, it can be shown that $\alpha_\mu(\tau)$ is a distribution fulfilling properties (i)–(iii). Considering only the case that the pulses are of the same frequency (and thus of the same width) for both pulse sequences, the autocorrelation function $\rho_S(\tau,s)$ of the sum of both sequences can be obtained from Eq. (33) with $\alpha_\mu(\tau)$ as in Eq. (48). Setting $\mu=0.3142$ results in a half-wave rectified pulse with the same center of gravity as the rectangular pulse with a width of 0.8; choosing $\mu=0.364$ gives a half-wave rectified cosine pulse with an amplitude half down for $t=0.4$, and the half-wave rectified cosine pulse with $\mu=0.8\pi^{-1}$ fits well into a rectangular pulse with a width of 0.8 ms. Again set $T=10$ ms. The calculated GCF for $\rho_S(\tau,s)$ in the case of a half-wave rectified cosine pulse with $\mu=0.8\pi^{-1}$ is shown in Fig. 3(e).

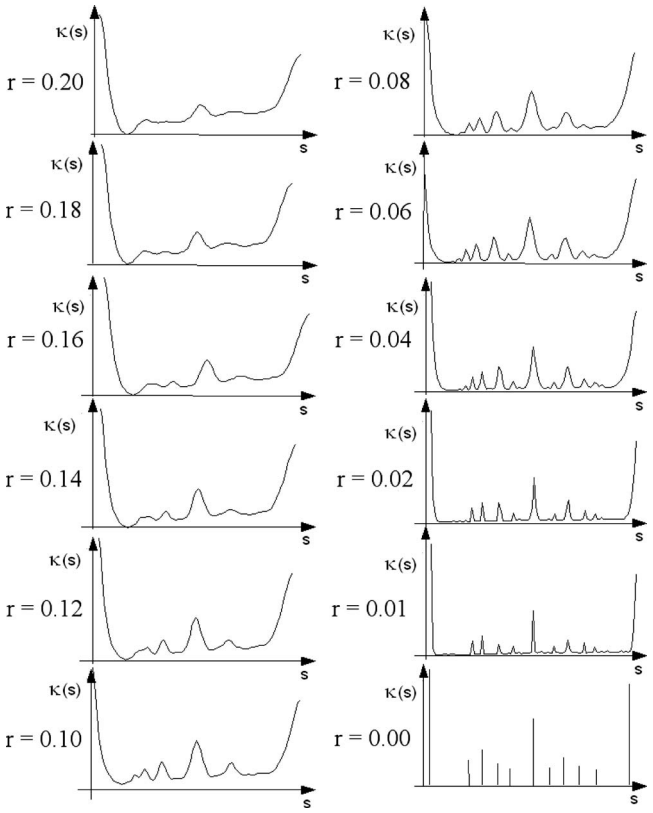


FIG. 4. (Color online) 12 GCFs are computed from sequences of rectangular pulse with the relative pulse width r converging from $r=0.2$ to zero, thus illustrating the convergence of the GCF.

C. The convergence of the generalized coincidence function

All pulses of the pulse sequences above are distributions with property (iii) so that every pulse has the generalized limit $\delta(t)$. Considering this property of the pulses, it becomes clear that all GCFs computed from different pulse forms approach the same limit function as parameter μ approaches zero. In Fig. 4, the convergence is illustrated for GCFs calculated from rectangular pulses of different widths with relative pulse widths ranging from $r=0.2$ to $r=0$. Taking into account the convergence against the same limit function, it becomes clear that the GCFs calculated from different pulse forms must become quite similar if the parameter μ is sufficiently small.

V. DISCUSSION

A. Properties of the generalized coincidence functions

The GCF shows a ranking of the musical intervals that is obviously equal to the traditional degrees of consonance as described by music theorists. Moreover, every consonant interval is surrounded by a region of still high coincidence. This is in accordance with the observation that slightly mistuned consonances preserve the characteristic of the interval and still sound consonant.

Though the model is formed in accordance with auditory pulse trains, the pulses of the model as applied in this paper do not reflect neuronal spiking but represent the probability of coincidence in an auditory autocorrelation process. Thus, the pulse shape is determined by properties of an auditory autocorrelator. The pulse width depends on the width of the coincidence window provided by the neuronal autocorrelation mechanism.

Provided that the pulses have a sufficiently small width, the shapes of the GCFs are only slightly affected by the pulse form. In spite of a presumable randomness of pulse forms in the auditory autocorrelation mechanism, the GCF obviously grants a great stability in coincidence detection.

Only symmetric pulses have been discussed so far. From Eq. (6), the symmetry of the autocorrelation function of any pulse is granted and from Eq. (8) it becomes clear that two related cross-correlation functions sum up to a symmetric pulse. So, on the level of the autocorrelation function, only symmetric pulses occur. Thus, it is totally sufficient to consider symmetric pulse forms exclusively.

As demonstrated in Sec. IV C, the shape of the GCF is sensitive to the relative pulse width, which is the ratio of pulse width to period. On the other hand, the number of considered spikes has a slight influence only on the shape of the GCF: The coincidences in the autocorrelation function of the interval are heightened. The squaring in the GCF leads to a more pronounced curve, but no general changes occur.

The assumption—that the sum of two pulse sequences [see Eq. (23)] is an adequate representation of firing patterns, evoked by musical intervals—is a linear assumption. Furthermore, the GCF is based on correlation functions in the time domain. The underlying mathematical techniques imply that the auditory system acts as a LTI system. Quite often LTIs are applied to describe neuronal processes. "In general, the success of most of these time series analysis methods in physiology is surprising considering that physiological processes are known to include significant nonlinearities. The explanation for this relative success is perhaps due to the cases in which physiological systems can be studied in a state where the linear behavior is most prominent …" (van Drongelen, 2007, p. 279). Each stage on the auditory pathway is a potential site of nonlinearity (see Hartmann, 2000, p. 511). However, first and foremost, nonlinearity affects signal strength and not the time patterns. If the effects of nonlinearity are comparatively small, especially in the dynamic range of music, the consideration of a LTI system is justified.

Wever *et al.* (1941), who discovered the cochlear microphonic, found that "the peripheral mechanism of the ear transmits vibrations with high fidelity" and considered "the most probable source of distortion to be the processes of the inner ear through which mechanical vibrations are transformed into electrical effects" (Wever *et al.*, 1941). The nonlinearities of the inner ear become evident from combination tones and suppression (de Boer, 1984). Plomp (1965) investigated the audibility of combination tones and determined that "all mean detectability thresholds found exceeded 40 dB, corresponding to a nonlinear distortion of below 1%." In experiments with musical interval stimuli, Tramo *et al.* (2001) found a strong correspondence between the pooled autocorrelogramms of neural spike rates and the computed autocorrelation functions of the stimuli, so that no significant suppression effect could be observed. Moreover, neuronal nonlinear effects, such as otoacoustic emissions (Zwicker and Fastl, 1999) or inhibition (Buzsáki, 2006), are believed to even enhance neural signal quality. Last, but not least, the evident correspondence between the GCF and Stumpf's "Stufen der Tonverschmelzung" may be regarded as a cue for a linearity in the perception of harmony.

B. The generalized coincidence functions and Stumpf's tonal fusion

Figure 3(a) shows the curve which Stumpf (1897) proposed to illustrate the levels of tonal fusion (Stufen der Tonverschmelzung) as he had investigated. The graph shows the degree of fusion for all intervals over a range of an octave. Not only do the consonant intervals, which are the prime (1:1), the minor third (5:6), the major third (4:5), the pure fourth (3:4), the pure fifth (2:3), the minor sixth (5:8), the major sixth (3:5), and the pure octave (1:2)—indicated by their vibration ratio—have a higher degree of fusion, but the slightly mistuned intervals nearby the consonant intervals also do.

The similarities of Stumpf's curve and the graph of the GCF are obvious, especially for relative pulse width in the range of $r=0.10-0.04$. Stumpf's curve is a schematic sketch showing an equal degree of coincidence for both sixths and thirds. He felt unsure about their order but had some evidence that the major third fuses more intensively than the minor third. The order of the minor and major sixth was not clear to him (Stumpf, 1897). In this respect, all GCFs in the range of $r=0.10-0.04$ clearly show a higher tonal fusion of the major sixth.

On the whole, one can conclude that tonal fusion and the degree of coinciding periods in the firing patterns for musical intervals are equivalent. Stumpf's concept of tonal fusion was an attempt to define consonance and dissonance psychologically. Considering the theory of periodicity detection in the midbrain as described by Langner (1983; Langner and Schreiner, 1988) and calculating the coincidence of musical intervals with a coincidence window of 0.8 ms as suggested by this theory result in a GCF closely resembling the curve Stumpf deduced from his extensive hearing experiments. Thus, Langner's model of periodicity analysis in the midbrain may provide a physiological correlate of tonal fusion.

Nevertheless, roughness is also an important sensation that is used to distinguish between consonance and dissonance. Pure tones produce roughness if they are close together or if the upper tone is close to the octave (Zwicker and Fastl, 1999). The intervals of the minor second, the minor seventh, and the minor ninth sound very harsh but show a high degree of coincidence, as can be seen in Figs. 3 and 4. They are dissonant only because of their roughness and in spite of their high degree of tonal fusion. On the other hand, we can also distinguish the degrees of consonance in the absence of roughness: for example, if an interval is formed by pure tones. In the absence of roughness, the sensation of consonance must be a consequence of tonal fusion. This leads to the conclusion that both roughness and tonal fusion determine the sensation of consonance and dissonance.

Burns, E. M., and Ward, W. D. (**1982**). "Intervals, scales, and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic, London), pp. 241–269.

Buzsáki, G. (**2006**). *Rhythms of the Brain* (Oxford, New York).

Cariani, P. A., and Delgutte, B. (**1996**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch Salience, II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch," J. Neurophysiol. **76**, 1698–1734.

Cedolin, L., and Delgutte, B. (**2005**). "Representation of the pitch of complex tones in the auditory nerve," in *Auditory Signal Processing*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, Danvers).

de Boer, E. (**1984**). "Auditory physics. physical principles in hearing theory. II," Phys. Rep. **105**, 141–226.

de Cheveigné, A. (**2005**). "Pitch perception models" in *Pitch: Neural Coding and Perception*, edited by C. Plack, R. R. Fay, A. J. Oxenham, and A. N. Popper (Springer, New York), pp. 169–233.

Goldstein, J. L., and Scrulovic, P. (**1977**). "Auditory-nerve spike intervals as an adequate basis for aural spectrum analysis," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 337–345.

Hartmann, W. M. (**2000**). *Signal, Sound, and Sensation*, 4th ed. (Springer, New York).

Hewitt, M. J., and Meddis, R. (**1994**). "A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus," J. Acoust. Soc. Am. **95**, 2145–2159.

Kaernbach Chr., and Demany L. (**1998**). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," J. Acoust. Soc. Am. **104**, 2298–2306.

Langner, G. (**1983**). "Evidence for neuronal periodicity detection in the auditory system of the guinea fowl implications for pitch analysis in the time domain," Exp. Brain Res. **52**, 333–355.

Langner, G. (**2005**). "Neuronal mechanisms underlying the perception of pitch and harmony," Ann. N. Y. Acad. Sci. **1060**, 50–52.

Langner, G. (**2007**). "Temporal processing of periodic signals in the auditory system: neuronal representation of pitch, timbre, and harmonicity," Z. Audiologie **46**, 8–21.

Langner, G., and Schreiner, C. E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms, II. Topographical organization," J. Neurophysiol.. **60**, 1799–1822 and 1823–1840.

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Cell. Mol. Life Sci. **7/4**, 128–134.

Meddis, R., and Hewitt, M. J. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. II: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2866–2894.

Moore, B. C. J., and Glasberg, B. R. (**1986**). "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).

Papoulis, A. (**1962**). *The Fourier Integral and Its Applications* (McGraw-Hill, New York).

Patterson, R. D., Allerhand, M. H., and Giguère Chr. (**1995**). "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.

Plomp, R. (**1965**). "Detectability threshold for combination tones," J. Acoust. Soc. Am. **37**, 1110–1123.

Plomp, R., and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**, 548–560.

Rhode, W. S. (**1995**). "Interspike intervals as correlate of periodicity pitch in cat cochlear nucleus," J. Acoust. Soc. Am. **97**, 2414–2429.

Stumpf, C. (**1890**). *Tonpsychologie ("Psychology of Tones")* (Knuf, Hilversum), reprinted 1965.

Stumpf, C. (**1897**). "Neueres über Tonverschmelzung," Zeitschift für Psychologie und Physiologie der Sinnesorgane **15**, 280–303.

Stumpf, C. (**1930**). "Autobiography of Carl Stumpf," in *History of Psychology in Autobiography*, edited by C. Murchison (Clark University Press, Worchester, MA), Vol. **1**, pp. 389–441.

Terhardt, E. (**1976,1977**). "Ein psychoakustisch begründetes Konzept der Musikalischen Konsonanz," ("A psychoacoustically substantiated concept of musical consonance,")," Acustica **36**, 121–137.

Terhardt, E. (**1984**). "The concept of musical consonance: A link between music and psychoacoustics," Music Percept. **1**, 276–226.

Tramo, M. J. P., Cariani, A., Delgutte, B., and Braida, L. D. (**2001**). "Neurobiological foundations for the theory of harmony in Western tonal music," in *The Biological Foundations of Music*, edited by J. Zatorre et I. Peretz (Annals of the New York Academy of Sciences, New York), Vol. **930**, pp. 92–116.

van Drongelen, W. (**2007**). *Signal Processing for Neuroscientists* (Elsevier, Amsterdam).

von Helmholtz H. (**1877**). in *On the Sensations of Tone*, edited by J. A. Ellis (Dover, New York), reprinted 1885 and 1954.

Voutsas, K., Langner, G., Adamy, J., and Ochse, M. (**2005**). "A brain-like neural network for periodicity analysis," IEEE Trans. Syst., Man, Cybern., Part B: Cybern. **35**, 12–22.

Wever, E. G. (**1949**). *Theory of Hearing* (Wiley, New York).

Wever, E. G., Bray, C. W., and Lawrence, M. (**1941**). "Combination tones: their nature and origin in the auditory mechanism," J. Acoust. Soc. Am. **12**, 468.

Wiener, N. (**1930**). "Generalized harmonic analysis," Acta Math. **55**, 117–258.

Yost, W. A., Patterson, R., and Sheft, St. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1077.

Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory–nerve fibers," J. Acoust. Soc. Am. **109**, 648–670.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics*, 2nd ed. (Springer, Berlin).

Cell viability viscoelastic measurement in a rheometer used to stress and engineer tissues at low sonic frequencies[a)]

Sarah A. Klemuk[b)] and Sanyukta Jaiswal
National Center for Voice and Speech at the Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, Iowa 52242

Ingo R. Titze
National Center for Voice and Speech at the Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, Iowa 52242 and The Denver Center for the Performing Arts, Denver, Colorado 80201

(Received 22 January 2008; revised 8 July 2008; accepted 10 July 2008)

Effects of vibration on human vocal fold extracellular matrix composition and the resultant tissue viscoelastic properties are difficult to study *in vivo*. Therefore, an *in vitro* bioreactor, simulating the *in vivo* physiological environment, was explored. A stress-controlled commercial rheometer was used to administer shear vibrations to living tissues at stresses and frequencies corresponding to male phonation, while simultaneously measuring tissue viscoelastic properties. Tissue environment was evaluated and adjustments made in order to sustain cell life for short term experimentation up to 6 h. Cell nutrient medium evaporation, osmolality, pH, and cell viability of cells cultured in three-dimensional synthetic scaffolds were quantified under comparably challenging environments to the rheometer bioreactor for 4 or 6 h. The functionality of the rheometer bioreactor was demonstrated by applying three vibration regimes to cell-seeded three-dimensional substrates for 2 h. Resulting strain was quantified throughout the test period. Rheologic data and cell viability are reported for each condition, and future improvements are discussed. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2973183]

PACS number(s): 43.80.Cs, 43.80.Ev [FD] Pages: 2330–2339

I. INTRODUCTION

Empirical observation of cellular response to vocal fold vibration is difficult to study in humans. An alternative is to grow a tissue surrogate from vocal fold derived cells *in vitro* and expose it to vibration force fields that are comparable to human phonation in a bioreactor. A bioreactor is a system that simulates physiological environments for the creation, physical conditioning and testing of cells, tissues, precursors, support structures, and organs *in vitro* (Barron *et al.*, 2003). Freed and Vunjak-Novakovic (2000) outlined that to be functional, a bioreactor should be capable of performing at least one of four functions: (1) establish a uniform distribution of cells on a three-dimensional (3-D) scaffold, (2) maintain gas and nutrient concentrations amenable to cell growth, (3) provide fluid circulation in order to move nutrients into and remove waste out of the growing tissue, and (4) expose developing tissue to physical stimuli.

Unique challenges of vocal fold tissue bioreactor design not only include the framework laid down by Freed and Vunjak-Novakovic but also include quantifying physical stimuli and measuring real-time viscoelastic tissue properties. These additional functions are necessary for accurate simulation of typically occurring phonation forces, and for examining biomechanical contributions of cellular and extracellular matrix (ECM) response to specific forces. It is important to find connections between macroscopic and microscopic vocal fold biomechanics. Viscoelastic properties of vocal fold tissues are macroscopic frequency-dependent quantities that directly affect mucosal wave form, phonation threshold pressure, and fundamental frequency (Chan and Titze, 2006; Perlman and Durham, 1991; Titze and Durham, 1991; Titze, 1988; Titze and Hunter, 2004). It is therefore essential to accurately determine the rheologic character of engineered vocal fold tissues while they are being tested or conditioned in a bioreactor. Characterization of microscopic cellular properties such as protein and gene expression, cell orientation, and molecular architecture, after the same exposures, can then help to better understand their contributions to biomechanical vocal fold vibration. Aims of the present investigation were to improve cell viability conditions of a cell-seeded three-dimensional substrate in a rheometer bioreactor, to deliver three different vibrational loading conditions where both input stress and output strain of the tissue were quantified, and to determine rheologic and cell viability response of three-dimensional cellularized tissues after vibration.

II. BIOREACTOR REQUIREMENTS

To date, some but not all of the stipulated bioreactor requirements have been met for the rheometer bioreactor used in the present study.

[a)]This paper includes material presented at the 6th ICVPB, Tampere, Finland, 6–9 August 2008.

[b)]Author to whom correspondence should be addressed. Electronic mail: sarah-klemuk@uiowa.edu

A. Uniform cell distribution

The first function of a bioreactor, to generate uniform cell distribution in a three-dimensional scaffold, is at this time a reflection of the scaffold. There is no expectation of substantial tissue development while in a bioreactor for experiments lasting several hours. Instead, cells and ECM should inhabit a synthetic substrate for several weeks while in an incubator. During this time mechanical properties of the scaffold should remain stable, and similar to vocal fold tissues, and the scaffold should support cell growth and protein production.

Tecoflex® (TFX) (Lubrizol Advanced Materials, Inc) is a nonbiodegradable, medical grade polymer. It was used in the present study, because TFX had similar properties to vocal fold tissues and was well characterized as biocompatible. TFX has been used for synthetic vascular grafts, *in vitro* tissue engineering studies (Klemuk *et al.*, 2001; Mulder *et al.*, 1998; Titze *et al.*, 2005; Titze *et al.*, 2004a; Vara *et al.*, 2005; Webb *et al.*, 2003), and has been fabricated in many shapes including electrospinning for controlled microfiber formation and increased cellular matrix formation (Andrews *et al.*, 2008; Andrews and Hunt, 2008; Mulder *et al.*, 1998). When TFX was fabricated as a 60%–70% porous scaffold with pore diameters of 66 μm or larger, its Young's modulus was similar to linearly functioning vocal fold ligament (20–30 kPa) (Min *et al.*, 1995) and remained constant up to 100% strain (Webb *et al.*, 2003). When this same material was seeded with tracheal fibroblast cells, as was done in the present study, cells distributed homogeneously and stained abundantly positive for collagen I, vimentin, and fibronectin. Fluorescein diacetate and ethidium homodimer staining indicated cell viability and no evidence of necroses up to 4 weeks in culture (Webb *et al.*, 2003).

B. Gas, nutrient, and fluid exchange

Direct environmental measurements in the rheometer bioreactor are needed. Evaporation of the cell nutrient medium was a unique challenge when using a stress-controlled rheometer to vibrate living tissues (Titze *et al.*, 2004b). Up to 22% of cell nutrient medium evaporated in 3 h and about half evaporated at 8 h with most fibroblasts being nonviable.

Use of a sealed incubator into which a bioreactor is placed is a common method for controlling gases and humidity, but this is not feasible for the rheometer bioreactor. The rheometer itself is large, measuring approximately 50 cm tall, 30 cm wide, and 30 cm deep, has purified air lines, water lines, and digital cables connected to peripheral devices, and cannot be moved without performing extensive calibrations. The user must also have access to stationary and moving parts for mounting, dismounting, raising, and lowering the plate. Because of these circumstances, our goal for the present investigation was to evaluate the extent to which the mounted cup, along with a newly designed acrylic cover, could sustain an adequate cell environment.

Osmotic pressure and pH change are indicators of medium evaporation and cell environment adequacy. Osmolality is the ratio of moles of solute to kilograms of solvent. De Zengotita, Schmelzer, and Miller (2002) found that increased osmolality, evaluated at 320–476 mOsm/kg, decreased hybridoma cell proliferation and increased apoptotic cell death. Potter and DeMarse (2001) reported that changes greater than 50 mOsm/day resulted in neural cell culture death. Osmolality of fresh cell nutrient medium from the Titze *et al.* (2004b) study, was 300.3 mOsm/kg (not published). After 3 h, the osmolality would have been 464.6 mOsm/kg, a value near the limits of the De Zengotita *et al.* study and exceeding Potter and DeMarse's maximum osmolality change by 26 times.

Potential of hydrogen (pH) of the nutrient medium used in the Titze *et al.* (2004b) study, a measure of adequacy of gas supplementation, rose from 7.2 to 8.69 in 3 h. This indicates a lack of CO_2 gas exchange within the medium despite CO_2 exposure at the upper surface of the medium. Hepes (N-2-hydroxyethylpiperazine-N′-2-ethenesulphonic acid), a widely used organic buffer, is resistant to rapid pH change. Buffer strength for cell culture applications is usually in the range of 10–20 mM, but concentrations can vary depending on the cell nutrient medium used and can be toxic if concentrations are too high (Freshney, 2000). In the Titze *et al.* (2004b) study, cell viability improved by including 1.5 mM of Hepes buffer in the medium, but pH was 7.71, still high after 2.5 h ambient air exposure, and quantitative osmolality and cell viability results were not reported (Titze *et al.*, 2004b).

C. Vibrational stresses quantified

Force application to tissues in a bioreactor is quite common, but quantifying vibrational stresses has been attempted in very few bioreactors (Bacabac *et al.*, 2006; Tanaka *et al.*, 2003; Titze *et al.*, 2004b; Titze *et al.*, 2005; Titze *et al.*, 2004a). Bacabac and colleagues determined regions of quasi-steady-state dynamic flow in a parallel plate flow bioreactor but the bioreactor was limited to 5–30 Hz vibrations. Tanaka used a compression bioreactor that delivered 1 second periodic compression, white noise compression of 0–50 Hz, or a combination of the two at 3000 and 300 μstrain, respectively (Tanaka *et al.*, 2003). Responding movement of the osteoblast seeded collagen was not reported. Titze and colleagues developed a bioreactor for vocal fold tissues that delivered low frequency (<5 Hz) elongation and high frequency (up to 100 Hz) transverse movement to engineered tissue ribbons (Titze *et al.*, 2004a). Analytical calculations of standing waves matched fairly well with experimental displacements (Titze *et al.*, 2005). Shear strain maxima were 0.1–0.2 up to 70 Hz and 0.05 at 130 Hz, with a strain gradient along the length of the ribbon. This shear strain and frequency is at the lower boundary of physiologic vocal fold vibration. Instrumental limitations precluded larger strain delivery, and viscous and elastic tissue measurements, while in the bioreactor, were still not feasible.

The rheometer bioreactor used in the present investigation is capable of delivering vibrational shear strains observed during vocal fold movement and directly measuring viscoelastic properties. A commercially built rheometer, the Bohlin CVO 120 (Malvern Instruments, Worcestershire, UK), exposes developing tissue to physical stimuli and mea-

sures rheologic properties. A custom designed stainless steel cup mounts directly onto the base plate of the rheometer so that the upper plate and cell-seeded substrate are submerged in nutrient medium. The upper plate then makes contact with the substrate and the torque motor is programmed to apply the desired shear stress or strain. The rheometer bioreactor is capable of delivering 1–2 Hz shear strains, on the time scale of pitch change, or dynamic shear strains at low speaking frequencies, currently limited to 150 Hz. Titze, Klemuk, and Gray (2004b) determined that a range of strains is possible depending on four parameters: shear elastic modulus of the sample G', sample radius R, sample thickness h, and system inertia I. These four terms shift the location of the motor/plate/sample (i.e., system) resonant frequency. This frequency is defined as the maximum strain per N m torque. It occurs when measuring properties of viscoelastic solids. The present study aimed to take advantage of the large strain potential at the system resonant frequency when exposing engineered substrates to vocal fold-like amplitudes and frequencies. By increasing radius and decreasing thickness, we hypothesized that this resonant frequency would be in the desired low sonic frequency range and that strains of near 0.5 could be administered to a living tissue.

D. Rheologic measurement in the rheometer bioreactor

Viscoelastic measurement of tissues is a required function of a vocal fold tissue bioreactor. The hypothesis is that cellular and extracellular response to vibration exposure, as is experienced in the vocal fold, will be reflected in rheologic change to the tissue as a whole. Work from several research groups support this hypothesis in terms of both passive and active environments. Cells had different shapes, elastic properties of the ensuing tissue, and mechano-transduction when viscoelastic properties of their seeded substrate were varied (Discher *et al.*, 2005; Engler *et al.*, 2006; Takai *et al.*, 2004). Cells also became more or less stiff depending on the frequency, amplitude, or duration of the applied force (Janmey and McCulloch, 2007; Matthews *et al.*, 2006). The present study shows that reliable rheologic properties of engineered vocal fold tissues can be measured in the bioreactor, a functionality that allows for future investigations of cellular mechanical response to vibration.

III. METHODS FOR MEDIUM EVALUATION

A. Fibroblast cell culture on 3-D porous Tecoflex® substrates

Established human tracheal scar fibroblast cell lines (T31) were obtained from the Division of Otolaryngology—Head and Neck Surgery, Department of Surgery, University of Utah, Salt Lake City. This cell line, originating from a 16 year old male, was shown to be phenotypically stable through passage 10 (Thibeault *et al.*, 2002), but was always used in lower passages, usually passage five to prevent senescence effects. Tracheal scar fibroblast cells were used rather than vocal fold fibroblasts, because the vocal fold fibroblasts were found to be phenotypically unstable at low passages (Thibeault *et al.*, 2002).

Elastomeric TFX substrates, made from TFX S-80, were supplied by Dr. Patrick Tresco at the Keck Tissue Engineering Center, University of Utah with dimensions of less than 1 mm thick, and 1 mm$\times$1 mm squares or 2 mm$\times$2 mm squares. The substrates were sterilized by cold ethylene oxide treatment. They were exposed overnight to 20 μg/ml fibronectin solution in Dulbecco's Phosphate Buffered Saline (PBS), then washed and placed in a petri dish. Cells were passaged, dissociated by trypsin treatment, counted in a hemocytometer, and resuspended at a concentration of 2×10^6 cells/ml. Cells were uniformly seeded onto each scaffold, as was described by Webb *et al.* (2003), by releasing two aliquots of cell suspension from one end to the other at constant speed with a pipette. The substrates were incubated 30 min at 37 °C, turned over, and another aliquot of cell suspension seeded in the same manner on the opposite side. Substrates were incubated for 30 min, and two more aliquots were introduced to the top surface in the same manner. The final concentration was 1.0×10^5 cells and 5.0×10^5 cells per matrix for the smaller and larger squares, respectively. Substrates were transferred to a fresh petri dish and cultured for 1–2 weeks with periodic exchange of the nutrient medium. Experiments commenced when cells were elongated across the scaffold pores, as observed under indirect microscopy.

B. Experimental conditions

Cell viability and medium characteristics were evaluated for six Hepes concentrations ranging from inadequate to excessive, and under two less than ideal environmental conditions. For the first environmental condition, Hepes concentrations of 0, 4.5, 13, 22, and 30 mM were included in cell nutrient media, seeded substrates were placed in the bioreactor stainless steel cup, and were immersed in a given conditional medium, all of which contained gentamicin and fungizone to guard against contaminants. A specially designed acrylic lid was used in all experiments. The lid diameter matched that of the cup, rested flatly on the top of the cup, and included a 2 mm lip around the perimeter to keep it in place. The lid also had a 15 mm diameter hole in the center so that, during a rheometer bioreactor experiment, contact between the moving rheometer shaft and the lid was avoided. The covered cup was then set in a waterbath environment for 4 h. For the second environmental condition, the 0 and 22 mM Hepes concentrations were repeated and additional concentrations of 0.89, 43, and 82 mM were included in cell nutrient media to better detect upper and lower limits of Hepes concentration in terms of pH and cell viability. Cell-seeded substrates were placed in the bioreactor cup, were immersed in the different media, covered with the same acrylic cover, and allowed to set in a laminar flow hood for a slightly longer period of time (6 h) to determine the cellular impact of an extra 2 h. Measurements of pH (Orion Triode pH electrode and model 420A benchtop pH meter, Thermo Fisher Scientific, Waltham, MA) were made at the start and end of each experiment. Osmolality (Advanced Micro-Osmometer model 3300, Advanced Instruments, Inc, Norwood, MA) was measured at the start of every experiment and at the end of the 4 h experiments.

For all cell counts, each Tecoflex®-seeded disk was trypsinized, released cells were centrifuged, resuspended, mixed with trypan blue, and counted using a hemocytometer. Viable cells excluded the dye and were transparent. Nonviable cells absorbed the dye and were blue in color. Viable cell fraction, defined as the ratio of counted cells with no dye uptake to total counted cells, was determined at 0 h for each condition, at 4 h for the waterbath conditions, and 6 h for the laminar flow hood conditions.

IV. RESULTS AND DISCUSSION FOR MEDIUM EVALUATION

Initial viable cell fractions, pH, and osmolality readings were taken for each of the conditions so that the n value was fairly high and the result quite accurate. At time=0 h, mean viable cell fraction and standard error (SEM) were 0.91 ± 0.02 ($n=16$), mean and standard error pH were 7.07 ± 0.01 ($n=18$), and mean and standard error osmolality were 294 ± 2 mOsm/kg ($n=5$). A control seeded substrate was kept in a 37 °C humidified and 5% CO_2 environment incubator with 0% Hepes in the medium for the 4 h experiment, and another control exposed to the same condition was kept in an incubator for the 6 h experiment. At the end of the two time periods, their pooled mean ± SEM living fraction was 0.87 ± 0.07 ($n=2$), mean ± SEM pH was 7.2 ± 0.1 ($n=4$), and mean ± SEM osmolality was 297 ± 2 mOsm/kg ($n=2$). As expected, under standard cell culture growth conditions, where environmental carbon dioxide and humidity were controlled, living cell fractions reflected physiologically ideal pH and osmotic values.

Cell viability, causative pH, and osmolality were then quantified under less ideal conditions, a temperature controlled waterbath, and a laminar flow hood. Figure 1(a) shows that Hepes concentrations of 0.89–43 mM maintained a 0.64–0.91 viable cell fraction for 4–6 h in a water bath or a laminar flow hood. Live fractions were 0.31–0.48 at the end of the experiment when no Hepes was used for either environmental conditions. When 82 mM Hepes was used the live fraction was 0.24. These values are much lower than the initial live fraction or the incubator control fractions, thus ruling out their use for rheometer bioreactor studies.

Satisfactory Hepes concentration range further reduced when considering pH and osmolality measurements. Figure 1(b) shows Hepes concentrations 4.5–82 mM maintained a pH of 7.1–7.5, but the 0.89 mM concentration returned a pH of 8.26 thus eliminating this low concentration. The 82 mM concentration is not acceptable because of the very low cell viability shown in Fig. 1(a). When considering osmolality readings, values increased regardless of the Hepes concentration after 4 h in the waterbath [Fig. 1(c)]. Change in osmolality during this time, shown in Fig. 1(d), was less than 20 mOsm/kg for the 4.5–22 mM Hepes conditions and reached a minimum change of 9 mOsm/kg when 4.5 mM Hepes was used. Based on the Potter and DeMarse (2001) study with neural cells, osmolality changes greater than 8.3 mOsm/kg in a 4 h period should be lethal. But when using 13.2 and 22 mM Hepes, both having osmotic changes twice the 8.3 value, human fibroblast cells exhibited maximum viability. For this reason, 4.5–22 mM Hepes concentrations were considered acceptable for less than ideal environmental conditions and for experiments lasting 4 h or less.

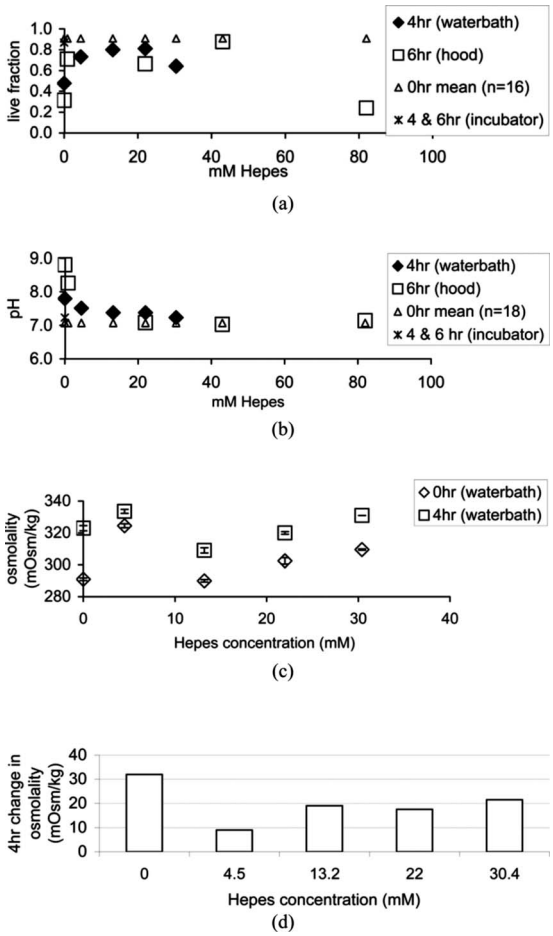


FIG. 1. Live cell fractions (a), as determined using trypan blue and hemocytometer counting, pH readings (b), and osmolality (c), as a function of Hepes concentration in cell culture medium. Environmental conditions were in a 37 °C waterbath for 4 h, in a laminar flow hood for 6 h, and in a 37 °C, 5% CO_2, humidified incubator at 4 and 6 h. The 0 hour control was plotted across Hepes concentrations as a reference. (d) Osmolality change for five different Hepes concentrations.

The two experimental environments each lacked elements of the ideal incubator condition. Both were equally deficient in atmospheric CO_2, where levels were presumably the same as the atmosphere, less than 1%. The waterbath was humid, nonsterile, and temperature controlled at 37 °C. The laminar flow hood was dry, sterile, and room temperature at approximately 23 °C. With "waterbath" cell viability being higher than "hood" cell viability at both the 0% and 22% Hepes conditions, lack of sterility in the waterbath was most likely managed by gentimicin and fungizone added to the medium. Neither the engineered tissue nor the surrounding medium ever made contact with the water in the waterbath, so it is unlikely that direct contamination occurred in the experiment. Alternatively, the dry air, lower temperature, and 2 h extension of the test in the hood seemed to create a harsher environment for cells in the TFX scaffold.

Environmental conditions of the rheometer bioreactor were also less than ideal. Atmospheric CO_2 levels were assumed to be low, humidity was that of the ambient environ-

ment, surfaces were wiped down with 70% ethanol for nominal sterility, and temperature was controlled at 37 °C. An additional complication was a steady flow of purified, dry air traveling through an air bearing around the motor chamber and exiting the chamber directly above the rheometer cup. The air flow allows the electromagnetic motor to float in a virtually frictionless environment for stress delivery and strain detection with high precision. Appliances or accessories making contact with the chuck or shaft from which the air originates, in an attempt to redirect airflow away from the cup, would compromise accurate motor function and possibly damage the instrument. The acrylic cover, placed on top of the cup, blocked some of the air flow from making contact with the cell environment. In order to compensate for additional medium evaporation due to the air flow, we tested medium exchange rates of every half hour and every 2 h over a 6 h period using 8.9 mM Hepes concentration in the medium. Pooled mean cell viability ± SEM for both exchange conditions at 2 and 4 h was 0.77 ± 0.005 ($n=4$). At 6 h, the viable cell fraction was 0.67 ± 0.01 ($n=2$).

Cell viability in the rheometer bioreactor could be improved in future studies by considering gas tensions. Hypoxia is a potential condition in the rheometer bioreactor, because tissues are compressed. Hypoxia occurs when oxygen tension in tissues is 5–15 mmHg O_2 rather than normal levels of 30–90 mmHg O_2 (Allen and Bhatia, 2003). The condition leads to anaerobic metabolism and reduction of mRNA translation, protein synthesis, and gene expression (Harris, 2002). Obradovic and colleagues found that gas exchange and medium replacement, three times per week, was adequate for generating a dense, well distributed cartilage tissue in a rotating bioreactor over a 5 week period, whereas no gas exchange over the same time period generated a sparse and poorly distributed cartilage tissue (Obradovic *et al.*, 1999). Our bioreactor design is not yet capable of maintaining tissues for this length of time. Based on these results, oxygen tension and medium nutrients, with the control of pH, for short term vibration dosing experiments of the present study were most likely adequate and trypan blue viability was sufficient for detecting lethal responses to the environment. Future improvements of the rheometer bioreactor, in order to conduct experiments lasting days to weeks, will require gas tension and metabolic measures.

V. SEEDED SUBSTRATES SUBJECTED TO THREE VIBRATION REGIMES

A. Instrument calibration

Instrument calibration procedures were performed in order to reduce systematic error. Temperature was set with a water-circulating temperature control unit ±0.5 °C. Its accuracy was confirmed using a Cannon DSR temperature probe (State College, PA). Concentricity and parallelism of the rheometer, was calibrated by Malvern Instruments technicians and by the first author on site as per manufacturer instructions. Trueness of machined plates was determined by The University of Iowa Medical Instruments staff. Potential errors in the empirically determined motor inertia value, phase error, and gap calibration were also accounted for (Klemuk, 2007). The motor inertia value and phase variabilities were each ±3%. Gap calibrations returned Newtonian viscosity measurement variabilities of ±7%. Total variability attributed to these calibrations is approximately ±15%.

B. Compression determination

Preliminary observations while measuring the viscoelastic properties of TFX substrates using a stress-controlled rheometer showed a significant lack of adherence between the sample and rheometer plates (Klemuk *et al.*, 2001). For the present investigation, the moving stainless steel parallel plate was bead-blasted in order to diminish slippage while still being able to sterilize the plate. 400grit wet/dry sandpaper was adhered in the bottom of the cup rather than cost prohibitive sandblasting. The sandpaper surface was rinsed with 70% ethanol and allowed to air dry prior to experimentation. Sample compression was also observed to reduce slippage and measurement variability, but possible drawbacks included changes to the substrate rheology and adverse cell viability. To evaluate sample compression, 30 mm diameter TFX disks were prepared in the same manner prior to cell seeding and allowed to incubate in cell nutrient medium for 48 h without cells. Thickness was determined using the rheometer and a 30 mm diameter plate. The plate was lowered onto a sample with a nominal gap of 1.00 mm. Gap setting was reduced incrementally until the normal force strain gauge reading rose, indicating that sample contact was attained. Sample thickness was then recorded as the gap setting and subsequent compressions, 5%, 10%, 20%, and 40%, were converted to the appropriate gap. Amplitude sweeps were also performed at each level of compression to select driving amplitude in the linear viscoelastic (LVE) region. LVE region was maintained for strains of 0.05 or less. Frequency sweeps were performed on each TFX disk at each sample compression. Input torque, output displacement, and phase were used to calculate linear viscoelastic properties.

Cell viability was also tested under varying compression conditions. T31 cells were seeded onto TFX substrates as described earlier. Each disk was placed in the rheometer-mounted cup (Bohlin CVO 120) and subjected to 0%, 10%, 20%, or 70% compression with a 30 mm parallel plate. Compression was applied for 6 h. Cell viability was measured by trypan blue staining.

C. Tissue preparation and vibration regimes

TFX substrates, having a 45 mm diameter, were seeded with T31 tracheal fibroblast cells as specified earlier and allowed to grow for approximately 2 weeks. Medium environment during the vibration experiments included the upper limit of Hepes, 22 mM, gentamicin and fungizone to protect against contamination, and a 50% medium exchange every 30 min. A sterilized, bead-blasted 45 mm diameter stainless steel plate and 400grit wet/dry sandpaper were used and sample compression was 10%, as determined by compression studies. Each substrate was subjected to one of three different vibration regimes for 2 h in the rheometer bioreactor where frequency, amplitude, and duty ratio (the on and off cycling of vibration) were specified. Using rheometer

software, amplitude was controlled and resulting tissue strain was measured. Peak stresses were adjusted at the beginning of a given experiment so that resulting tissue strains were near 0.5 for each of the vibration regimes. Duty ratio was also controlled using the rheometer software. A 25% duty ratio cycled vibration on for 45 s and off for 135 s, and the 75% duty ratio cycled vibration on for 135 s and off for 45 s. These duty ratios were selected to test the machine's ability to deliver two different duty ratios rather than constant vibration, as was done in other studies. The three vibration conditions were 1 Hz 25% duty ratio to test for minimal tissue disruption, high frequency 25% duty ratio, and high frequency 75% duty ratio to test for maximal tissue disruption.

High frequency for a given sample was at or near the resonant frequency of the motor/plate/sample system, the frequency at which maximum strain occurs with the least amount of torque. The resonant angular frequency ω_0 for an undamped oscillator is written in terms of the shear elastic modulus G', the sample radius R, the sample thickness h, and the motor and plate inertia I:

$$\omega_0 = \sqrt{\frac{\pi G' R^4}{2hI}}. \tag{1}$$

This is the same as Eq. (12) in Titze *et al.* (Titze *et al.*, 2004b). Resonant frequency of the motor/plate/sample system was determined using Eq. (15) from Titze *et al.* (2004b). A frequency sweep was performed prior to initiating the vibration experiment at strains in the LVE region: 0.01–0.015. The resulting G' and G'' values at each frequency ω were then inserted into the following equation, and magnitude of strain per N m torque MAG$[\gamma/T]$ was calculated.

$$\mathrm{MAG}\left[\frac{\gamma}{T}\right] = \frac{1.5}{G'\pi R^3} \frac{1}{\sqrt{\left[1 + \left(\frac{\omega}{\omega_0}\right)^4 - 2\left(\frac{\omega}{\omega_0}\right)^2 + \left(\frac{G''}{G'}\right)^2\right]}}. \tag{2}$$

Using values from Eq. (2), each vibration condition was subsequently performed on three different substrates.

No-vibration controls in the bioreactor were seeded substrates measuring 30 mm in diameter ($n=3$). Controls were subjected to 10% compression for 2 h, were not vibrated, and viscoelastic data were subsequently collected at 0.01–100 Hz. Each substrate was then cut into smaller pieces. For each substrate piece, cells were trypsanized, centrifuged, mixed with trypan blue, and counted using a hemocytometer.

VI. RESULTS AND DISCUSSION FOR THREE VIBRATION REGIMES

A. Compression effects to TFX rheology and cell viability

Compression directly affects rheologic properties of porous TFX substrate, saturated with cell nutrient medium. Displayed in Fig. 2 are mean elastic and viscous moduli ($n=3$) at four different frequencies. Magnitudes increased with increasing compression, where the largest increase occurred

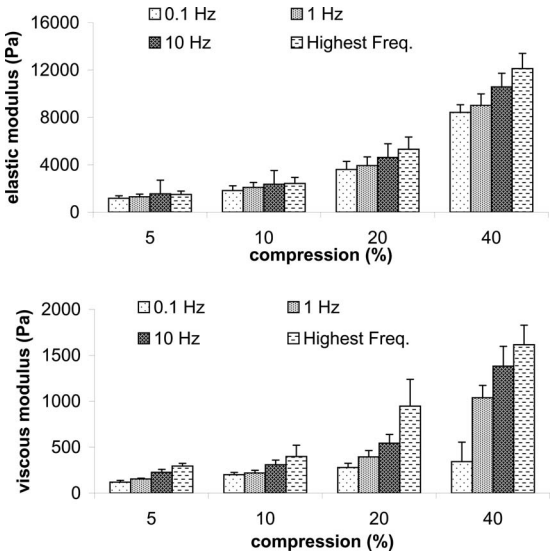


FIG. 2. Mean and standard error of elastic and viscous moduli ($n=3$) as a function of TFX substrate compression at four frequencies. Highest frequency was 35, 43, 67, and 70 Hz for 5%, 10%, 20%, and 40% compression, respectively.

from 20% to 40% compression. For this compression doubling, G' and G'' more than doubled for every frequency with only one exception: G'' at 0.1 Hz. The least rheologic increase occurred with a 5% to 10% compression increase, where values were about half again as much at 10% compression as compared to 5% compression. G' variability (SEM/mean $\times$ 100) was, on average, $\pm$18%, $\pm$20%, $\pm$19%, $\pm$10% as compression increased, and G'' variability was $\pm$11%, $\pm$18%, $\pm$20%, and $\pm$26% as compression increased. Within a given compression, gap settings were within 0.2 mm for each of the three samples. If the gap had been kept constant rather than percent compression, 20% compression data from samples 1 and 2 would have been averaged with 40% compression data of the third sample, all having a gap of 0.5 mm. In that case, mean elastic and viscous moduli for the 20% compression would nearly double and average variabilities of G' and G'' would be $\pm$42% and $\pm$45%, respectively.

A natural log fit $y=a \ln x+b$ to mean elastic moduli versus frequency showed that the slope a steadily increased with increasing compression as well (Table I). Viscous moduli of medium-saturated TFX also increased with increasing compression, but when fitting the dynamic moduli $\eta'=G''/\omega$ to a power law function $y=ax^b$ (Table II), the declination rates were about the same. The steeper elastic moduli slopes with increasing compression may indicate a change in proportional solid to liquid presence, or changes in transient hydrogen bonding and van der Waals forces. With increased compression, relative positions of hard and soft molecular chains within the TFX may have clumped, causing the structure to become stiffer. It is unlikely that compression or rheology broke chemical bonds. If the substrate had torn, it would be more likely that property values decrease, which they did not. Repeated rheological measures on the same sample and scanning electron microscopy imaging would help in the future to determine the cause of change. TFX is described as an

TABLE I. Coefficients a and b and corresponding coefficient of determination R^2 for natural log curve fit $y = a \ln x + b$ to each mean elastic modulus data set as a function of frequency.

Elastic modulus fits for the following conditions		$y = a \ln x + b$		
		a	b	R^2
TFX compression conditions	5%	58.59	1331	0.869
	10%	97.42	2078	0.944
	20%	204.7	4053	0.971
	40%	604.5	9217	0.984
Cell-seeded TFX pre-vibration at 10% compression	Bioreactor control	—	—	—
	1 Hz 25% duty ratio	—	—	—
	High freq. 25% duty ratio	152.6	2960	0.980
	High freq. 75% duty ratio	172.7	3459	0.983
Cell-seeded TFX post-vibration at 10% compression	Bioreactor control	165.2	3433	0.969
	1 Hz 25% duty ratio	156.2	2994	0.986
	High freq. 25% duty ratio	166.8	3192	0.987
	High freq. 75% duty ratio	205.8	3846	0.989

elastomer where, under tension, it behaves as an elastic solid for up to 100% strain. We therefore expect that viscosity originated from the cell nutrient medium and hydrophilic interactions between the medium and the substrate. This is supported by magnitude increases in viscosity without a noticeable change in the declination rate. As compression increases, free medium, having the viscosity near water, leaves the porous structure. Remaining medium molecules are too tightly bound to the substrate to exit but probably soften the substrate structure and give it a more viscous property.

Another effect of compression is an increase in the upper frequency at which valid rheologic data occur. The highest frequency for each compression was 35, 43, 67, and 70 Hz for 5%, 10%, 20%, and 40% compression, respectively. This high frequency limit was determined by locating the system resonant frequency using the original torque and displacement signal, and then doubling that frequency (Titze *et al.*, 2004b). In the case of 40% compression, a doubling of the resonant frequency was 116, but G' and G'' values above 70 Hz were in rapid decline, an indicator of an inertially controlled system. A 10%–20% compression boosts frequency a bit over the 5% compression and material properties are still similar. A 10%–20% compression also holds the sample in place when high strains are applied. When no compression and smooth plates were used in preliminary studies, high strains on TFX caused the material to roll up and be ejected from between the plates. The system resonant frequency also increases when using a larger diameter plate and sample, as in the vibration experiments.

The resulting viable cell fractions were 0.94, 0.82, 0.85, and 0.53 for 0%, 10%, 20%, and 70% compression respectively. The 0% compression result was in line with the baseline cell viability reported in Fig. 1, the 10% and 20% compression viabilities were similar, and the 70% compression was not conducive to cell survival. It was concluded that a 10%–20% sample compression was acceptable, and every sample should be compressed the same amount because of the strong rheologic dependence on compression.

Accurate rheological data collection is a critical component for evaluating engineered vocal fold tissues in a biore-

TABLE II. Coefficients a and b and corresponding coefficient of determination R^2 for power law curve fit $y = ax^b$ to each mean dynamic viscosity data set as a function of frequency.

Dynamic viscosity fits for the following conditions		$y = ax^b$		
		a	b	R^2
TFX compression conditions	5%	26.56	−0.8523	0.994
	10%	40.39	−0.9034	0.996
	20%	67.29	−0.8587	0.998
	40%	154.8	−0.8682	0.992
Cell-seeded TFX pre-vibration at 10% compression	Bioreactor control	—	—	—
	1 Hz 25% duty ratio	—	—	—
	High freq. 25% duty ratio	54.48	−0.8987	0.998
	High freq. 75% duty ratio	62.40	−0.8894	0.998
Cell-seeded TFX post-vibration at 10% compression	Bioreactor control	63.32	−0.9024	0.999
	1 Hz 25% duty ratio	59.28	−0.9333	0.996
	High freq. 25% duty ratio	62.24	−0.9176	0.998
	High freq. 75% duty ratio	74.58	−0.909	0.999

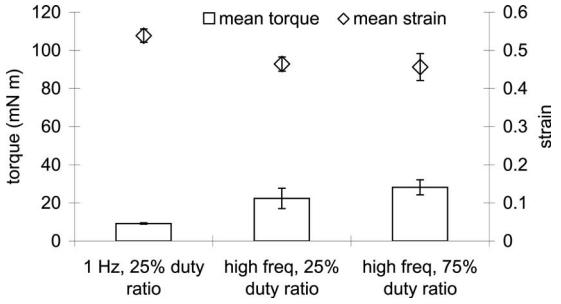


FIG. 3. Mean applied torque and responding mean strain ± SEM ($n=3$) of fibroblast-seeded TFX scaffold for three vibration conditions: 1 Hz and 25% duty ratio, high freq (62, 56, and 56 Hz) 25% duty ratio, and high frequency (68, 56, and 56 Hz) 75% duty ratio.

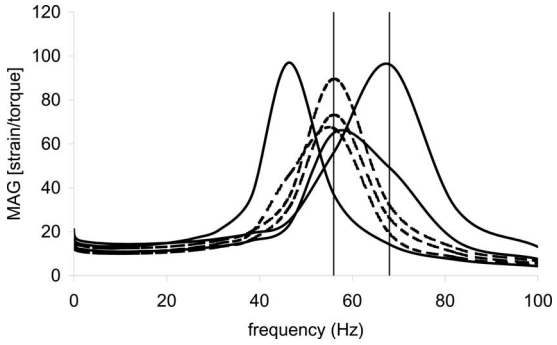


FIG. 4. Motor/plate/sample response, strain per N m torque, as a function of frequency for the high frequency samples. Solid curves are for 25% duty ratio samples and dashed curves are for 75% duty ratio samples.

actor. Biological tissues, composites of solid and liquid such as biocompatible substrates in cell nutrient medium, and dispersions used for vocal fold injection, are all problematic when collecting their rheologic data with a stress-controlled rheometer. Sample slippage is an unresolved problem in the field of rheology, has been observed by the current authors, and remains a difficult problem for complex material measurement (Klemuk, 2007; Sanchez-Reyes and Archer, 2003; Tapadia *et al.*, 2006). Sample compression results in the present study lead to the particular conclusion that TFX sample should be exposed to the same compression, 10%–20% rather than the same gap, because of compression dependencies and variations in substrate thickness. But the results also suggest that any porous biocompatible scaffold or biological tissue, where free water or medium can escape from pores and between parallel plates, should be evaluated in a similar way for the exact same reasons.

B. Quantifying vibration stimulation

Experimental results up to this point were now put to the test. T31 seeded TFX scaffolds were immersed in 22 mM Hepes gentamicin fungizone-added cell nutrient medium, pre-cut as a large diameter disk of 45 mm, and subjected to 10% compression using a bead-blasted parallel plate with sandpaper cup bottom. After calculating and viewing response curves using Eq. (2) and Excel software, vibration frequencies were chosen at or near the peak response frequency. For the 25% duty ratio condition, one sample vibrated at 62 Hz and two at 56 Hz. Peak response frequencies were 46, 56, and 68 Hz, respectively. For the 75% duty ratio condition, one sample vibrated at 68 Hz and two at 56 Hz. Peak response frequency was 56 Hz for each sample. Figure 3 shows mean ± SEM applied torque for each vibration condition ($n=3$) and the mean resulting strain ± SEM ($n=3$) reported by the rheometer software. A total of 30–40 points were recorded for each sample during the 2 h test. Torque was 10–35 mN m for all conditions—well below the maximum torque capacity of the rheometer, 120 mN m. The mean ratio of strain to torque for the 1 Hz condition was more than twice that of the mean ratio of the 56–68 Hz conditions, 58.6 and 21.6 $(\text{N m})^{-1}$, respectively. This was an unexpected result. System response curves calculated with Eq. (2) prior to high frequency vibration experiments indicated that 1–7 times more strain should have been achievable at or near the resonant frequency compared to 1 Hz vibration (Fig. 4). Later, system response was calculated using the raw torque and strain signal. The peaks were slightly shifted but there still should have been the 1–7 times more strain available at the higher frequency vibration regimes.

A nonplausible explanation is that vertical standing waves developed in the sample. If the gap h is sufficiently small with respect to the sample viscosity η, density ρ, and frequency f, vertical standing waves will not develop within the sample, thus satisfying the gap loading condition. Schrag (1977) and later Ferry (1980) wrote the criterion as

$$h \ll \sqrt{\frac{2\pi\eta}{\rho f}}. \tag{3}$$

Assuming a density of 1 g/cm^3 and complex viscosity of 10 Pa s and 4 Pa s at 56 Hz and 150 Hz, respectively, for seeded TFX, the largest gap of 0.86 mm in the present investigation is 39 times smaller and 12 times smaller than the right hand side of Eq. (3) at 56 Hz and 150 Hz, respectively. According to Schrag (1977), these both satisfy the gap loading criterion for "reasonable approximations." Instead, it is suspected that the increased torque at higher frequency is due to slippage between the sample and the bead-blasted plate and 400grit sandpaper covered cup base. Preliminary slippage studies showed that a lower grit sandpaper (i.e., rougher sandpaper), reduced rheologic variability of TFX measurements (Klemuk, 2007). It was not used, however, because particles might have infiltrated the sample during vibration experiments and affected biological outcomes. Despite this potential slip effect, the desired strains of 0.4–0.5 were achieved using about $\frac{1}{4}$ of the available torque. These strains occurred at 56–68 Hz, frequencies much lower than the 150 Hz limit of the Bohlin CVO electromagnetic torque motor. It is therefore likely that large strains will be attainable at 150 Hz. In this study we did not customize rheometer controls beyond what is available to the typical rheometer user. More advanced controls will extend large amplitude vibrations up to 250 Hz in future investigations. Improvements to strain limits for vibration stimulation will include the use of more adherent, prototypic biomaterials that remain mechanically stable but eventually degrade, and use of alternative hard surface treatments for proper adhesion.

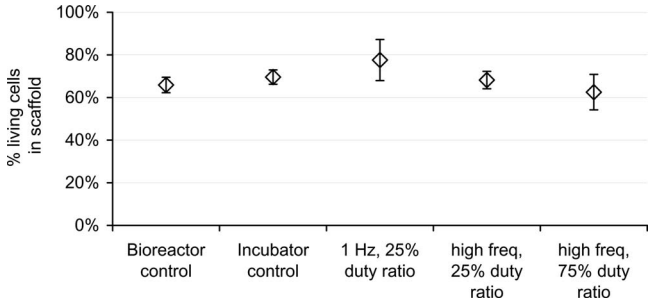


FIG. 5. Mean viable cell fractions (SEM), as determined using trypan blue and hemocytometer, at the completion of each 2 h test condition.

C. Cell viability under vibration

Viable cell fractions for vibration regime studies were comparable across all conditions, as shown in Fig. 5. The bioreactor control and the 1 Hz 25% duty ratio each had 10 samples, and the incubator control, the high frequency 25% duty ratio, and the high frequency 75% duty ratio each had five samples. All error bars corresponded to standard error, where standard deviation was normalized to the square root of n. The range of mean values was 0.63–0.78 and standard error bars overlapped for all conditions. Viability was similar to the 6 hour laminar flow hood, 22 mM condition (0.66) shown in Fig. 1, leading to the conclusion that viability was not affected by vibration conditions. Results from the present study indicate suitability and continued use of the rheometer bioreactor for tissue engineering studies.

D. Tissue rheology

Viscoelastic properties through 50 Hz were attainable while living tissues were in the rheometer bioreactor before and after vibration experiments. Values were in the range of vocal fold related biomaterial properties and they were in the range of unseeded TFX substrate properties. Shear elastic and viscous moduli increased with increasing frequency with a G' range of 2300–4600 Pa and a G'' range of 280–630 Pa across all vibration and control conditions (Fig. 6). Coefficients of natural log and power law fits to mean elastic moduli and to mean dynamic viscosities, respectively, were similar pre- and post-vibration and were coincident with unseeded TFX data at compressions between 10 and 20% (Tables I and II). G' and G'' values did not remain constant before and after vibration. This is not a function of the machine but rather that of sample changes, and will be a topic to consider for a future study.

G' and G'' magnitudes were $\frac{1}{2}$ to $\frac{1}{4}$ that of Cymetra and Radiesse, two biocompatible vocal fold dispersions typically injected into the lateral portion of the vocal fold to medialize tissue (Caton *et al.*, 2007; Chan and Titze, 1999). G' values were nearly an order of magnitude higher than hyaluronic acid based injectates, targeted for vocal fold mucosal injection, and cadaveric vocal fold mucosa. Dynamic viscosity values were 1–10 Pa s at 10 Hz and higher, corresponding to dynamic viscosities of two out of the ten male vocal fold mucosal samples from Chan and Titze (1999). G' variability (SEM/mean$\times$100) was, on average, $\pm$6.2%, $\pm$15%, $\pm$10%, $\pm$1.6% for the control, 1 Hz, high frequency 25% duty ratio, and high frequency 75% duty ratio, respectively. G'' variability was $\pm$1.4%, $\pm$8.7%, $\pm$4.2%, and $\pm$1.5% as compression increased. The loss tangent G''/G' was also in an acceptable range, consistently 0.13. Upper frequency limits of rheologic measurement are only a current limitation. We have not extended to piezo motors in this paper, but with this technology, measurement and small amplitude vibration up to 1000 Hz will be possible.

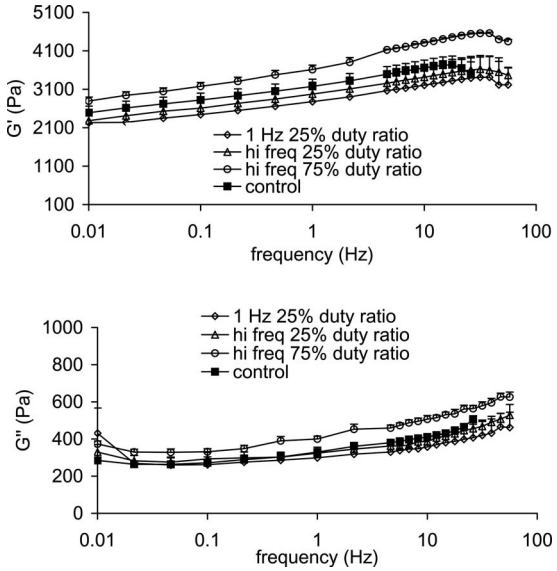


FIG. 6. Mean elastic moduli (top) and mean viscous moduli (bottom) of fibroblast-seeded TFX substrates after exposure to various vibration regimes. Error bars represent one standard error.

VII. SUMMARY AND CONCLUSION

Functionality of a rheometer as a bioreactor has been demonstrated for short term vibration experiments in the sonic region and with large strains. A stable biocompatible nondegradable substrate, known to support tracheal fibroblast cell growth, uniform distribution, and ECM production, was successfully seeded and exposed to three different vibration regimes. Cell fraction viability was maintained at approximately 0.7 by using 22 mM Hepes. Input stresses and output tissue deformation were quantified at 56–68 Hz and for strains of 0.4–0.5. Given the amount of torque used to generate the strains, it is likely that the rheometer can also generate the same strains at frequencies up to the instrument's capacity of 150 Hz. Finally, rheologic measurements were made while in the bioreactor before and after vibration exposure, demonstrating a unique and important measuring capability of this rheometer bioreactor. Improvements to the cell environment are ongoing to increase experimental times from several hours to several days, and expanded technologies are being incorporated to increase vibration and measurement frequencies well into the sonic region.

ACKNOWLEDGMENTS

This work was supported by NIH Grant Nos. DC004224 and DC008047 from the National Institute on Deafness and Other Communication Disorders.

Allen, J. W., and Bhatia, S. N. (**2003**). "Formation of steady-state oxygen gradients in vitro," Biotechnol. Bioeng. **82**, 253–262.
Andrews, K. D., and Hunt, J. A. (**2008**). "Upregulation of matrix and adhesion molecules induced by controlled topography," J. Mater. Sci.: Mater. Med. **19**, 1601–1608.
Andrews, K. D., Hunt, J. A., and Black, R. A. (**2008**). "Technology of electrostatic spinning for the production of polyurethane tissue engineering scaffolds," Polym. Int. **57**, 203–210.
Bacabac, R. G., Smit, T. H., Van Loon, J. J. W. A., Boulabi, B. Z., Helder, M., and Klein-Nulend, J. (**2006**). "Bone cell responses to high-frequency vibration stress: Does the nucleus oscillate within the cytoplasm?," FASEB J. **20**, 858–864.
Barron, V., Lyons, E., Stenson-Cox, C., McHugh, P. E., and Pandit, A. (**2003**). "Bioreactors for cardiovascular cell and tissue growth: Areview," Ann. Biomed. Eng. **31**, 1017–1030.
Caton, T., Thibeault, S. L., Klemuk, S., and Smith, M. E. (**2007**). "Viscoelasticity of hyaluronan and nonhyaluronan based vocal fold injectables: Implications for mucosal versus muscle use," Laryngoscope **117**, 516–521.
Chan, R. W., and Titze, I. R. (**1999**). "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021.
Chan, R. W., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.
De Zengotita, V. M., Schmelzer, A. E., and Miller, W. M. (**2002**). "Characterization of hybridoma cell responses to elevated pCO_2 and osmolality: Intracellular pH, cell size, apoptosis, and metabolism," Biotechnol. Bioeng. **77**, 369–380.
Discher, D. E., Janmey, P., and Wang, Y. (**2005**). "Tissue cells feel and respond to the stiffness of their substrate," Science **310**, 1139–1143.
Engler, A. J., Sen, S., Sweeney, H. L., and Discher, D. E. (**2006**). "Matrix elasticity directs stem cell lineage specification," Cell **126**, 667–689.
Ferry, J. D. (**1980**). *Viscoelastic Properties of Polymers*, John Wiley & Sons, New York.
Freed, L. E., and Vunjak-Novakovic, G. (**2000**). *Tissue Engineering Bioreactors*, in Principles of Tissue Engineering, edited by R. P. Lanza, R. Langer, and J. Vacanti (Academic Press, San Diego), pp. 143–156.
Freshney, I. R. (**2000**). *Culture of Animal Cells; A Manual of Basic Technique*, Wiley-Liss, New York.
Harris, A. L. (**2002**). "Hypoxia—A key regulatory factor in tumour growth," Nat. Rev. Cancer **2**, 38–47.
Janmey, P. A., and McCulloch, C. A. (**2007**). "Cell mechanics: Integrating cell responses to mechanical stimuli," Annu. Rev. Biomed. Eng. **9**, 1–34.
Klemuk, S. A. (**2007**). *Methodological Improvements in Mechanical Measurement of Vocal Fold-Related Tissues*, The University of Iowa, Iowa City, IA.
Klemuk, S. A., Titze, I. R., and Gray, S. (**2001**). "Methodology for straining vocal fold tissues at audio frequencies," J. Acoust. Soc. Am. **109**, 2412.
Matthews, B. D., Overby, D. R., Mannix, R., and Ingber, D. E. (**2006**). "Cellular adaptation to mechanical stress: Role of integrins, Rho, cytoskeletal tension and mechanosensitive ion channels," J. Cell Biol. **119**, 508–518.
Min, Y., Titze, I. R., and Alipour-Haghighi, F. (**1995**). "Stress-strain response of the human vocal fold ligament," Ann. Otol. Rhinol. Laryngol. **104**, 563–569.
Mulder, M. M., Hitchcock, R. W., and Tresco, P. A. (**1998**). "Skeletal myogenesis on elastomeric substrates: Implications for tissue engineering," J. Biomater. Sci., Polym. Ed. **9**, 731–748.
Obradovic, B., Carrier, R. L., Vunjak-Novakovic, G., and Freed, L. E. (**1999**). "Gas exchange is essential for bioreactor cultivation of tissue engineered cartilage," Biotechnol. Bioeng. **63**, 197–205.
Perlman, A., and Durham, P. (**1991**). *In Vitro Studies of Vocal Fold Mucosa During Isometric Conditions*, in Laryngeal Function in Phonation and Respiration, edited by T. Baer (Little, Brown, & Co, Boston), pp. 291–303.
Potter, S. M., and DeMarse, T. B. (**2001**). "A new approach to neural cell culture for long-term studies," J. Neurosci. Methods **110**, 17–24.
Sanchez-Reyes, J., and Archer, L. A. (**2003**). "Interfacial slip violations in polymer solutions: Role of microscale surface roughness 1," Langmuir **19**, 3304–3312.
Schrag, J. L. (**1977**). "Deviation of velocity gradient profiles from the "gap loading" and "surface loading" limits in dynamic simple shear experiments," Trans. Soc. Rheol. **21**, 399–413.
Takai, E., Costa, K. D., Shaheen, A., Hung, C. T., and Guo, X. E. (**2004**). "Osteoblast elastic modulus measured by atomic force microscopy is substrate dependent," Annu. Rev. Biomed. Eng. **33**, 963.
Tanaka, S. M., Li, J., Duncan, R. L., Yokota, H., Burr, D. B., and Turner, C. H. (**2003**). "Effects of broad frequency vibration on cultured osteoblasts," J. Biomech. **36**, 73–80.
Tapadia, P., Ravindranath, S., and Wang, S. Q. (**2006**). "Banding in entangled polymer fluids under oscillatory shearing," Phys. Rev. Lett. **96**, 196001.
Thibeault, S. L., Li, W., Gray, S. D., and Chen, Z. (**2002**). "Instability of extracellular matrix gene expression in primary cell culture of fibroblasts from human vocal fold lamina propria and tracheal scar," Ann. Otol. Rhinol. Laryngol. **111**, 8–14.
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R., Broadhead, K., Tresco, P. A., and Gray, S. (**2005**). "Strain distribution in an elastic substrate vibrated in a bioreactor for vocal fold tissue engineering," J. Biomech. **38**, 2406–2414.
Titze, I., and Durham, P. L. (**1991**). *Passive Mechanisms Influencing Fundamental Frequency Control*, in Laryngeal function in phonation and respiration, edited by Baer, Sasaki, and Harris (Little, Brown and Company, Boston), pp. 304–319.
Titze, I. R., Hitchcock, R. W., Broadhead, K., Webb, K., Li, W., Gray, S. D., and Tresco, P. A. (**2004a**). "Design and validation of a bioreactor for engineering vocal fold tissues under combined tensile and vibrational stresses," J. Biomech. **37**, 1521–1529.
Titze, I. R., and Hunter, E. (**2004**). "Normal vibration frequencies of the vocal ligament," J. Acoust. Soc. Am. **115**, 2264–2269.
Titze, I. R., Klemuk, S. A., and Gray, S. (**2004b**). "Methodology for rheological testing of engineered biomaterials at low audio frequencies," J. Acoust. Soc. Am. **115**, 392–401.
Vara, D. S., Salacinski, H. J., Kannan, R. Y., Bordenave, L., Hamilton, G., and Seifalian, A. M. (**2005**). "Cardiovascular tissue engineering: State of the art," Pathol. Biol. (Paris) **53**, 599–612.
Webb, K., Li, W., Hitchcock, R. W., Smeal, R. M., Gray, S. D., and Tresco, P. A. (**2003**). "Comparison of human fibroblast ECM-related gene expression on elastic three-dimensional substrates relative to two-dimensional films of the same material," Biomaterials **24**, 4681–4690.

Estimation of the acoustic impedance of lung versus level of inflation for different species and ages of animals

Michael L. Oelze[a] and Rita J. Miller
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

James P. Blue, Jr. and James F. Zachary
Department of Pathobiology, University of Illinois, 2001 South Lincoln Avenue, Urbana, Illinois 61802

William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

(Received 3 December 2007; revised 14 July 2008; accepted 17 July 2008)

In a previous study, it was hypothesized that ultrasound-induced lung damage was related to the transfer of ultrasonic energy into the lungs (W. D. O'Brien *et al.* 2002, "Ultrasound-induced lung hemorrhage: Role of acoustic boundary conditions at the pleural surface," J. Acoust. Soc. Am. **111**, 1102–1109). From this study a technique was developed to: 1) estimate the impedance (Mrayl) of fresh, excised, *ex vivo* rat lung versus its level of inflation (cm H_2O) and 2) predict the fraction of ultrasonic energy transmitted into the lung (M. Oelze *et al.* 2003, "Impedance measurements of *ex vivo* rat lung at different volumes of inflation." J. Acoust. Soc. Am. **114**, 3384–3393). In the current study, the same technique was used to estimate the frequency-dependent impedance of lungs from rats, rabbits, and pigs of various ages. Impedance values were estimated from lungs under deflation (atmospheric pressure, 0 cm H_2O) and three volumes of inflation pressure [7 cm H_2O (5 cm H_2O for pigs), 10 cm H_2O, and 15 cm H_2O]. Lungs were scanned in a tank of degassed 37 °C water. The frequency-dependent acoustic pressure reflection coefficient was determined over a frequency range of 3.5–10 MHz. From the reflection coefficient, the frequency-dependent lung impedance was calculated with values ranging from an average of 1.4 Mrayl in deflated lungs (atmospheric pressure) to 0.1 Mrayl for fully inflated lungs (15 cm H_2O). Across all species, deflated lung (i.e., approximately 7% of the total lung capacity) had impedance values closer to tissue values, suggesting that more acoustic energy was transmitted into the lung under deflated conditions. Finally, the impedance values of deflated lungs from different species at different ages were compared with the thresholds for ultrasound-induced lung damage. The comparison revealed that increases in ultrasonic energy transmission corresponded to lower injury threshold values.
 [DOI: 10.1121/1.2973186]

PACS number(s): 43.80.Cs, 43.80.Gx, 43.80.Jz [FD] Pages: 2340–2352

I. INTRODUCTION

Ultrasound-induced lung hemorrhage has been observed in animal experiments at exposure levels corresponding to those used for ultrasound examinations in humans (incidental exposure) (Child *et al.*, 1990; Hartman *et al.*, 1990; Penney *et al.*, 1993; Raeman *et al.*, 1993, 1996; Frizzell *et al.*, 1994, 2003;Tarantal and Canfield, 1994; Zachary and O'Brien, 1995; Baggs *et al.*, 1996; Holland *et al.*, 1994; O'Brien and Zachary, 1997; Dalecki *et al.*, 1997a, 1997b; AIUM, 2000; O'Brien *et al.*, 2000, 2001a, 2001b, 2003a, 2003b, 2004, 2006; Zachary *et al.*, 2001a, 2001b). Specifically, sharply defined thresholds on the order of 1 MPa were found with 10 μs length pulses and larger thresholds at 4 MHz than at 1 MHz (Child *et al.*, 1990; Hartman *et al.*, 1990). In addition, the lung damage was not found to be progressive, i.e., hemorrhage occurred only during exposure and did not continue after the termination of exposure (Penney *et al.*, 1993). In further experiments, the threshold for lung damage was found to depend mainly on the pressure level and not on the exposure duration (Raeman *et al.*, 1993, 1996). Similar results were obtained in larger animals, i.e., monkeys and swine (Tarantal and Canfield, 1994; Baggs *et al.*, 1996). The dependence of ultrasound-induced lung damage was examined in terms of the age dependence of animals (Dalecki *et al.*, 1997a, 1997b; O'Brien *et al.*, 2003b). Although the thresholds for ultrasound-induced lung damage were similar for neonatal, juvenile, and adult mice, the sizes of the suprathreshold hemorrhages were significantly larger in adult mice than in neonatal or juvenile mice (Dalecki *et al.*, 1997b). Furthermore, lung damage thresholds were examined across species with similar trends observed (O'Brien and Zachary, 1997). Ultrasound-induced lung damage was examined versus various ultrasound parameters including pulse duration (O'Brien *et al.*, 2001a, 2003a), pulse repeti-

[a]Author to whom correspondence should be addressed; electronic mail: oelze@uiuc.edu

tion frequency (O'Brien *et al.*, 2001a), beamwidth (O'Brien *et al.*, 2001b), and use of contrast agents (O'Brien *et al.*, 2004).

As a result of previous studies, several mechanisms for ultrasound-induced lung hemorrhage have been hypothesized including heating, inertial cavitation, and breaking of tissue layers from acoustic transfer of energy. In the case of tissue heating, several studies have indicated that heating cannot explain lung damage (Hartman *et al.*, 1992; Zachary *et al.*, 2006). The hypothesis that inertial cavitation was the mechanism underlying ultrasound-induced lung hemorrhage was also tested (Frizzell *et al.*, 2003; O'Brien *et al.*, 2000, 2004). These studies quantified lung damage thresholds when pulse polarity, energy, and overpressure values were varied. The studies suggested that inertial cavitation was not responsible for ultrasound-induced lung hemorrhage; however, the debate on the role of inertial cavitation in lung damage remains (Raeman *et al.*, 1997; Carstensen *et al.*, 2000; O'Brien *et al.*, 2000; Frizzell *et al.*, 2001a, 2001b, 2003; Apfel, 2001a, 2001b).

The third mechanism hypothesized to be responsible for ultrasound-induced lung damage is the breaking of tissue layers due to the transfer of acoustic energy. This could be due to a radiation-type force where tight junctions are transiently or permanently opened due to ultrasound exposure allowing hemorrhaging in the lung. In the case of a transiently opened junction, the hemorrhaging would stop after the exposure has terminated. This result would support early reports that ultrasound-induced lung damage was not progressive and did not continue after the termination of exposure (Penney *et al.*, 1993). Later reports of superthreshold behavior in rats indicated that following exposure to pulsed ultrasound, lesions at zero days postexposure were characterized by areas of acute alveolar hemorrhage under the visceral pleural surface but without pleural or septal injury, resulting in no long-term residual effects on the lung (Zachary *et al.*, 2001a). Superthreshold corresponds to pressure levels used in exposure above the threshold determined in previous studies to yield lung damage in rats (O'Brien *et al.*, 2001a, 2001b).

In vivo experiments in rats demonstrated that deflated lung increased the occurrence and degree of hemorrhage when compared to a fully inflated lung (O'Brien *et al.*, 2002). In the *in vivo* experiments, deflated lung corresponded to functional residual capacity or about 40% of total lung capacity (TLC). In this study, the respiratory systems of 12 rats were arrested (chemically inhibited and then ventilator used), and then the lungs at three volumes of inflation (inflated: approximately tidal volume; half-deflated: half-tidal volume; deflated: lung volume at functional residual capacity or about 40% of TLC) were exposed to two previously determined (from anesthetized breathing rats) superthreshold ultrasound conditions. Deflated lungs were more easily damaged than half-deflated lungs, and half-deflated lungs were more easily damaged than inflated lungs (inflated lungs had no lesions). The acoustic impedance difference between intercostal tissues and lungs was hypothesized to be much less for the deflated lung condition, suggesting that the extent of lung damage was related to the amount of ultrasonic energy that was propagated across the pleural surface boundary (tissue) into the lung (composed of tissue and fraction of air). These findings led to a better understanding of the ultrasonic properties of lungs as a function of inflation. As a result, it was suggested that ultrasound would be reflected from an inflated lung and less acoustic energy would be transferred into it.

The lung inflation studies led to the development of a model of the lung in terms of impedance (O'Brien *et al.*, 2002). The lung was modeled as a two-component acoustic structure of air and bulk material. The impedance of the lung depended on the volume fraction of air in the lung. As the lung became more inflated, the lung impedance would approach that of air; however, as the lung became more deflated, the lung impedance would approach that of tissue. Based on estimates of the fraction of air and lung at the deflated and inflated cases in the *in vivo* experiments [volume fraction of air ~0.76 for a deflated lung and ~0.80 for an inflated lung (Gil *et al.*, 1979)], the impedance estimated from the two-component model increases by 67% from an inflated to a deflated lung. Several other experiments have been conducted on ultrasound interaction with the lung that support the two-component bulk model of the lung (Dunn and Fry, 1961; Bauld and Schwan, 1974; Dunn, 1974, 1986; Pederson and Ozcan, 1986; Mikhak and Pederson, 2002). In these experiments, pressure reflection coefficients, attenuation values, and sound speed values were determined in lungs with varying levels of inflation. Table I summarizes the results of the lung reflection coefficient, attenuation, and sound speed values from these previous studies. The results indicate that the attenuation tends to be larger for higher air content in the lung. The sound speed tends to increase with decreasing air content while the density increases. The impedance tends to decrease with increasing air content, suggesting that more acoustic energy is transmitted into a lung with less air. Furthermore, the sound speed and attenuation both increase with increasing frequency at any one inflation level. In summary, the cumulative interpretation of results from these studies suggests that the amount of ultrasonic energy transmitting into and propagating within a deflated lung is much greater than in an inflated lung.

In our previous study, a technique was introduced to measure directly the reflection coefficient from the lung surface incident from water and to calculate the impedance value of intact fresh lungs, *ex vivo*, from rats for different volumes of inflation (Oelze *et al.*, 2003). In all of the previous studies on the acoustic properties of lungs, the reflection coefficient and impedance values were not estimated from intact fresh lungs. It was hypothesized that fresh lungs would have an impedance value dependent on the ultrasonic frequency and volume of inflation.

Measurements of the reflected ultrasound were made on seven rat lungs, *ex vivo*, under deflation (atmospheric pressure, 0 cm H_2O) and three volumes of inflation pressure (7, 10, and 15 cm H_2O) (Oelze *et al.*, 2003). From the measurements, the frequency-dependent pressure reflection coefficient for the lung was estimated, and the frequency-dependent impedance of the lung was then estimated from the pressure reflection coefficient. The estimates indicated

TABLE I. Summary of reflection coefficient, attenuation, and sound speed values for lung.

Air content (%)	Density (kg/m^3)	Sound speed (m/s)	Attenuation coefficient (Np/cm)	Impedance (Mrayl)
100[a]				0.17[a,b]
67[c,d]	350–360[c,d]	875.0,[c] 583.0[d]	8.83[c]	0.31,[b,c] 0.21[d]
62.5[b,c]	400[c]	918.7[c]	7.40[c]	0.37[b,c]
60[a,d]	430[d]	640.0[d]	7.40[a,b]	0.28,[d] 0.80[a,b]
57.7[b,c]	450[c]	965.3[c]	6.83[c]	0.43[b,c]
53[b,c]	500[c]	1027[c]	5.83[c]	0.51[b,c]
49[d]	540[d]	733.0[d]		0.40[d]
41.6[a]				0.92[a,b]
40[d]	640[d]	842.0[d]		0.54[d]
34[b,c]	700[c]	1202[c]	3.13[c]	0.84[b,c]
32[d]	720[d]	948.0[d]		0.68[d]
16[a]			7.40[a,b]	1.40[a,b]

[a]Values obtained from study by Bauld and Schwan (1974) using a frequency range of 2.4–7.4 MHz.
[b]Values estimated from parameters given in the study.
[c]Values obtained from study by Dunn (1986) using a frequency range of 1–5 MHz.
[d]Values obtained from study by Pederson and Ozcan (1986) using a frequency range of 10–800 kHz.

that the greater the volume of the inflation, the larger the magnitude of the reflection coefficient and, correspondingly, the smaller the impedance value of the lung. Estimates also indicated that the impedance value of the lung was always less than the characteristic impedance of water at 37 °C (pressure release surfaces relative to water). The frequency-dependent lung impedance ranged from an average of 1 Mrayl in deflated lungs to 0.2 Mrayl for fully inflated lungs; deflated lungs had an impedance closer to water (1.52 Mrayl) than inflated lungs. Using the two-component model of the lung, as the level of inflation increases more air fills the lung, and the impedance value of the lung approaches that of air (O'Brien *et al.*, 2002). As a result of the increased impedance mismatch between the inflated lung (which approaches that for air) and the surrounding fluid or tissue, more ultrasonic energy is reflected from the lung surface and less is transmitted. Therefore, more ultrasonic energy was transferred into the deflated lungs than into the inflated lungs.

In the case of an incidental clinical exposure of the lungs by ultrasound through intervening tissues, the degree of lung inflation will determine the amount of ultrasonic energy transferred into the lungs. Most of the sound will be reflected from the lung surface with just a little inflation. With the deflated lung, the pressure reflection coefficient is small (~0.25 for the rat lungs), the impedance is closely matched to the surrounding tissue, and more acoustic energy is transmitted into the lung. Because the deflated lung was less dense than water, the pressure transmission coefficient was ~0.75. If increased ultrasonic energy transmitted into the lung causes an increase in the incidence of lung damage, then the increased ultrasonic energy transmitted into the deflated lung due to more closely matched impedance supports earlier observations of higher probability of lung lesions in rats with deflated lungs.

In the present study, the reflection coefficient and impedance of intact fresh lungs, *ex vivo*, were estimated at four different levels of inflation from three species of animals, i.e., rats, rabbits, and pigs, in populations with varying ages. Using the same experimental techniques and statistical approach to examine the threshold for ultrasound-induced lung damage across species, the thresholds for ultrasound-induced lung damage were found to be consistent across species (O'Brien *et al.*, 2006). Furthermore, the morphology of ultrasound-induced lesions was similar in all species and age groups studied. In studies of the age-dependent effect of ultrasound-induced lung damage, threshold values were found to have a dependence on age in pigs (O'Brien *et al.*, 2003b). If the mechanism of ultrasound-induced lung damage results from the fraction of ultrasonic energy transmitted into the lung, then the impedance values of the lung should also correlate to species and age-dependent threshold values.

The study was conducted to further elaborate the dependence of ultrasonic energy transfer into the lungs and the effects of different species and age on the transfer of ultrasonic energy into the lung. Corresponding lung damage thresholds for different species of animals at different ages were then correlated to their respective lung impedance values under deflated conditions. Section II briefly discusses the theory and methodology used to estimate the reflection coefficient and the corresponding impedance value of the lung at different levels of inflation. Section III presents the results of the experiments. Section IV discusses the results of the experiments, and the final section ends with some concluding remarks.

II. METHODOLOGY

A. Measurements

Estimates of the impedance of the *ex vivo* lung were obtained by measuring the pressure reflected from the lung surface. The reflected ultrasound would come from the interface between the lung surface and the surrounding water. Figure 1 shows a diagram of the interface between the lung and water. The acoustic pressure reflection coefficient (for plane waves) of sound reflecting at normal incidence from a planar boundary of two different media (media I and II) with a frequency, f, is defined as (Kinsler *et al.*, 2000)

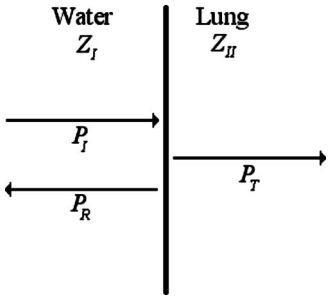


FIG. 1. Diagram representing the interface between water with impedance Z_{I} and the lung with impedance Z_{II} and corresponding incident (P_{I}), reflected (P_R), and transmitted (P_T) pressure waves.

$$R(f) = \frac{P_R(f)}{P_{\mathrm{I}}(f)}, \tag{1}$$

where $P_R(f)$ is the pressure amplitude of the reflected sound and $P_{\mathrm{I}}(f)$ is the amplitude of the incident pressure wave. The sign of the reflection coefficient is positive if the impedance of medium I is smaller than the impedance of medium II and is negative if the impedance of medium I is greater than that of medium II. Actual estimates of the pressure reflection coefficient were obtained through a planar reference technique outlined in a previous study (Oelze *et al.*, 2003).

The pressure reflection coefficient is related to the impedance of the two media by (Kinsler *et al.*, 2000),

$$R(f) = \frac{Z_{\mathrm{II}}(f) - Z_{\mathrm{I}}(f)}{Z_{\mathrm{II}}(f) + Z_{\mathrm{I}}(f)}, \tag{2}$$

where $Z_{\mathrm{I}}(f)$ and $Z_{\mathrm{II}}(f)$ represent the frequency-dependent impedance of media I and II, respectively. Solving Eq. (2) for the impedance of medium II yields

$$Z_{\mathrm{II}}(f) = Z_{\mathrm{I}}(f)\frac{1 + R(f)}{1 - R(f)}. \tag{3}$$

If the impedance of medium I is known (i.e., water) and the pressure reflection coefficient at the boundary of media I and II is determined, then the impedance of medium II can be estimated by Eq. (3).

A single-element weakly focused ultrasonic transducer (GE Panametrics Inc., Waltham, MA) was used to make measurements of the pressure reflection coefficient from the planar reference material (a smooth Plexiglas™ plate) and the lung tissues. The transducer had an aperture diameter of 19 mm and a focal length of 51 mm measured from the planar reference material. The center frequency of the transducer was 7.1 MHz with a −6 dB pulse-echo frequency bandwidth of 5.5 MHz and a −6 dB pulse-echo beamwidth of ~550 μm. The analysis bandwidth ranged from 3.5 to 10 MHz (−9 dB pulse-echo bandwidth).

The transducer was operated in a pulse-echo mode by using a Panametrics 5800 pulser/receiver (GE Panametrics, Inc., Waltham, MA). The incident amplitude of the pulse used for the reflection coefficient study was determined to be 1.4 MPa at the focus of the transducer with a calibrated Polyvinylidene Fluoride (PVDF) membrane hydrophone (Marconi Model Y-34-6543, Chelmsford, UK). Measurements were digitized on an oscilloscope (LeCroy 9354TM, Chestnut Ridge, NY) that had a dynamic range of 48 dB and downloaded to a PC for postprocessing. The sampling rate was 50 MHz. Reflection measurements from both the planar reference and the lung surface were taken near the focus of the transducer where the wave fronts could be approximated as a plane wave within the narrow beamwidth (Lizzi *et al.*, 1983).

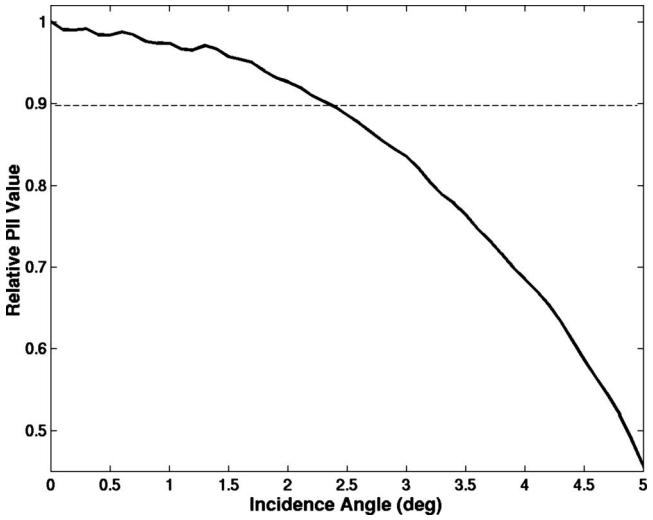


FIG. 2. Relative loss in magnitude of the PII value of the reflected signal as the angle of incidence moves away from normal.

While the planar reflector is a flat surface, the lung is not a flat surface. One of the difficulties in the experimental setup was in aligning the lung with the transducer so that the sound reflected from the lung was at normal incidence. Failure to align the lung properly or a local curvature of the lung surface would lead to a reduced reflected signal relative to sound reflected at normal incidence. To quantify the possible alignment error, the plastic plate was used to record the relative reflected signal at several angles of incidence. A smooth plastic plate was placed normally incident to the transducer at the focus and rotated axially by several degrees. The plastic plate was rotated about the z-axis, and the magnitude of the signal decrease was then recorded. Figure 2 shows the relative pulse intensity integral (PII) of the reflected signal as the angle of incidence changes from normal. The angle of incidence could be off normal by 2.5° and still have a PII value within 10% error of true normal.

The lung surface was measured similarly to the reflection from the plastic plate. In order to obtain a smooth flat surface on the lung that would be near-normal incidence, the lung was gently pressed up against a plastic holder that was aligned perpendicular to the beam axis. The holder had a small circle cut out in the middle with a diameter of 18 mm that allowed for a scanning window onto the lung surface; the lung surface was in direct contact with the water. The large size of the scanning window (18 mm) compared to the beamwidth (~550 μm) ensured negligible diffraction effects. The lung was setup so that its surface would be near the focus of the transducer. A 5×5 mm^2 square perpendicular to the beam axis of the transducer (y-z plane) was scanned on the surface of the lung. RF data waveforms were acquired at intervals of 100 μm in the y and z directions.

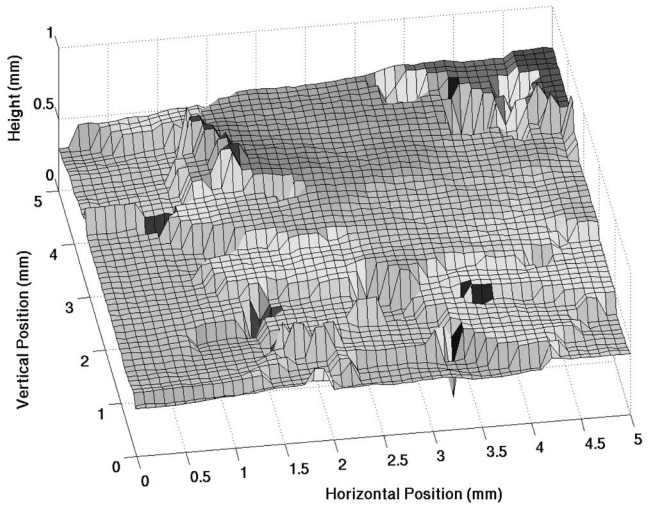


FIG. 3. Surface plot of a deflated rabbit lung recorded from the maximum peak magnitude of waveforms scanned over a 5×5 mm^2 window.

In order to reduce the measurement error of reflections from off-normal incidence, waveform sections were used in the calculations that coincided with flat areas at near-normal incidence on the lung surface. A surface plot of the lung over the 5×5 mm^2 scanned area was constructed from waveforms measured off the lung surface. The surface plot was constructed by finding the distance from the source to the peak with a maximum magnitude in each reflected waveform. Figure 3 shows a graph of a surface plot from one of the lungs. The flat regions shown by the surface plots were used to determine smooth regions at near-normal incidence from which the reflection coefficient calculations were made. The angle of incidence at a particular point in the scan region was calculated from the average of the arctangents of the slopes between the particular scan point and four immediately surrounding scan points. Only reflected pulses from scan points that showed angles of incidence less than 0.5° from normal were used. This procedure does not account for surface roughness smaller than a beamwidth, which could result in a frequency-dependent loss of some reflected signal relative to the reference signal. However, because the size of the roughness would be small compared to the beamwidth (and wavelength), the contribution from the roughness was assumed to be negligible compared to the direct reflection from the lung surface.

Reference pulses were measured from a smooth plastic plate of known reflectivity. The transducer was aligned perpendicular to the reflecting surface. The same equipment and settings were used for the pulse/echo measurement of the reference pulse as used for the lung reflection measurements. Reflection measurements from the planar surfaces were made at different depths relative to the transducer by scanning axially along the transducer beam. The scan measured reflected values over an axial length of 1 cm centered about the focus with a step size of 100 μm (approximately half the wavelength).

B. Animal procedures

The experimental protocol was approved by the Institutional Animal Care and Use Committee at the University of Illinois at Urbana-Champaign and satisfied all campus and National Institutes of Health (NIH) rules for the humane use of laboratory animals. Animals were housed in an Association for Assessment and Accreditation of Laboratory Animal Care, Rockville, MD (AAALAC)-approved animal facility. Water was provided *ad libitum* for all species. Rabbits and pigs were fed a measured amount, and rats were provided food *ad libitum*.

Age-dependent lung impedances (reflection coefficients) were determined in four 2-week-old, four 3-week-old, four 7-week-old, seven 17-week-old, and five 52-week-old Sprague Dawley rats (Harlan, Indianapolis, IN); seven 12-week-old, three 3-yr-old, and three 5-yr-old New Zealand White rabbits (Myrtle's Rabbitry, Thompson Station, TN); and seven 1-week-old, seven 4- to 5-week-old, and seven 7–8-week-old crossbred pigs (University of Illinois Veterinary Research Farm, Urbana, IL). Animals were weighed and anesthetized. Rats and rabbits were anesthetized with ketamine hydrochloride (87.0 mg/kg, rats; 50.0 mg/kg, rabbits) and xylazine (13 mg/kg, rats; 10 mg/kg, rabbits) administered intraperitoneally in rats and subcutaneously in rabbits. Pigs were anesthetized with ketamine hydrochloride (2.2 mg/kg), xylazine (2.2 mg/kg), and Telazol (4.4 mg/kg) administered intramuscularly. The animals were immediately euthanized by cervical dislocation (rats) or carbon dioxide (rabbits and pigs) while under anesthesia. The lungs (with heart attached) and trachea (cut transversely at the thoracic inlet) were removed by opening the thorax with a midline incision extending from the thoracic inlet to the caudal most aspect of the sternum. Following the removal, the lungs were rinsed with a 0.9% (isotonic) saline solution. Saline was not allowed to enter the trachea. All lungs were used immediately after harvest.

A 20-gauge needle (2- and 3-week-old rats), a 16-gauge needle (7-week, 17-week, and 52-week-old rats), a 0.25-in. diameter stainless steel tube (1-week-old pigs and rabbits), and a 0.50-in.-diameter stainless steel tube (4–5- and 7–8-week-old pigs), with the tips blunted, were inserted into the open end of the trachea and sealed airtight. Plastic hosing was then used to attach the needle to the lung inflation apparatus. The lung inflation apparatus consisted of a plastic tube connected to two pressure regulators, a tank of compressed air, and a ruler to determine the pressure level (for a diagram of the apparatus, see Fig. 4). Pressure in the *ex vivo* lungs was measured in terms of centimeters of H_2O (1 cm H_2O=98 Pa). When a pressure level of 30 cm H_2O in an *ex vivo* lung was reached, the lung was considered fully inflated or inflated to TLC (Bachofen *et al.*, 1982). In terms of common lung parameters (i.e., the TLC), 15-, 10-, 7-, 5-, and 0 cm H_2O were estimated to correspond to 90%, 82%, 70%, 61%, and 7% TLCs, respectively (Gil *et al.*, 1979).

All measurements were made in degassed water at 37 ± 0.5 °C. Water temperature was controlled by a proportional temperature controller (Yellow Springs Instrument Co., Inc., Yellow Springs, OH). A digital thermometer (Control Company, Friendswood, TX), with an accuracy of 0.2 °C, in the water tank served as the temperature standard.

Measurements for the pressure reflection coefficient were taken at four different volumes of inflation for each

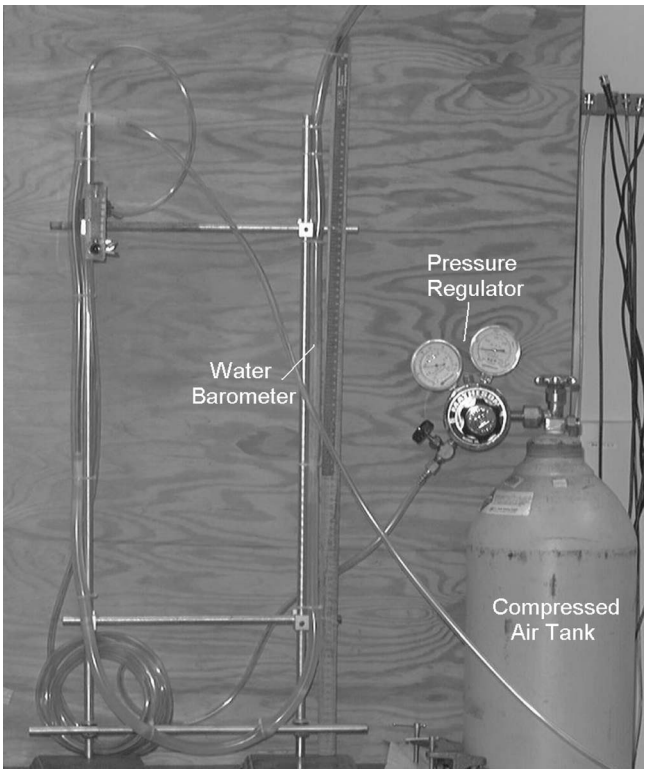


FIG. 4. Photograph of the inflation apparatus including a compressed air tank, pressure gauge regulators, and a water barometer.

lung. The waveforms reflected from the lung surface were normal to the axis of the transducer. The first measurements were taken on the deflated lungs (0 cm H_2O). After the deflated measurement was taken, the lung was slowly inflated at a rate of 1 cm H_2O/min, to 20 cm H_2O. The lung was held at that pressure for a minimum of 5 min to allow the pressure and air inside the lung to equilibrate. The pressure was then reduced to 15 cm H_2O at 2 cm H_2O/min and held for 15 min to ensure pressure equilibrium. Measurements were then taken from the inflated lung at 15 cm H_2O. For a normal animal, 15–20 cm H_2O pressure is required for adequate but not full inflation (Hayatdavoudi *et al.*, 1980). The next two pressure volumes (7 and 10 cm H_2O) were used to observe lung impedance at intermediate levels of inflation. The pressure was reduced to 10 cm H_2O and 7 cm H_2O (5 cm H_2O for the pigs) at 2 cm H_2O/min, and held for 15 min at each pressure level to ensure pressure equilibrium. Measurements were then taken from the inflated lung at 10 and 7 cm H_2O (5 cm H_2O for the pigs). 7 cm H_2O (5 cm H_2O for the pigs) was the lowest pressure that could be maintained without the lung visibly deflating during the measurement procedure.

C. Data analysis

The acquired echo waveforms were downloaded to a computer for postprocessing. MATLAB® (The Mathworks Inc., Natick, MA) was used to perform the waveform analysis. The frequency spectrum of each waveform reflected from the lung was calculated by taking its Fourier transform. Each reflection waveform was made up of 2000 points. A Hanning window was constructed based on the width of the pulse reflected from the reference surface. The Hanning window was applied to each pulse reflected from the lung by centering the window at the point in the waveform where the peak had its maximum magnitude. To account for the spectral influence of the Hanning window, the reference pulse was also multiplied by the same Hanning window. A 5 $\times 5$ mm^2 section from each lung was scanned at a 100 μm step size, yielding many waveforms per lung examined. The absolute values of the Fourier transforms from the waveforms were taken, and all the Fourier spectra were averaged for each lung to obtain an overall representation of the pressure reflection coefficient from the lung. The exact number of waveforms acquired per inflation level depended on the amount of surface area of the lung that was flat and perpendicular to the transducer axis. Therefore, while the minimum number of waveforms from each animal was greater than 10, the number of waveforms acquired for each animal was different. The impedance was then calculated over a frequency range of 3.5–10 MHz according to Eq. (3), where $Z_I(f)$ represented the acoustic impedance of water at 37 °C [$Z_I(f)$ is the product of the density of water, 998 kg m^{-3}, and the speed of sound in water at 37 °C, 1525 m s^{-1}].

III. RESULTS

Estimates of the pressure reflection coefficient from *ex vivo* lungs of different species of animals at various ages for four levels of inflation were acquired with the ultrasonic system. The impedance values of the lung for the four levels of inflation were calculated from the pressure reflection coefficients. Because the TLC and functional residual capacity of the lung scale approximately linearly with animal weight across species (Stahl, 1967), estimates of the fraction of air in the lung from previous studies of rabbit lung in terms of TLC versus pressure (cm H_2O) were used to predict the impedance at each pressure level (Gil *et al.*, 1979). Based on the study by Gil *et al.* (1979), the fraction of air corresponding to 0, 5, 7, 10, and 15 cm H_2O in the lung is 0.39, 0.84, 0.85, 0.86, and 0.89, respectively. Using the two-component model of the lung, this leads to predictions of 0.26, 0.013, 0.011, 0.099, and 0.0062 Mrayl for the acoustic impedance of the lung at the different inflation levels, respectively (O'Brien *et al.*, 2002).

A. Rats

Figure 5 shows the absolute value of pressure reflection coefficients, $|R(f)|$, from rat lungs at the four levels of inflation as a function of frequency. For the deflated lungs for all ages of rats, $|R(f)|$ ranged between 0.1 and 0.3 across all frequencies examined, indicating that a large percent of the ultrasonic energy was transmitted into the lung. $|R(f)|$ at higher levels of inflation suggests an age-dependent effect. $|R(f)|$ in the inflated lung is greater in the older rats. In the case of the 2-week-old rats, no differences are observed between $|R(f)|$ at deflation and various levels of inflation. Furthermore, minimal changes in $|R(f)|$ versus inflation level are observed for the 3-week-old rats. Based on our hypothesis, this suggests that the young rats would have a lower thresh-

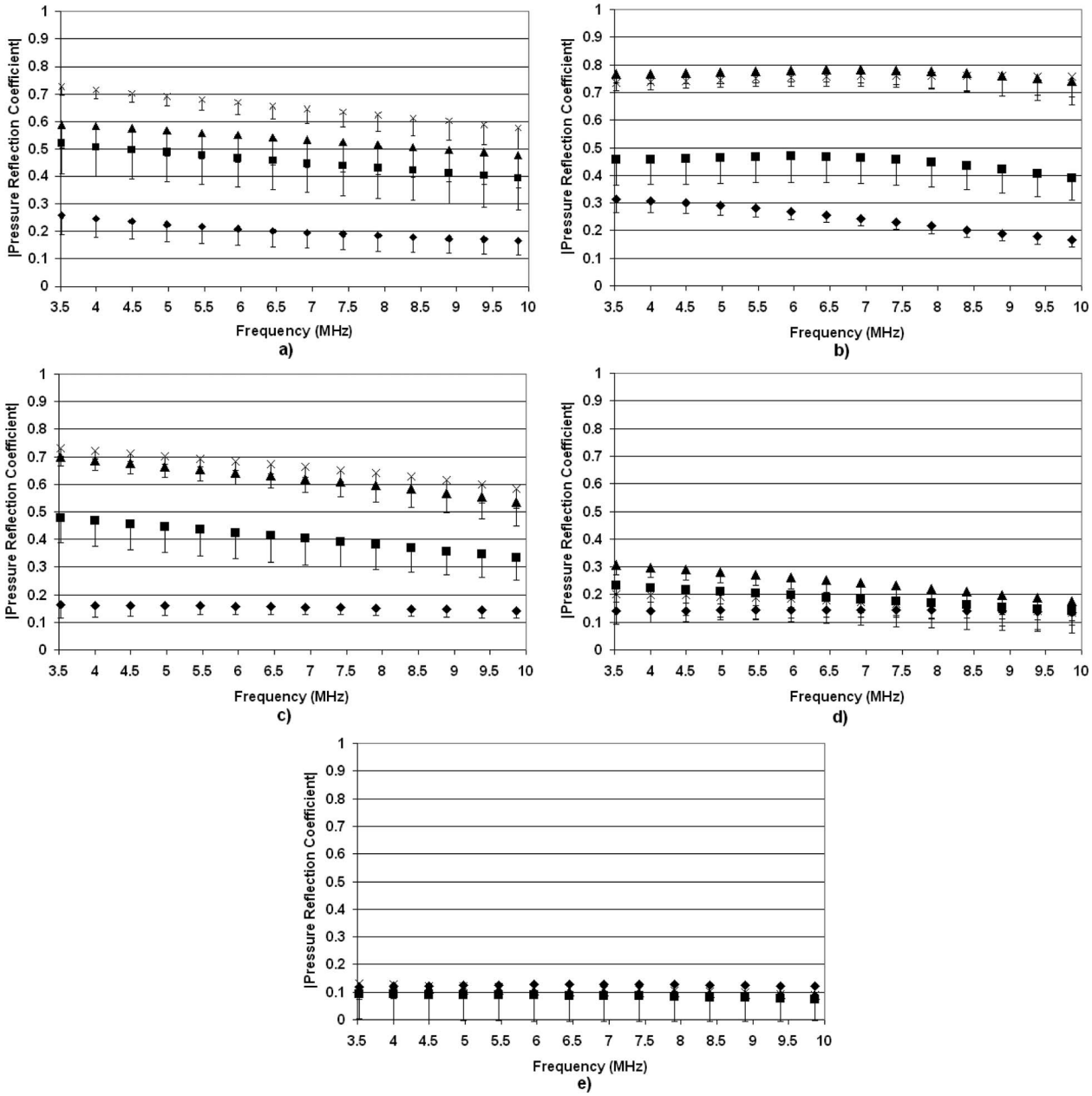


FIG. 5. Pressure reflection coefficients vs frequency for deflated rat lungs (♦) and inflated lungs at 7 cm H_2O (■), 10 cm H_2O (▲), and 15 cm H_2O (×). The measurements were taken from (a) 52-week-old rats, (b) 17-week-old rats, (c) 7-week-old rats, (d) 3-week-old rats, and (e) 2-week-old rats. The error bars represent one standard error.

old for ultrasound-induced lung damage because no matter the inflation level, a larger percentage of ultrasonic energy would be transmitted into the lung.

The impedance values for the deflated and fully inflated (15 cm H_2O) rat lungs reveal similar trends (Fig. 6). The impedance values for the deflated lungs range between 0.8 and 1.25 Mrayl across all age ranges. However, in the case of the inflated lungs, the 2- and 3-week-old rats have impedance values larger than 0.7. Note also that the 2-week-old rats have essentially the same impedance values for both inflation levels. The 17- and 52-week-old rats have the lowest impedance values for the deflated and inflated cases. Based on the hypothesis that lung damage is caused by the transfer of ultrasonic energy into the lung, this would suggest that the threshold for lung damage would be lower for less developed (younger) rats. Finally, the predicted impedance values based on extrapolating measurements and estimates from Gil *et al.* (1979) were consistently smaller than estimates of impedance derived from reflection coefficient measurements from ex vivo lungs at all levels of inflation.

B. Rabbits

Figure 7 shows $|R(f)|$ versus ultrasonic frequency for the rabbit lungs at different stages of development. Across all ages of rabbits, $|R(f)|$ is consistent with the level of inflation and $|R(f)|$ increases with increasing level of inflation. This would suggest that an age-dependent threshold is not ob-

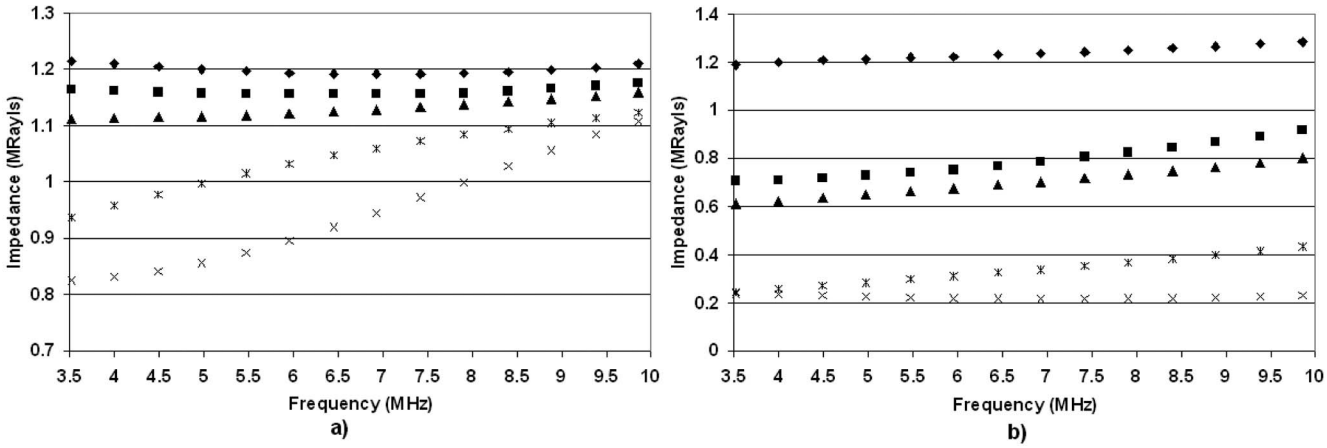


FIG. 6. Estimates of the rat lung impedance vs frequency for (a) deflated lungs and (b) fully inflated lungs from 52-week-old rats (×), 17-week-old rats (*)s, 7-week-old rats (▲), 3-week-old rats (■), and 2-week-old rats (♦).

served for the rabbits (at least spanning the ages explored in these experiments). The only biological differences observed between the rabbits and rats were the smaller thickness of the pleural surface for the rats. Therefore, we hypothesize that the thickness of the pleural membrane, which will increase with increasing maturity of the animal, might be responsible for the differences observed between the rats and rabbits. However, at this time the hypothesis has not been tested.

The impedance values (Fig. 8) reveal that for all ages of rabbits, the deflated lungs have impedances typically greater than 1.0 Mrayl and close to that of tissue or water (1.4–1.6 Mrayl). The close impedance match between the deflated lung and tissue suggests that a large percentage of incident ultrasonic energy would be transferred into the lung. The impedance values for the deflated lungs also increase with frequency by as much as 15% from 3.5 to 10 MHz. The

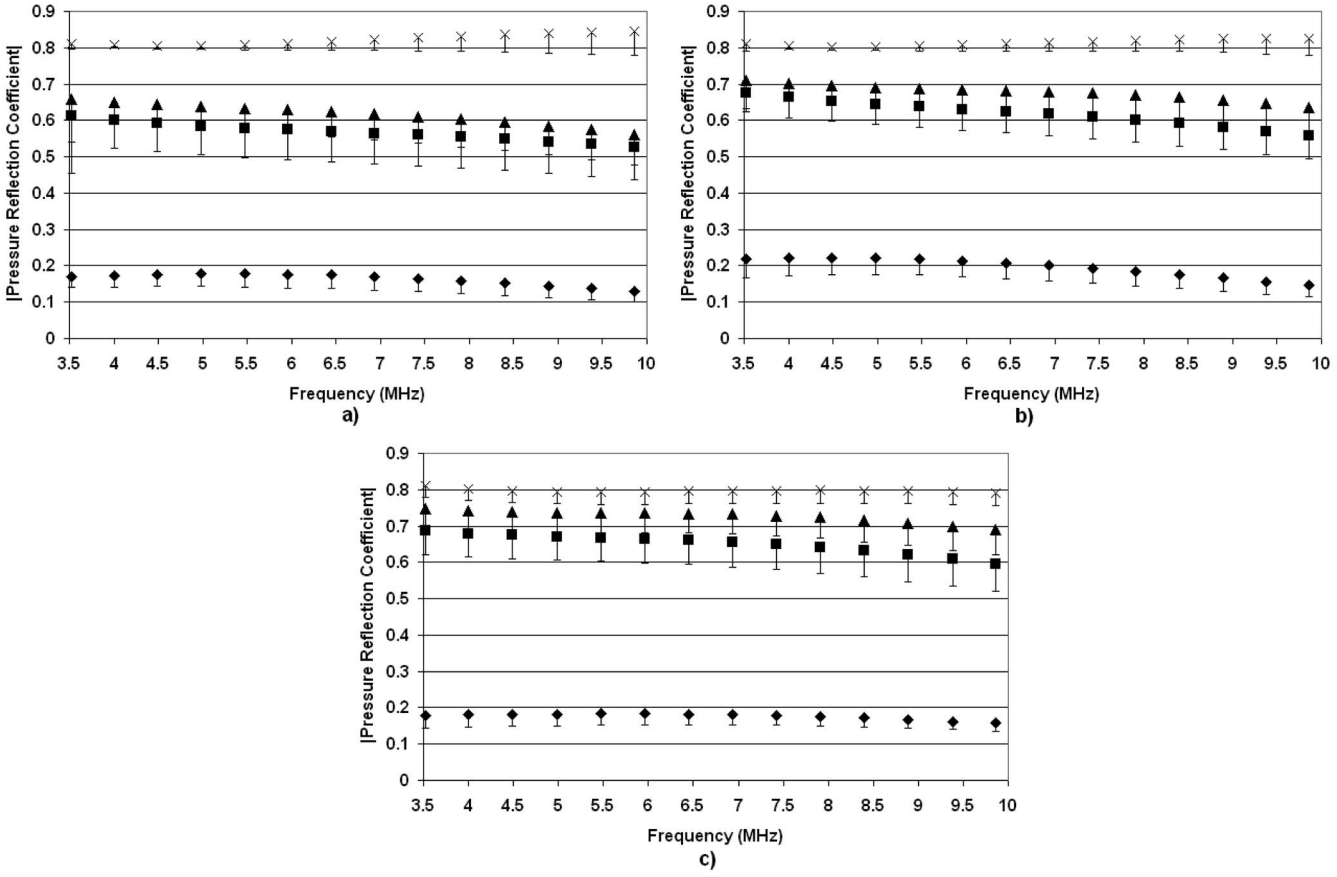


FIG. 7. Pressure reflection coefficients vs frequency for deflated rabbit lungs (♦) and inflated lungs at 7 cm H_2O (■), 10 cm H_2O (▲), and 15 cm H_2O (×). The measurements were taken from (a) 5-yr-old rabbits, (b) 3-yr-old rabbits, and (c) 12-week-old rabbits. The error bars represent one standard error.

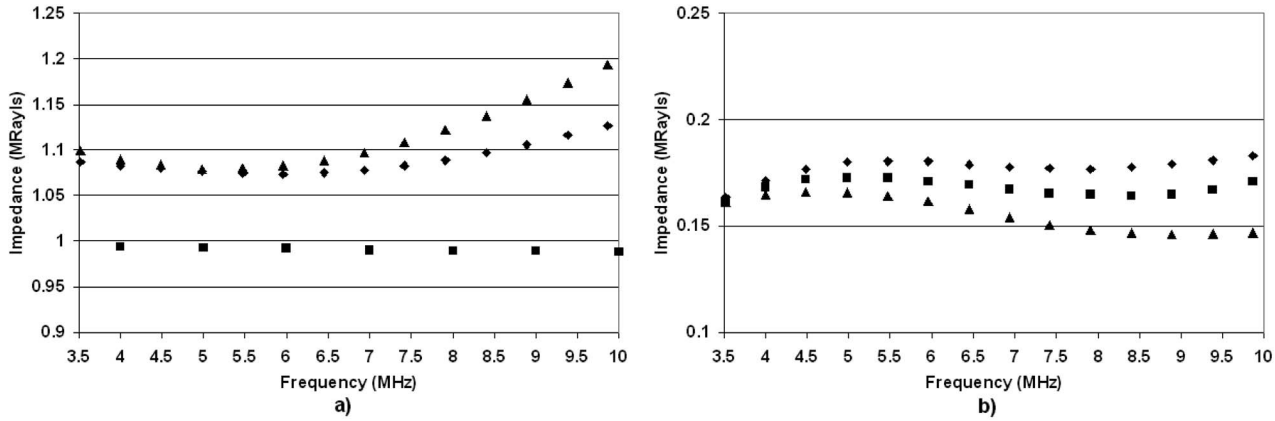


FIG. 8. Estimates of the rabbit lung impedance vs frequency for (a) deflated lungs and (b) fully inflated lungs from 5-yr-old rabbits (▲), 3-yr-old rabbits (■), and 12-week-old rabbits (◆).

impedance values for the fully inflated (15 cm H_2O) lungs range from 0.15 to 0.18 Mrayl, or almost an order of magnitude smaller than the impedance of tissue. The frequency dependence of the inflated rabbit lungs does not reveal a significant increase in impedance values with higher ultrasonic frequency. The large impedance mismatches between the inflated lungs and tissue would suggest little transfer of ultrasonic energy into the lungs. Furthermore, as in the case of the rats, the predicted impedance values based on extrapolating measurements and estimates from Gil *et al.* (1979) were consistently smaller than estimates of impedance derived from reflection coefficient measurements from *ex vivo* lungs at all levels of inflation.

C. Pigs

The $|R(f)|$ versus ultrasonic frequency for the pigs is shown in Fig. 9. Results for the pigs are similar to those of the rabbit in that $|R(f)|$ is observed to increase with increasing inflation, and similar trends were observed for each age group of pigs examined. Like the rabbits, the largest increase in the pigs $|R(f)|$ is observed from deflated to just partially

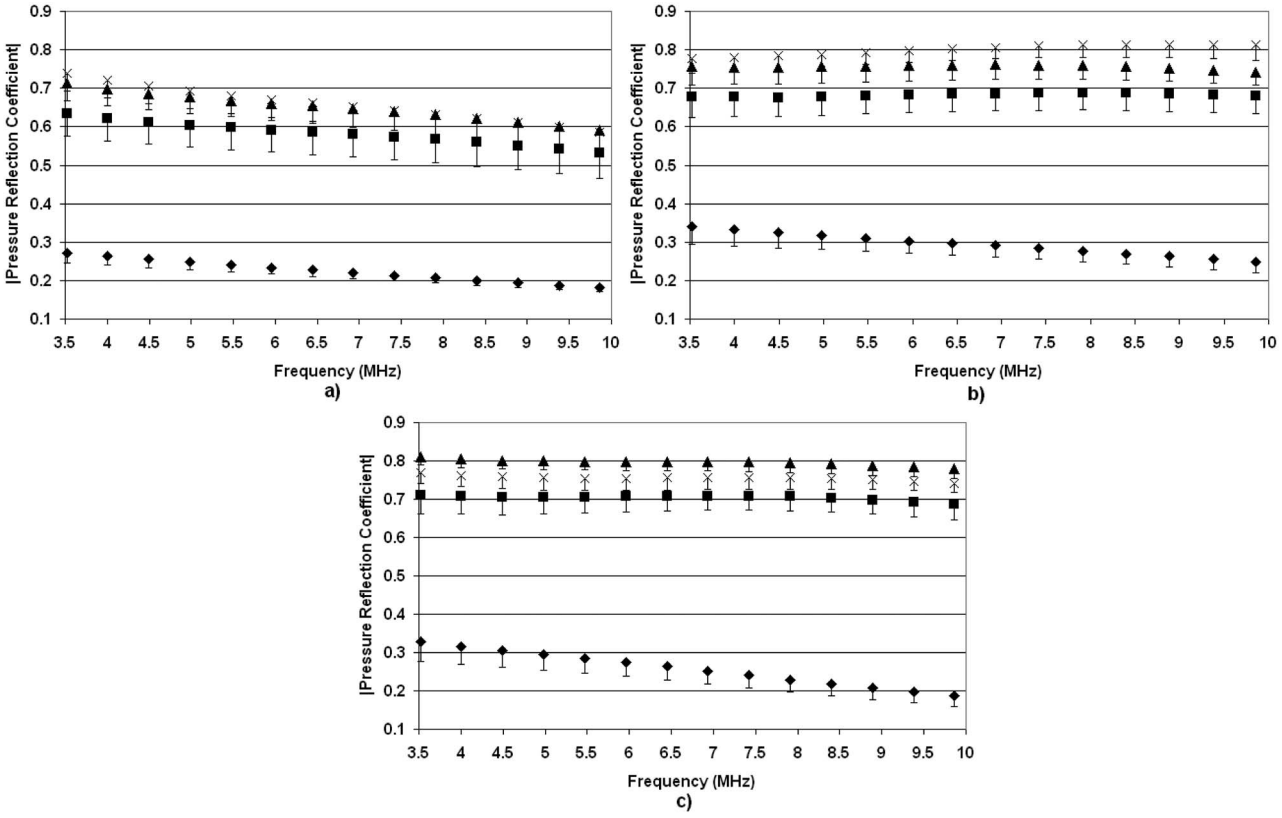


FIG. 9. Pressure reflection coefficients vs frequency for deflated pig lungs (◆) and inflated lungs at 5 cm H_2O (■), 10 cm H_2O (▲), and 15 cm H_2O (×). The measurements were taken from (a) 7–8-week-old pigs, (b) 4–5-week-old pigs, and (c) 1-week-old pigs. The error bars represent one standard error.

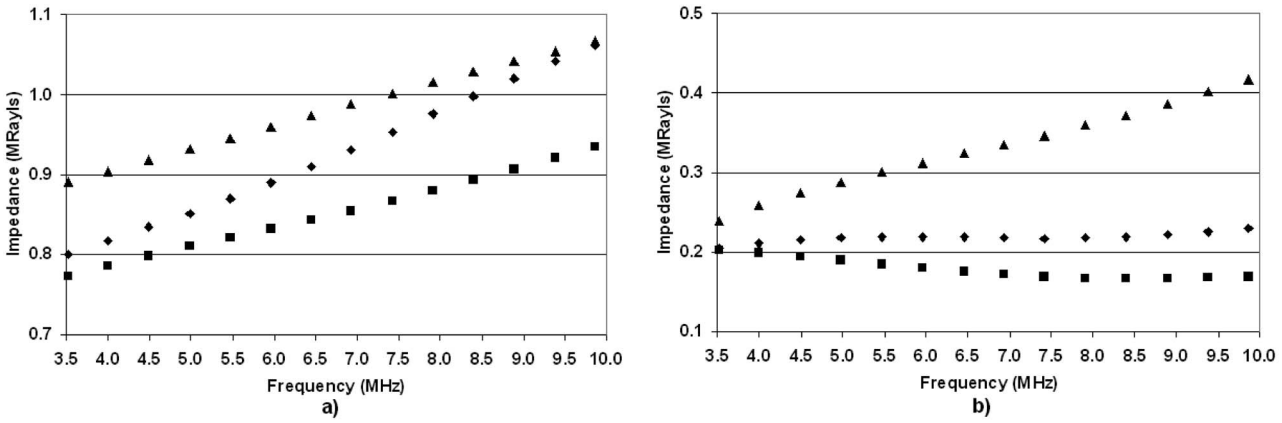


FIG. 10. Estimates of the pig lung impedance vs frequency for (a) deflated lungs and (b) fully inflated lungs from 7–8-week-old pigs (▲), 4–5-week-old pigs (■), and 1-week-old pigs (▲).

inflated (5 cm H_2O) lungs. Similar to the rabbits, the thickness of the pleura for the pigs will be larger than for the rats and may explain the differences observed between the lungs of the rats and the lungs of the pigs and rabbits.

The impedance values for the pig lungs are shown in Fig. 10. Again, the impedance values for all ages in the deflated case are always within 20% of each other. However, compared with the rabbit lung, deflated pig lungs exhibit greater frequency dependence. The impedance values increase by more than 20% from 3.5 to 10 MHz with the pigs, whereas with the rabbits, the increase in impedance values over the same range is at most 15%. The rat impedance values reveal similar trends of increasing impedance with frequency in the deflated lungs for the older rat populations. In the case of the fully inflated (15 cm H_2O) lungs, the impedance values for the pigs have very small slopes except for the oldest pigs, where the impedance value increases by almost 77% from 3.5 to 10 MHz. In terms of lung impedance and age, the impedance is greatest for the oldest pigs, least for the 4–5-week-old pigs, and in between for the youngest pigs. Furthermore, as in the case of the rats and rabbits, the predicted impedance values based on extrapolating measurements and estimates from Gil *et al.* (1979) were consistently smaller than estimates of impedance derived from reflection coefficient measurements from *ex vivo* lungs at all levels of inflation.

IV. DISCUSSION

In most cases examined, $|R(f)|$ increased with increasing level of inflation. The increase in $|R(f)|$ corresponded to a decrease in the impedance of the lung with increasing inflation level. These results corroborate earlier results from a study on rat lungs alone (Oelze *et al.*, 2003). The results further support the acoustic model of the lung as a two-component structure whose impedance is determined by a combination of the bulk properties of air and tissue (O'Brien *et al.*, 2002). The higher the level of inflation of the lung, the more the lung impedance approaches that of air and subsequently the smaller the fraction of ultrasonic energy transmitted into the lung.

However, while the two-component model of the lung predicted an increase in the acoustic impedance with decreasing inflation levels, the actual predicted values of acoustic impedance were much smaller than impedance values estimated from the reflection coefficient measurements. The largest impedance value predicted by the two-component model when the fraction of air is zero is 0.77 Mrayl. Compared to the acoustic impedance values estimated for the deflated lungs, whose fraction of air is estimated to be approximately 0.39, all estimates were above 0.77 Mrayl (with the exception of the 4–5-week old pigs), while the predicted value from the model was 0.26 Mrayl. This discrepancy between predicted acoustic impedance values and acoustic impedance values estimated for the deflated and inflated lungs suggests that the two-component model of the lung, while conceptually sound, needs improvement to better predict the acoustic impedance versus the level of lung inflation.

The only exceptions to the observations that the acoustic impedance values decreased with increasing level of inflation were found in the measurements from the 2- and 3-week-old rats. The impedance values did not appear to decrease as much with increasing level of inflation, suggesting that with 2- and 3-week-old rats the threshold of lung damage could be less than for 7-, 17-, and 52-week-old rats. The thickness of the pleura was smallest for the youngest rats, which has been hypothesized to play a role in the threshold value for lung damage. Therefore, the difference in the pleural thickness with the age of rats may play a role in the age dependence observed in the rats.

One way to test this hypothesis is to examine the frequency-dependent impedance from the lung surface. The frequency dependence of the impedance may yield some clues to the underlying mechanism of the age-dependent effect. In the case of the deflated lungs, the impedance values were typically observed to increase with increasing frequency, the exception being the youngest rats. One hypothesis explaining this effect is the presence of the pleural surface surrounding the freshly excised lungs, a mechanism similar to that used to explain the frequency dependence of ultrasonic thresholds in a mammalian brain (Johnston and Dunn, 1976). In rats and rabbits, the thickness of the pleura

TABLE II. Comparison between the sound power transmission coefficient (SPTC) calculated from the deflated (0 cm H_2O) lung impedance values (Z_{def}) determined herein at the indicated frequency f using an incident (water) impedance value of 1.5 Mrayl and the comparable literature-based ultrasound-induced lung damage ED_{05} occurrence threshold ($ED_{05\ occ}$). Selected rat Z_{def} values were interpolated (Int) because the literature-based rat ages were slightly different from those evaluated herein. Also, the Z_{def} values at frequencies less than 3.5 MHz were extrapolated to either 2.8 or 3.1 MHz.

From impedance study			From threshold studies			
	Z_{def} (Mrayl)	SPTC	⟨age⟩	ED_{05} occ (MPa)	f (MHz)	Reference
1 week pigs	0.79	0.904	5 d	3.60	3.1	O'Brien *et al.*, 2003b
4–5 week pigs	0.76	0.893	39 d	5.83	3.1	O'Brien *et al.*, 2003b
7–8 week Pigs	0.88	0.932	58 d	2.87	3.1	O'Brien *et al.*, 2003b
2 week rats	1.22	0.989	12–14 d	2.44	2.8	O'Brien *et al.*, 2008
2 week rats	1.20				5.6	
3 week rats	1.17	0.985	22–24 d	3.00	2.8	O'Brien *et al.*, 2008
3 week rats	1.16				5.6	
7 week rats	1.11	0.978	53–62 d	3.24	2.8	O'Brien *et al.*, 2008
7 week rats	1.12				5.6	
Int (rats)	1.00	0.960	72 d	3.38	2.8	O'Brien *et al.*, 2008
Int (rats)	1.00	0.960	10–11 week	2.30	2.8	Zachary *et al.*, 2001a, 2001b
Int (rats)	1.05	0.969	10–11 week	2.80	5.6	Zachary *et al.*, 2001a, 2001b
Int (rats)	1.00	0.960	10–11 week	3.60	2.8	O'Brien *et al.*, 2001a, 2001b
Int (rats)	1.05	0.969	10–11 week	3.40	5.6	O'Brien *et al.*, 2001a, 2001b
Int (rats)	1.00	0.960	10–11 week	3.13	2.8	O'Brien *et al.*, 2003a
17 week rats	0.81				2.8	
17 week Rats	0.88				5.6	
52 week rats	0.91				2.8	
52 week rats	1.02				5.6	
12 week rabbits	1.07	0.972	76 d	3.43	5.6	O'Brien *et al.*, 2006
3 year rabbits	1.00				5.6	
5 year rabbits	1.08				5.6	

is 8–25 μm, and for pigs and humans this thickness is 100–150 μm (AIUM, 2000; Tyler and Julian, 1992). The thickness of the pleura will increase with age from young to adult animals. Mice, rats, rabbits, cats, dogs, and monkeys, which are members of the thin group, have scant but varying quantities of collagen fibers within the pleura. Sheep, pigs, humans, cattle, horses, and other large animals, which are members of the thick group, have abundant but varying quantities of collagen fibers within the pleura (AIUM, 2000; Tyler and Julian, 1992). A thin layer surrounding the lung will act as a filter for ultrasonic energy, allowing ultrasound at certain frequencies to pass through more efficiently than other frequencies. A simple equation for the pressure reflection coefficient of ultrasound through a thin layer is given by (Kinsler *et al.*, 2000)

$$R(f) = \frac{(1 - Z_1/Z_3)\cos k_2 L + j(Z_2/Z_3 - Z_1/Z_2)\sin k_2 L}{(1 + Z_1/Z_3)\cos k_2 L + j(Z_2/Z_3 + Z_1/Z_2)\sin k_2 L}, \quad (4)$$

where k_2 refers to the acoustic wave number in the layer, L is the layer thickness, and the subscripts 1, 2, and 3 refer to the front medium, layer, and back medium, respectively. From Eq. (4), the frequency-dependent impedance can be calculated from Eq. (3).

Another hypothesis that might explain these differences is roughness of the pleural surface. Roughness will cause some of the ultrasonic energy incident on the surface to scatter in many directions. Therefore, the backscattered energy will be less if the surface was smooth (Sagar *et al.*, 1978). In addition, the backscatter from a rough surface will be frequency dependent (Oelze *et al.*, 2001). Future studies will be required to thoroughly test both hypotheses.

Finally, literature-based thresholds of ultrasound-induced lung damage were related to the impedance values determined herein from the different species of animals examined (Table II). Assuming that the surrounding material (water or tissue) had an impedance of 1.5 Mrayl, the normal-incidence sound power transmission coefficient (SPTC) was calculated (Kinsler *et al.*, 2000) for each lung (using the mean lung impedance values). Furthermore, the SPTC was calculated for each of the animals for which ED_{05} occurrence thresholds were assessed. Figure 11 shows the ED_{05} occurrence threshold versus SPTC. For the deflated lung, there is a general trend whereby as the SPTC increases, the ED_{05} threshold decreases. For the fully inflated (plot not shown), no clear trend was observed. These observations support the view that the ultrasonic energy transmitted into the lung (deflated lung received the most ultrasound energy) determines the lung hemorrhage threshold.

Furthermore, as noted earlier, for the pigs the impedance was greatest for the oldest pigs, least for the 4–5-week-old pigs, and in between for the youngest (neonate) pigs. This suggests that the oldest pigs would have the lowest threshold

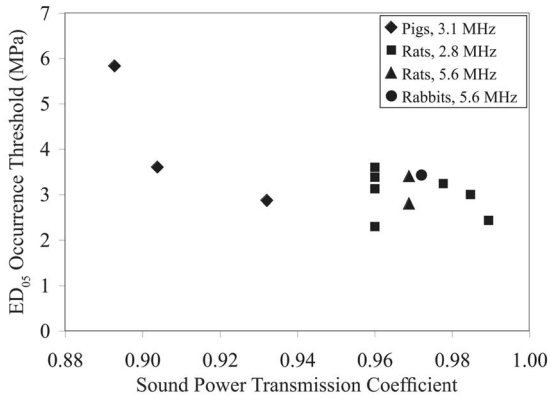


FIG. 11. ED_{05} occurrence threshold vs SPTC under deflated lung conditions.

for lung damage (i.e., lung damage occurs at lower pressure values), and the 4–5-week-old pigs would have the highest threshold for lung damage. A comparison of these impedance values with ultrasound-induced lung damage thresholds in pigs reveals the same trend (O'Brien *et al.*, 2003b). The oldest pigs had the lowest threshold for damage (2.87 ± 0.29 MPa), followed by the neonate pigs (3.60 ± 0.44 MPa), followed lastly by the 4–5-week old pigs (5.83 ± 0.52 MPa).

V. CONCLUSION

The pressure reflection coefficient was determined, and the impedance values versus frequency were estimated from freshly excised lungs at different levels of inflation from rats, rabbits, and pigs of various ages. The data revealed consistent trends whereby lungs that were deflated had smaller pressure reflection coefficient values than inflated lungs. The only exceptions were with the rats, where rats younger than 3 weeks old were observed to have minimal increases in pressure reflection coefficients versus level of inflation. These observed trends corresponded to earlier experiments that revealed that ultrasound-induced lung damage was related to the level of inflation; i.e., deflated lungs were more likely to have damage from ultrasound. The impedance values of deflated lungs across all species and ages were found to be much closer to tissue impedance values than inflated lungs. This supported the hypothesis that ultrasound-induced lung damage was related to the transfer of ultrasonic energy into the lung.

A few mechanisms were proposed to explain the frequency dependence of the pressure reflection coefficients and the impedance values. These mechanisms were the modeling of the lung with a thin layer, i.e., the visceral pleura, the roughness of the lung surface, and a combination of both. Future studies will be required to determine the extent to which the proposed mechanisms contribute to the frequency dependence of the estimated pressure reflection coefficients and impedance values.

ACKNOWLEDGMENTS

The authors would like to thank Ade Oshinowa for her assistance in conducting the experiments and Mike Babbitt, Justin Boike, Yuihui Lui, Deepti Narla, and Grace Wang, for their technical assistance. This work was supported by NIH Grant No. R37EB02641.

American Institute of Ultrasound in Medicine (AIUM) (**2000**). "Mechanical bioeffects from diagnostic ultrasound: AIUM consensus statements," J. Ultrasound Med., **19**, 67–168.

Apfel, R. E. (**2001a**). "Comment on 'Ultrasound-induced lung hemorrhage is not caused by inertial cavitation,'" J. Acoust. Soc. Am., **110**, 1737.

Apfel, R. E. (**2001b**). "Reply to Frizzell *et al.*'s comment to our comment," J. Acoust. Soc. Am. **110**, 1740–1741.

Bachofen, H., Amman, A., Wangensteen, D., and Weibel, E. R. (**1982**). "Perfusion fixation of lungs for structure-function analysis: credits and limitations," J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **53**, 528–533.

Baggs, R., Penney, D. P., Cox, C., Child, S. Z., Raeman, C. H., Dalecki, D., and Carstensen, E. L. (**1996**). "Thresholds for ultrasonically induced lung hemorrhage in neonatal swine," Ultrasound Med. Biol. **22**, 119–128.

Bauld, T. J., and Schwan, H. P. (**1974**). "Attenuation and reflection of ultrasound in canine lung tissue," J. Acoust. Soc. Am. **56**, 1630–1637.

Carstensen, E. L., Gracewski, S., and Dalecki, D. (**2000**). "The search for cavitation *in vivo*," Ultrasound Med. Biol. **26**, 1377–1385.

Child, S. Z., Hartman, C. L., Schery, L. A., and Carstensen, E. L. (**1990**). "Lung damage from exposure to pulsed ultrasound," Ultrasound Med. Biol. **16**, 817–825.

Dalecki, D., Child, S. Z., Raeman, C. H., Cox, C., and Carstensen, E. L. (**1997a**). "Ultrasonically induced lung hemorrhage in young swine," Ultrasound Med. Biol. **23**, 777–781.

Dalecki, D., Child, S. Z., Raeman, C. H., Penney, D. P., Cox, C., and Carstensen, E. L. (**1997b**). "Age dependence of ultrasonically induced lung hemorrhage in mice," Ultrasound Med. Biol. **23**, 767–776.

Dunn, F. (**1974**). "Attenuation and speed of ultrasound in lung," J. Acoust. Soc. Am. **56**, 1638–1639.

Dunn, F. (**1986**). "Attenuation and speed of ultrasound in lung: Dependence upon frequency and inflation," J. Acoust. Soc. Am. **80**, 1248–1250.

Dunn, F., and Fry, W. J. (**1961**). "Ultrasonic absorption and reflection by lung tissue," Phys. Med. Biol. **5**, 401–410.

Frizzell, L. A., Chen, E., and Lee, C. (**1994**). "Effects of pulsed ultrasound on the mouse neonate: Hind limb paralysis and lung hemorrhage," Ultrasound Med. Biol. **20**, 53–63.

Frizzell, L. A., Kramer, J. M., Zachary, J. F., and O'Brien, W. D., Jr. (**2001a**). "Comment on Apfel's second comment," J. Acoust. Soc. Am. **110**, 1742.

Frizzell, L. A., Kramer, J. M., Zachary, J. F., and O'Brien, W. D., Jr. (**2001**). "Response to Comment on 'Ultrasound-induced lung hemorrhage is not caused by inertial cavitation.'" J. Acoust. Soc. Am. **110**, 1738–1739.

Frizzell, L. A., Zachary, J. F., and O'Brien, W. D. Jr. (**2003**). "Effect of pulse polarity and energy on ultrasound-induced lung hemorrhage in adult rats," J. Acoust. Soc. Am. **113**, 2912–2918.

Gil, J., Bachofen, H., Gehr, P., and Weibel, E. R. (**1979**). "Alveolar volume-surface area relation in air- and saline-filled lungs fixed by vascular perfusion," J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **47**, 990–1001.

Hartman, C., Child, S. Z., Mayer, R., Schenk, E., and Carstensen, E. L. (**1990**). "Lung damage from exposure to fields of an electrohydraulic lithotripter," Ultrasound Med. Biol. **16**, 675–679.

Hartman, C. L., Child, S. Z., Penney, D. P., and Carstensen, E. L. (**1992**). "Ultrasonic heating of lung tissue," J. Acoust. Soc. Am. **91**, 513–516.

Hayatdavoudi, G., Crapo, J. D., Miller, F. J., and O'Neil, J. J. (**1980**). "Factors determining degree of inflation in intratracheally fixed rat lungs," J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **48**, 389–393.

Holland, C. K., Sandstrom, K., Zheng, X., Rodriguey, J., and Roy, R. A. (**1994**). "The acoustic field of a pulsed Doppler diagnostic ultrasound system near a pressure release surface," J. Acoust. Soc. Am. **95**, 2855.

Johnston, R. L., and Dunn, F. (**1976**). "Influence of subarachnoid structures on transmeningeal ultrasonic propagation," J. Acoust. Soc. Am. **60**, 1225–1227.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).

Lizzi, F. L., Greenebaum, M., Feleppa, F. J., Elbaum, M., and Coleman, D. J. (**1983**). "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373.

Mikhak, Z., and Pederson, P. C. (**2002**). "Acoustic attenuation properties of

the lung: An open question," Ultrasound Med. Biol. **28**, 1209–1216.
O'Brien, W. D. Jr., Frizzell, L. A., Schaeffer, D. J., and Zachary, J. F. (**2001a**). "Superthreshold behavior of ultrasound-induced lung hemorrhage in adult mice and rats: Role of pulse repetition frequency and exposure duration," Ultrasound Med. Biol. **27**, 267–277.
O'Brien, W. D. Jr., Frizzell, L. A., Weigel, R. M. *et al.* (**2000**). "Ultrasound-induced lung hemorrhage is not caused by inertial cavitation," J. Acoust. Soc. Am. **108**, 1290–1297.
O'Brien, W. D. Jr., and Zachary, J. F. (**1997**). "Lung damage assessment from exposure to pulsed-wave ultrasound in the rabbit, mouse, and pig," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 473–485.
O'Brien, W. D. Jr., Kramer, J. M., Waldrop, T. G., Frizzell, L. A., Miller, R. J., Blue, J. P., and Zachary, J. F. (**2002**). "Ultrasound-induced lung hemorrhage: Role of acoustic boundary conditions at the pleural surface," J. Acoust. Soc. Am. **111**, 1102–1109.
O'Brien, W. D. Jr., Simpson, D. G., Frizzell, L. A., and Zachary, J. F. (**2001b**). "Superthreshold behavior and threshold estimates of ultrasound-induced lung hemorrhage in adult rats: Role of beamwidth," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1695–1705.
O'Brien, W. D. Jr., Simpson, D. G., Frizzell, L. A., and Zachary, J. F. (**2003a**). "Threshold estimates and superthreshold behavior of ultrasound-induced lung hemorrhage in adult rats: Role of pulse duration," Ultrasound Med. Biol. **29**, 1625–1634.
O'Brien, W. D. Jr., Simpson, D. G., Frizzell, L. A., and Zachary, J. F. (**2004**). "Effect of contrast agent on the incidence and magnitude of ultrasound-induced lung hemorrhage in rats," Echocardiogr. **21**, 417–422.
O'Brien, W. D. Jr., Simpson, D. G., Ho, M. H., Miller, R. J., Frizzell, L. A., and Zachary, J. F. (**2003b**). "Superthreshold behavior and threshold estimation of ultrasound-induced lung hemorrhage in pigs: Role of age dependency," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 153–169.
O'Brien, W. D. Jr., Yan, Y., and Simpson, D. G. (**2008**). "Threshold estimation and superthreshold behavior of ultrasound-induced lung hemorrhage in rats: Role of age dependency," Ultrasound Med. Biol. in press.
O'Brien, W. D. Jr., Yan, Y., Simpson, D. G., Frizzell, L. A., Miller, R. J., Blue, J. P. Jr., and Zachary, J. F. (**2006**). "Threshold estimation of ultrasound-induced lung hemorrhage in adult rabbits, and comparison of thresholds in mice, rats, rabbits and pigs," Ultrasound Med. Biol. **32**, 1793–1804.
Oelze, M. L., Miller, R. J., Blue, J. P. Jr., Zachary, J. F., and O'Brien, W. D. Jr. (**2003**). "Impedance measurements of *ex vivo* rat lung at different volumes of inflation," J. Acoust. Soc. Am. **114**, 3384–3393.
Oelze, M. L., Sabatier, J. M., and Raspet, R. (**2001**). "Roughness characterization of porous soil with acoustic backscatter," J. Acoust. Soc. Am. **109**, 1826–1832.
Pederson, P. C., and Ozcan, H. S. (**1986**). "Ultrasound properties of lung tissue and their measurements," Ultrasound Med. Biol. **12**, 483–499.
Penney, D. P., Schenk, E. A., Maltby, K., Hartman-Raeman, C., Child, S. Z., and Carstensen, E. L. (**1993**). "Morphologic effects of pulsed ultrasound in the lung," Ultrasound Med. Biol. **19**, 127–135.
Raeman, C. H., Child, S. Z., and Carstensen, E. L. (**1993**). "Timing of exposures in ultrasonic hemorrhage of murine lung," Ultrasound Med. Biol. **19**, 507–517.
Raeman, C. H., Child, S. Z., Dalecki, D., Cox, C., and Carstensen, E. L. (**1996**). "Exposure-time dependence of the threshold for ultrasonically induced murine lung hemorrhage," Ultrasound Med. Biol. **22**, 139–141.
Raeman, C. H., Dalecki, D., Child, S. Z., *et al.* (**1997**). "Albunex does not increase the sensitivity of the lung to pulsed ultrasound," Echocardiogr. **6**, 553–557.
Sagar, K. B., Rhyne, T. L., Myers, G. S., and Lees, R. S. (**1978**). "Characterization of normal and abnormal pulmonary surface by reflected ultrasound," Chest **74**, 29–33.
Stahl, W. R. (**1967**). "Scaling of respiratory variables in mammals," J. Appl. Physiol. **22**, 453–460.
Tarantal, A. F., and Canfield, D. R. (**1994**). "Ultrasound-induced lung hemorrhage in the monkey," Ultrasound Med. Biol. **20**, 65–72.
Tyler, W. S., and Julian, M. D. (**1992**). "Gross and subgross anatomy of lungs, pleura, connective tissue septa, distal airways, and structural units," in *Comparative Biology of the Normal Lung*, edited by R. A. Parent (CRC, Boca Raton, FL), Vol. **1**, 37–47.
Zachary, J. F., Blue, J. P. Jr., Miller, R. J., Ricconi, B. J., Eden, J. G., and O'Brien, W. D. Jr. (**2006**). "Lesions of ultrasound-induced lung hemorrhage are not consistent with thermal injury," Ultrasound Med. Biol. **32**, 1763–1770.
Zachary, J. F., Frizzell, L. A., Norrell, K. S., Blue, J. P. Jr., Miller, R. J., and O'Brien, W. D. Jr. (**2001a**). "Temporal and spatial evaluation of lesion reparative responses following superthreshold exposure of rat lung to pulsed ultrasound," Ultrasound Med. Biol. **27**, 829–839.
Zachary, J. F., and O'Brien, W. D. Jr. (**1995**). "Lung lesions induced by continuous- and pulsed-wave (diagnostic) ultrasound in mice, rabbits, and pigs," Vet. Pathol. **32**, 43–54.
Zachary, J. F., Sempsrott, J. M., Frizzell, L. A., Simpson, D. G., and O'Brien, W. D. Jr. (**2001b**). "Superthreshold behavior and threshold estimation of ultrasound-induced lung hemorrhage in adult mice and rats," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 581–592.

Born iterative reconstruction using perturbed-phase field estimates

Jeffrey P. Astheimer
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

Robert C. Waag
Department of Electrical and Computer Engineering and Department of Imaging Sciences, University of Rochester, Rochester, New York 14627

(Received 24 January 2008; revised 16 June 2008; accepted 23 June 2008)

A method of image reconstruction from scattering measurements for use in ultrasonic imaging is presented. The method employs distorted-wave Born iteration but does not require using a forward-problem solver or solving large systems of equations. These calculations are avoided by limiting intermediate estimates of medium variations to smooth functions in which the propagated fields can be approximated by phase perturbations derived from variations in a geometric path along rays. The reconstruction itself is formed by a modification of the filtered-backpropagation formula that includes correction terms to account for propagation through an estimated background. Numerical studies that validate the method for parameter ranges of interest in medical applications are presented. The efficiency of this method offers the possibility of real-time imaging from scattering measurements. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967482]

PACS number(s): 43.80.Ev, 43.60.Pt, 43.35.Wa, 43.80.Qf [TDM] Pages: 2353–2363

I. INTRODUCTION

In medical ultrasound, reconstruction of sound speed and density variations from scattering measurements promises higher resolution than the usual *b*-scan images that are produced by pulse-echo scanning. In fact, uniqueness results[1,2] imply that these medium parameters are completely determined by measurements of monochromatic scattering. However, an image resolution that improves on the diffraction limit of the apparatus (sometimes called superresolution) relies on the nonlinear relationship that exists between medium variations and scattering and is unraveled by inverting a sequence of linearizations. The first linear reconstruction has a resolution determined by the *f*-number for the apparatus. However, once an estimate of the medium is available, attention can focus on deviations from the estimate, and these deviations can be imaged using a revised linear model that incorporates the estimate. In effect, the initial estimate of the medium becomes part of the imaging apparatus, and the *f*-number of this new "apparatus" is smaller than the *f*-number of the original.

The Born approximation is the basis for the linear models used in most iterative reconstruction methods. Examples of such methods are the algebraic technique described in Ref. 3, the sinc-basis moment method discussed in Ref. 4, and the eigenfunction expansion methods detailed in Ref. 5–7. One reason for continuing to pursue new approaches is the high cost in computation time involved in forming nonlinear solutions. Distorted-wave Born estimates described in Ref. 8 are essentially Newton iterations toward the solution of the nonlinear system of Lippman–Schwinger equations[9] that relate the medium variations to the scattering measurements. Each Newton step requires linearization of the equations in the vicinity of the current estimate, and this linearization entails propagation of multiple incident fields through the estimated scattering medium.

To make the computation of propagation more efficient, the medium variations can be expanded in terms of a well-chosen finite-dimensional basis of functions. Products of retransmitted eigenfields of the scattering operator were proposed as such a basis in Ref. 5 because these fields concentrate in regions where the scattering is strong and because the resulting solution is characterized by an appealing constrained minimization property [Eqs. (15) and (16) of Ref. 5]. In Refs. 6 and 7, the cost of solving forward problems was eliminated by using estimates of the background medium that were homogeneous cylinders because analytic expressions exist for fields propagated through this object. Although a homogeneous cylinder may be a natural choice for breast imaging by a ring transducer system, other backgrounds (e.g., concentric cylinders) can also be used in similar circumstances.

This study takes a different tack. Instead of requiring the background estimates to have explicit analytic solutions for propagated fields, the estimates are required to be sufficiently smooth so that the propagated fields can be approximated using phase perturbations found by integrating the variations in acoustic path length along rays. This approach is limited to the use of scattering measurements from plane-wave illuminations rather than the eigenfield illuminations used in Refs. 5–7. Because propagated fields are only approximated, modest inaccuracies are introduced in the reconstructions. However, these inaccuracies are offset by significant efficiencies that accrue from the approximations. For example, solving the large systems of linear equations that usually appear in Newton or Born iterations is not necessary. Thus, while the individual iterations are not quite as accurate, the

overall efficiency of the iterations is better and has the improved possibility of being implemented in a way that permits real-time imaging.

Born iterations for large-scale backgrounds can be very specific. For example, if preliminary measurements identify the background as a homogeneous region or as multiple homogeneous regions with prescribed boundaries, then initial iterations can concentrate on determining the constant medium parameters within these regions and can defer a detailed resolution until the final steps. In such cases, the question of convergence is specific to the geometry of the background. This work focuses on an efficient reconstruction of fine details after the background parameters have been firmly established.

The experimental setting for which the methods of this paper are described is the ring transducer system described in Ref. 10. This system consists of an array of transducer elements that extend around the full circumference of a circle and that can be assigned independent complex amplitudes for illumination as well as independent complex weights for accumulating received measurements. The scattering object is placed in the center of the ring and is imaged in two dimensions. The transducer elements are sufficiently close that transmitted amplitudes can be assigned to produce very nearly monochromatic plane waves incident throughout the vicinity of the scattering object.

The theory that supports the reconstruction method is presented in Sec. II. Section III describes numerical experiments that validate the approach. The performance and accuracy of the method are discussed in Sec. IV. Section V gives conclusions from the study. In the Appendix, the exact field that results from propagating an incident plane wave through a homogeneous cylinder is compared with the approximate field that is obtained by phase perturbation for cylinder parameters matching those of the cylindrical object studied in Ref. 7.

II. THEORY

A. Estimation of medium variations

Let $e(\mathbf{x},\theta_n)=e^{ik\mathbf{u}(\theta_n)\cdot\mathbf{x}}$ $(n=1,2,\ldots,M)$ be monochromatic plane waves with spatial-frequency magnitude k traveling in directions $\theta_1,\ldots,\theta_M$ that are evenly distributed around the unit circle so that each $e_n(\mathbf{x})$ satisfies the homogeneous Helmholtz equation

$$(\Delta+k^2)e(\mathbf{x},\theta_n)=0. \tag{1}$$

Then, for each n, let $u_B(\mathbf{x},\theta_n)$ and $u(\mathbf{x},\theta_n)$ be the solutions to the equations

$$(\Delta+k^2)u_B(\mathbf{x},\theta_n)=k^2u_B(\mathbf{x},\theta_n)\,\eta_B(\mathbf{x}) \tag{2}$$

and

$$(\Delta+k^2)u(\mathbf{x},\theta_n)=k^2u(\mathbf{x},\theta_n)[\,\eta_B(\mathbf{x})+\Delta\,\eta_B(\mathbf{x})], \tag{3}$$

where $\eta(\mathbf{x})=\eta_B(\mathbf{x})+\Delta\,\eta_B(\mathbf{x})$ is a decomposition of the unknown medium variations as the sum of an assumed background term and a deviation term and where $u_B(\mathbf{x},\theta_n)$ and $u(\mathbf{x},\theta_n)$ are both assumed to have a nonradiating component equal to $e(\mathbf{x},\theta_n)$. All three medium variation terms are assumed to vanish outside a bounded region Ω. The fields $u_B(\mathbf{x},\theta_n)$ and $u(\mathbf{x},\theta_n)$ are total fields that result from $\eta_B(\mathbf{x})$ and $\eta(\mathbf{x})$ by applying the incident field $e(\mathbf{x},\theta_n)$ to the scattering region Ω, respectively.

Now subtract Eq. (1) from Eq. (2) to obtain

$$(\Delta+k^2)[u_B(\mathbf{x},\theta_n)-e(\mathbf{x},\theta_n)]=k^2u_B(\mathbf{x},\theta_n)\,\eta_B(\mathbf{x}) \tag{4}$$

and also subtract Eq. (2) from Eq. (3) to obtain

$$(\Delta+k^2[1-\eta_B(\mathbf{x})])[u(\mathbf{x},\theta_n)-u_B(\mathbf{x},\theta_n)]=k^2u(\mathbf{x},\theta_n)\Delta\,\eta_B(\mathbf{x}). \tag{5}$$

The assumptions made earlier regarding the asymptotic behavior of fields $u_B(\mathbf{x},\theta_n)$ and $u(\mathbf{x},\theta_n)$ imply that $u_B(\mathbf{x},\theta_n)-e(\mathbf{x},\theta_n)$ and $u(\mathbf{x},\theta_n)-u_B(\mathbf{x},\theta_n)$ satisfy the Sommerfeld radiation conditions, so integrating both sides of Eq. (4) against the radiating Green's function $G_0(\mathbf{x},\mathbf{y})$ for the operator $(\Delta+k^2)$ gives

$$u_B(\mathbf{x},\theta_n)-e(\mathbf{x},\theta_n)=k^2\iint u_B(\mathbf{y},\theta_n)G_0(\mathbf{x},\mathbf{y})\,\eta_B(\mathbf{y})d^2\mathbf{y} \tag{6}$$

and integrating both sides of Eq. (5) against the radiating Green's function $G_B(\mathbf{x},\mathbf{y})$ for the operator $(\Delta+k^2[1-\eta_B(\mathbf{x})])$ gives

$$u(\mathbf{x},\theta_n)-u_B(\mathbf{x},\theta_n)=k^2\iint u(\mathbf{y},\theta_n)G_B(\mathbf{x},\mathbf{y})\Delta\,\eta_B(\mathbf{y})d^2\mathbf{y}. \tag{7}$$

The left side of Eq. (6) is the change in the field at $\mathbf{x}$ that is caused by the introduction of the medium variations $\eta_B(\mathbf{x})$ and will be denoted as the scattered field $u_B^s(\mathbf{x})$. Similarly, the left side of Eq. (7) is the further change in the field at $\mathbf{x}$ that is caused by the addition of the medium variations $\Delta\,\eta_B(\mathbf{x})$ and will be denoted as the scattered field $u_\Delta^s(\mathbf{x})$. Using this notation, Eqs. (6) and (7) become

$$u_B^s(\mathbf{x},\theta_n)=k^2\iint u_B(\mathbf{y},\theta_n)G_0(\mathbf{x},\mathbf{y})\,\eta_B(\mathbf{y})d^2\mathbf{y} \tag{8}$$

and

$$u_\Delta^s(\mathbf{x},\theta_n)=k^2\iint u(\mathbf{y},\theta_n)G_B(\mathbf{x},\mathbf{y})\Delta\,\eta_B(\mathbf{y})d^2\mathbf{y}, \tag{9}$$

respectively.

Equations (8) and (9) give expressions for the scattered fields at location $\mathbf{x}$, but scattering measurements are formed as weighted accumulations of field values across the array of transducer elements at locations $\mathbf{x}_1,\ldots,\mathbf{x}_N$ as follows:

$$\sum_{i=1}^{N}\alpha_iu_B^s(\mathbf{x}_n,\theta_n)=k^2\iint u_B(\mathbf{y},\theta_n)\times\left[\sum_{i=1}^{N}\alpha_iG_0(\mathbf{x}_i,\mathbf{y})\right]\eta_B(\mathbf{y})d^2\mathbf{y} \tag{10}$$

and

$$\sum_{i=1}^{N} \alpha_i u_\Delta^s(\mathbf{x}_i, \theta_n) = k^2 \iint u(\mathbf{y}, \theta_n) \times \left[\sum_{i=1}^{N} \alpha_i G_B(\mathbf{x}_i, \mathbf{y}) \right] \Delta\eta_B(\mathbf{y}) d^2\mathbf{y}. \quad (11)$$

The weights $\alpha_1, \ldots, \alpha_N$ may be chosen in different ways to shape different receiver sensitivity patterns for the bracketed sum inside the integral in Eq. (10). The weights are chosen here to form the M sensitivity patterns

$$\sum_{i=1}^{N} \alpha_i^{(n)} G_0(\mathbf{x}_i, \mathbf{y}) = e(\mathbf{y}, \theta_n), \quad n = 1,2, \ldots, M \quad (12)$$

that mirror the M incident fields. Reciprocity then implies that the bracketed term in Eq. (11) produces the sensitivity patterns

$$\sum_{i=1}^{N} \alpha_i^{(n)} G_B(\mathbf{x}_i, \mathbf{y}) = u_B(\mathbf{y}, \theta_n), \quad n = 1,2, \ldots, M. \quad (13)$$

Pairing each incident field with each receiver sensitivity pattern results in the M^2 set of measurements given by

$$M_{nm}^B = k^2 \iint u_B(\mathbf{y}, \theta_n) e(\mathbf{y}, \theta_m) \eta_B(\mathbf{y}) d^2\mathbf{y} \quad (14)$$

and also the M^2 measurements given by

$$M_{nm}^\Delta = k^2 \iint u(\mathbf{y}, \theta_n) u_B(\mathbf{y}, \theta_m) \Delta\eta_B(\mathbf{y}) d^2\mathbf{y}. \quad (15)$$

The right sides of Eqs. (14) and (15) express the scattering measurements in terms of the medium variations $\eta_B(\mathbf{y})$ and $\Delta\eta_B(\mathbf{y})$, respectively. However, these relationships are nonlinear because of the implicit dependency of the fields $u_B(\mathbf{y}, \theta_n)$ and $u(\mathbf{y}, \theta_n)$ on $\eta_B(\mathbf{y})$ and $\Delta\eta_B(\mathbf{y})$, respectively. Equations (14) and (15) can be linearized by invoking the Born approximation that replaces $u_B(\mathbf{y}, \theta_n)$ in Eq. (14) with the homogeneous field $e(\mathbf{y}, \theta_n)$ and replaces $u(\mathbf{y}, \theta_n)$ in Eq. (15) with the field $u_B(\mathbf{y}, \theta_n)$ that propagates through the background variations. This gives the equations

$$M_{nm}^B = k^2 \iint e(\mathbf{y}, \theta_n) e(\mathbf{y}, \theta_m) \eta_B(\mathbf{y}) d^2\mathbf{y} \quad (16)$$

and

$$M_{nm}^\Delta = k^2 \iint u_B(\mathbf{y}, \theta_n) u_B(\mathbf{y}, \theta_m) \Delta\eta_B(\mathbf{y}) d^2\mathbf{y}. \quad (17)$$

Equation (17) is a general relationship between scattering measurements and medium variations that will be used for reconstruction. Equation (16) that applies to the reconstruction of medium variations in an empty background is a special case of Eq. (17) because if $\eta_B(\mathbf{x})=0$, then the terms $u_B(\mathbf{y}, \theta_n)$ and $u_B(\mathbf{y}, \theta_m)$ in the integrand of Eq. (17) become $e(\mathbf{y}, \theta_n)$ and $e(\mathbf{y}, \theta_m)$, respectively. Thus, Eq. (16) is used for an initial estimate and Eq. (17) is used for successive refinements.

Equations (16) and (17) can be solved for the medium variations by estimating a pseudoinverse for the integral kernel given by the imaginary part of the Green's function for the medium.[11] A simpler approach used here is to adopt specific finite-dimensional expansions for $\eta_B(\mathbf{y})$ and $\Delta\eta_B(\mathbf{y})$ and the forms

$$\eta_B(\mathbf{y}) = \sum_{i \leqslant j=1}^{M} Q_{ij}^B \overline{e(\mathbf{x}, \theta_j) e(\mathbf{x}, \theta_k)} \quad (18)$$

and

$$\Delta\eta_B(\mathbf{y}) = \sum_{i \leqslant j=1}^{M} Q_{ij}^\Delta \overline{u_B(\mathbf{x}, \theta_j) u_B(\mathbf{x}, \theta_k)} \quad (19)$$

suggested by the inner products in Eqs. (16) and (17) that define the measurements.[a] The limitation $i \leqslant j$ for the summation indices is imposed to avoid redundancy in the field products. Substituting these expansions into Eqs. (16) and (17) gives

$$M_{nm}^B = k^2 \sum_{i \leqslant j=1}^{M} \left[\iint e(\mathbf{y}, \theta_n) e(\mathbf{y}, \theta_m) \times \overline{e(\mathbf{x}, \theta_j) e(\mathbf{x}, \theta_k)} d^2\mathbf{y} \right] Q_{ij}^B \quad (20)$$

and

$$M_{nm}^\Delta = k^2 \sum_{i \leqslant j=1}^{M} \left[\iint u_B(\mathbf{y}, \theta_n) u_B(\mathbf{y}, \theta_m) \times \overline{u_B(\mathbf{x}, \theta_j) u_B(\mathbf{x}, \theta_k)} d^2\mathbf{y} \right] Q_{ij}^\Delta, \quad (21)$$

where the measurement indices are also limited by the constraint $n \leqslant m$ to avoid redundancy. This yields a system of $M(M+1)/2$ linear equations in the same number of unknowns. Solving these equations provides values for the Q_{ij}^B and Q_{ij}^Δ coefficients that can be used in Eqs. (18) and (19) to estimate the medium variations. The estimate given by Eq. (18) may be described as a Born estimate of the medium variations $\eta_B(\mathbf{x})$ relative to a homogeneous background, while the estimate given by Eq. (19) may be described as a distorted-wave Born estimate of the deviations $\Delta\eta_B(\mathbf{x})$ of the medium variations from the background $\eta_B(\mathbf{x})$.

B. Explicit solutions

The above methods for the estimates of medium variations from scattering measurements require solving the large matrix equations given in Eqs. (20) and (21). However, some modest approximations lead to analytic solutions and eliminate the need for lengthy numerical inversions. Equation (20) is the simpler of the two equations and serves as a model for the calculation.

Substituting exponential expressions for the e fields in the double integral in Eq. (20) gives

$$\int_\Omega e^{jk[\mathbf{u}(\theta_n)+\mathbf{u}(\theta_m)-\mathbf{u}(\theta_i)-\mathbf{u}(\theta_j)]\cdot\mathbf{x}} d^2\mathbf{x} = \hat{\chi}_\Omega(k[\mathbf{u}(\theta_i) + \mathbf{u}(\theta_j) - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]), \quad (22)$$

where $\chi_\Omega(\mathbf{x})$ is the characteristic function of the scattering volume, i.e., $\chi_\Omega(\mathbf{x})=1$ when $\mathbf{x} \in \Omega$ and is 0 otherwise, and

$\hat{\chi}_\Omega(\boldsymbol{\nu})$ denotes the Fourier transform of $\chi_\Omega(\mathbf{x})$. Using the right side of Eq. (22) in place of the integral in Eq. (20) gives

$$M^B_{nm} = k^2 \sum_{i \leqslant j=1}^{M} \hat{\chi}_\Omega(k[\mathbf{u}(\theta_i) + \mathbf{u}(\theta_j) - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]) Q^B_{ij}. \tag{23}$$

Since the size of the scattering volume Ω is large relative to the size of the local variations and since the Q^B_{ij} coefficients in Eq. (18) act as Fourier coefficients for the local variations, the Q^B_{ij} coefficients are nearly constant over the range of i and j indices where $\hat{\chi}_\Omega(k[\mathbf{u}(\theta_i)+\mathbf{u}(\theta_j)-\mathbf{u}(\theta_n)-\mathbf{u}(\theta_m)])$ is appreciable. Hence, Eq. (23) can be approximated by the diagonal system of equations

$$M^B_{nm} = \left[k^2 \sum_{i \leqslant j=1}^{M} \hat{\chi}_\Omega \times (k[\mathbf{u}(\theta_i) + \mathbf{u}(\theta_j) - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]) \right] Q^B_{nm} \tag{24}$$

with solution

$$Q^B_{nm} = M^B_{nm} \Big/ \left[k^2 \sum_{i \leqslant j=1}^{M} \hat{\chi}_\Omega \times (k[\mathbf{u}(\theta_i) + \mathbf{u}(\theta_j) - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]) \right]. \tag{25}$$

As Ω grows larger, this result becomes increasingly accurate. An even more explicit expression for the denominator may be obtained by interpreting the sum as a Riemann approximation for the integral

$$\frac{1}{2}\left(\frac{M}{2\pi}\right)^2 \int_0^{2\pi} \int_0^{2\pi} \hat{\chi}_\Omega \times (k[\mathbf{u}(\theta') + \mathbf{u}(\theta'') - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]) d\theta' d\theta''. \tag{26}$$

The $1/2$ factor that appears outside the integral is included to compensate for extending the limit of the θ' integral to the full 0 to 2π range rather than setting the upper limit to θ'', which is indicated by the limit $i \leqslant j$ of the j index in the sum. Applying the change in variables

$$\nu_x = \cos\theta' + \cos\theta''$$

$$\nu_y = \sin\theta' + \sin\theta'', \tag{27}$$

with Jacobian

$$\begin{vmatrix} \partial_{\theta'}\nu_x & \partial_{\theta''}\nu_x \\ \partial_{\theta'}\nu_y & \partial_{\theta''}\nu_y \end{vmatrix} = \begin{vmatrix} -\sin\theta' & -\sin\theta'' \\ \cos\theta' & \cos\theta'' \end{vmatrix} = |\sin(\theta' - \theta'')| \tag{28}$$

allows Eq. (26) to be rewritten as

$$\frac{1}{2}\left(\frac{M}{2\pi}\right)^2 \int \hat{\chi}_\Omega(k[\boldsymbol{\nu} - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]) \frac{d^2\boldsymbol{\nu}}{|\sin(\theta' - \theta'')|}. \tag{29}$$

However, because $|\sin(\theta' - \theta'')| = |\sin(\theta_n - \theta_m)|$, where $\hat{\chi}_\Omega(k[\boldsymbol{\nu} - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)])$ is appreciable, Eq. (29) simplifies to

$$\frac{(M/2\pi)^2}{2|\sin(\theta_n - \theta_m)|} \int \hat{\chi}_\Omega(k[\boldsymbol{\nu} - \mathbf{u}(\theta_n) - \mathbf{u}(\theta_m)]) d^2\boldsymbol{\nu} = \frac{M^2}{2k^2|\sin(\theta_n - \theta_m)|}, \tag{30}$$

where the integral is evaluated as $4\pi^2\chi_\Omega(0)/k^2 = 4\pi^2/k^2$. Substituting the right side of Eq. (30) for the denominator of Eq. (25) then gives

$$Q^B_{nm} = (2/M^2)|\sin(\theta_n - \theta_m)| M^B_{nm}, \tag{31}$$

and substituting the coefficient values given by Eq. (31) into the expansion in Eq. (18) gives

$$\eta_B(\mathbf{x}) = (2/M)^2 \sum_{n \leqslant m=1}^{M} |\sin(\theta_n - \theta_m)| M^B_{nm} e^{-jk[\mathbf{u}(\theta_n)+\mathbf{u}(\theta_m)]\cdot\mathbf{x}}. \tag{32}$$

Equation (32) is the well-known filtered-backpropagation reconstruction formula[12] that gives an explicit solution for the medium variations in terms of the scattering measurements. The above derivation is included here to provide a framework for establishing a similar expression for $\Delta\eta_B(\mathbf{x})$ based on the distorted-wave Born iteration to which attention is now given.

To simplify the integral in Eq. (21), the u_B fields must be given a more explicit form. This requires making assumptions about the behavior of $\eta_B(\mathbf{x})$. If $\eta_B(\mathbf{x})$ is slowly varying, then the u_B fields can be approximated inside Ω by plane waves with phase corrections that compensate for changes in the speed of propagation along rays. Thus,

$$u_B(\mathbf{x}, \theta_n) \approx \tilde{u}_B(\mathbf{x}, \theta_n) = e^{jk[\mathbf{u}(\theta_n)\cdot\mathbf{x} + \delta(\mathbf{x}, \theta_n)]}, \tag{33}$$

where the $\tilde{}$ indicates approximation and

$$\delta(\mathbf{x}, \theta_n) = \int_0^\infty \eta_B(\mathbf{x} - \beta\mathbf{u}(\theta_n)) d\beta. \tag{34}$$

These phase corrections are easily calculated from the assumed background $\eta_B(\mathbf{x})$.

Substituting these field expressions into the integral that forms the matrix in Eq. (22) gives

$$\int_\Omega e^{jk[\mathbf{u}(\theta_n)+\mathbf{u}(\theta_m)-\mathbf{u}(\theta_i)-\mathbf{u}(\theta_j)]\cdot\mathbf{x}} \times e^{jk[\delta(\mathbf{x},\theta_n)+\delta(\mathbf{x},\theta_m)-\delta(\mathbf{x},\theta_i)-\delta(\mathbf{x},\theta_j)]} d^2\mathbf{x}, \tag{35}$$

and oscillations in the left term produce cancellations that null the integral except when the direction angles θ_n and θ_m are close to the angles θ_i and θ_j. This implies that only elements near the diagonal of the matrix are appreciable, and the phase correction terms for those elements will nearly cancel. Thus, the integral in Eq. (35) can be replaced with the right side of Eq. (22). This approximation gives the system of equations in Eq. (21) the same form as Eq. (20). Hence, the same steps can be used to arrive at the solution. This yields

$$Q_{nm}^{\Delta} = (2/M)^2 |\sin(\theta_n - \theta_m)| M_{nm}^{\Delta}, \tag{36}$$

and substituting these coefficients and u_B field approximations into the expansion in Eq. (19) gives

$$\Delta\eta_B(\mathbf{x}) = (2/M^2) \sum_{n\leq m=1}^{M} |\sin(\theta_n - \theta_m)| M_{nm}^{\Delta} e^{-jk[\mathbf{u}(\theta_n)+\mathbf{u}(\theta_m)]\cdot\mathbf{x}} e^{-jk[\delta(\mathbf{x},\theta_n)+\delta(\mathbf{x},\theta_m)]}. \tag{37}$$

This startlingly simple result indicates that distorted-wave Born iterations can also be estimated using the filtered-backpropagation formula for a free space reconstruction provided that the expansion is adjusted with the appropriate phase correction terms.[b]

C. Incremental scattering measurements

Equation (37) depends on the measurements M_{nm}^{Δ} that are defined in Eq. (10) as the weighted sums,

$$M_{nm}^{\Delta} = \sum_{i=1}^{M} \alpha_i^{(m)} u_{\Delta}^{s}(\mathbf{x}_i, \theta_n), \tag{38}$$

where the $\alpha_i^{(m)}$ weights are chosen in accordance with Eq. (12) and the scattered field $u_{\Delta}^{s}(\mathbf{x}_i, \theta_n)$ is

$$u_{\Delta}^{s}(\mathbf{x}_i, \theta_n) = u(\mathbf{x}, \theta_n) - u_B(\mathbf{x}, \theta_n). \tag{39}$$

The difference on the right side of Eq. (39) may be further expanded using Eq. (6) to obtain

$$u_{\Delta}^{s}(\mathbf{x}_i, \theta_n) = u(\mathbf{x}_i, \theta_n) - e(\mathbf{x}_i, \theta_n) - k^2 \iint u_B(\mathbf{y}, \theta_n) G_0(\mathbf{x}_i, \mathbf{y}) \eta_B(\mathbf{y}) d^2\mathbf{y}, \tag{40}$$

and substituting the right side of this expansion in place of $u_{\Delta}^{s}(\mathbf{x}_i, \theta_n)$ in Eq. (38) gives

$$M_{nm}^{\Delta} = \left[\sum_{i=1}^{N} \alpha_i^{(m)} u(\mathbf{x}_i, \theta_n)\right] - \left[\sum_{i=1}^{N} \alpha_i^{(m)} e(\mathbf{x}_i, \theta_n)\right] - k^2 \iint u_B(\mathbf{y}, \theta_n) e(\mathbf{y}, \theta_m) \eta_B(\mathbf{y}) d^2\mathbf{y}. \tag{41}$$

The first term on the right side of Eq. (41) represents an explicit field measurement for a scattering experiment in which the incident field is $e(\mathbf{x}, \theta_n)$, and the second term is the same measurement in an experiment with the same incident field but without the scattering medium. The third term, however, represents scattering due to the assumed background and cannot be obtained experimentally. This term can, however, be evaluated by approximating the field $u_B(\mathbf{y}, \theta_m)$ by $e^{jk[\mathbf{u}(\theta_m)\cdot\mathbf{y}+\delta(\mathbf{y},\theta_n)]}$ as in Eq. (34) and replacing the $e(\mathbf{y}, \theta_n)$ term with the exponential $e^{jk\mathbf{u}(\theta_n)\cdot\mathbf{y}}$ to give

$$k^2 \iint u_B(\mathbf{y}, \theta_n) e(\mathbf{y}, \theta_m) \eta_B(\mathbf{y}) d^2\mathbf{y} = k^2 \iint e^{jk[\mathbf{u}(\theta_n)+\mathbf{u}(\theta_m)]\cdot\mathbf{y}} e^{jk\delta(\mathbf{y},\theta_n)} \eta_B(\mathbf{y}) d^2\mathbf{y}. \tag{42}$$

The right side of Eq. (42) is a spatial Fourier transform of $e^{jk\delta(\mathbf{y},\theta_n)}\eta_B(\mathbf{y})$, and the values for different θ_m ($m = 1, 2, \ldots, M$) can be efficiently extracted by interpolating values from a single two-dimensional fast Fourier transform.

III. NUMERICAL STUDIES

The reconstruction procedure described above uses the far-field measurements M_{nm} for the total field, together with an assumed background $\eta_B(\mathbf{x})$, to form a correction $\Delta\eta_B(\mathbf{x})$ to the medium variations. The steps in this process are as follows:

(1) For each incident direction θ_n ($n = 1, \ldots, M$), the incident plane wave $e^{jk_0\mathbf{u}(\theta_n)\cdot\mathbf{x}}$ is propagated through the background $\eta_B(\mathbf{x})$ to obtain the field $u_B(\mathbf{x}, \theta_n)$.
(2) The far-field pattern M_{nm}^{B} for the background $\eta_B(\mathbf{x})$ is computed.
(3) The correction $\Delta\eta_B(\mathbf{x})$ is estimated using the modified filtered-backpropagation expansion

$$\widetilde{\Delta\eta_B}(\mathbf{x}) = (2/M^2) \sum_{m\leq n=1}^{M} |\sin(\theta_n - \theta_m)| (M_{nm} - M_{nm}^{B}) \overline{u_B(\mathbf{x}, \theta_n) u_B(\mathbf{x}, \theta_m)}. \tag{43}$$

Equation (43) represents a significant improvement in efficiency over the more complete reconstructions carried out in Refs. 5–7 because the calculation of the inner products $\langle u_B(\mathbf{x}, \theta_i) u_B(\mathbf{x}, \theta_j), \overline{u_B(\mathbf{x}, \theta_n) u_B(\mathbf{x}, \theta_m)}\rangle$ and the inversion of the resulting matrix are unnecessary, but a wavy overbar has been added to the symbol for the correction potential to acknowledge that Eq. (43) is an approximation.

To improve the efficiency of the calculation, two additional approximations are considered. First, the u_B fields in Eq. (43) are replaced with the approximate fields given by Eq. (33). This results in the perturbed-phase filtered-backpropagation expansion

$$\tilde{\tilde{\Delta\eta}}_B(\mathbf{x}) = (2/M^2) \sum_{m\leq n=1}^{M} |\sin(\theta_n - \theta_m)| \times (M_{nm} - M_{nm}^{B}) \overline{\tilde{u}_B(\mathbf{x}, \theta_n) \tilde{u}_B(\mathbf{x}, \theta_m)}, \tag{44}$$

where an additional wavy overbar has been added to the term on the left side to indicate the additional level of approximation, and the use of approximate fields on the right side is also reflected by wavy overbars. This expression retains exact values for the far-field pattern M_{nm}^{B}. A second approximation entails substituting approximate values $\tilde{M}_{nm}^{B}$ that are obtained by using $\tilde{u}_B$ fields in place of u_B fields in the right side of Eq. (41). This gives the third expansion

$$\tilde{\tilde{\Delta\eta}}_B(\mathbf{x}) = (2/M^2) \sum_{m\leq n=1}^{M} |\sin(\theta_n - \theta_m) \times |(M_{nm} - \tilde{M}_{nm}^{B}) \overline{\tilde{u}_B(\mathbf{x}, \theta_n) \tilde{u}_B(\mathbf{x}, \theta_m)}. \tag{45}$$

To determine the effectiveness of these estimates, studies were conducted using both simulated and measured data. In the first study, Eq. (43) was used to reconstruct medium variations $\Delta\eta_B(\mathbf{x})$ in an assumed background medium $\eta_B(\mathbf{x})$

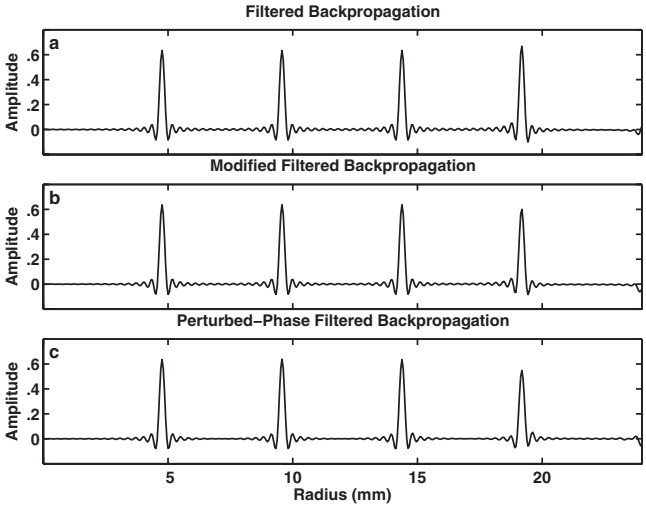


FIG. 1. Radial profiles of reconstructed unit-amplitude point reflectors relative to an assumed background. (a) Filtered backpropagation in an empty background. (b) Modified filtered backpropagation, $\Delta\tilde{\eta}_B(\mathbf{x})$, for a background cylinder with a 24 mm radius and a sound speed of 1.574 mm/μs using Bessel-function expansions for the internal fields. (c) Modified filtered backpropagation, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, for the same background cylinder except using phase perturbation to approximate the internal fields.

consisting of a homogeneous cylinder with a 48 mm diameter. The variations consisted of five unit-amplitude point reflectors that were positioned along a ray at radii that were spaced by one-fifth of the cylinder radius. The far-field measurements $M_{nm}^{\Delta}=M_{nm}-M_{nm}^{B}$ were computed numerically using the weak scattering formula in Eq. (17). Figures 1(a) and 1(b) are plots of the reconstruction $\Delta\tilde{\eta}_B(\mathbf{x})$ along the ray where the points are located. In Fig. 1(a), the sound speed for the background cylinder was chosen to be 1.509 mm/μs, which matched the sound speed for the ambient medium, so this plot illustrates the effectiveness of conventional filtered backpropagation in the absence of an assumed background. The sound speed for the background cylinder in Fig. 1(b) was 1.574 mm/μs, which is a high-contrast level relative to the ambient medium that is well beyond the range of variations that can be reconstructed using conventional filtered backpropagation. The same cylinder was used as the background for the reconstruction shown in Fig. 1(c) except that the estimate $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$ given Eq. (44) was used instead of $\Delta\tilde{\eta}_B(\mathbf{x})$. These calculations were all performed at a frequency of 2.5 MHz for 1024 incident plane waves with directions that were spaced evenly around the unit circle. The reconstructions were expanded on a polar grid with 500 radii ranging from 0 to 24 mm and with an angular refinement of $2\pi/1024$ radians. Since the incremental scattering measurements used in this simulation were generated as weak linear effects and the reconstructions are also linear, the plots in Fig. 1 are linear responses to the unit-amplitude point reflectors.

Further calculations were carried out using far-field measurements that were obtained experimentally. These measurements were also used in Ref. 7 and are described in greater detail in that article. The measurements were made using a ring transducer system with 2048 transducer elements. Scattering was produced by a tissue-mimicking phantom with dimensions and parameters shown in Fig. 2. The phantom included two scatter-free cylinders and three 0.1 mm diameter nylon filaments.

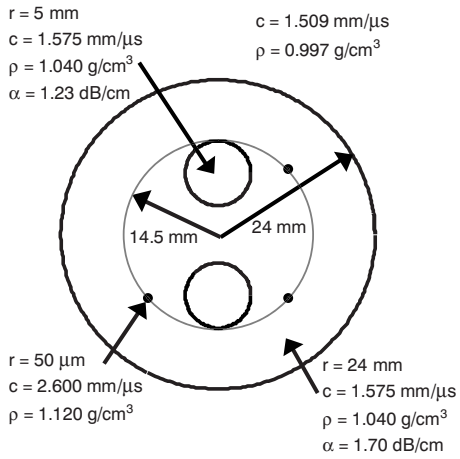


FIG. 2. Tissue-mimicking phantom used as a large-scale high-contrast scattering object for measurements.

The scattering measurements were used to produce four different reconstructions. An exact weak scattering result was found using an expansion of products of retransmitted eigenfields as described in Ref. 5. This result may also be interpreted as an expansion in the form of Eq. (19) where the Q_{nm} coefficients are found by solving the system of equations in Eq. (21). This result may be regarded as optimal in the least-squares sense described in Ref. 5 and, in that respect, may be regarded as the most accurate Born estimate. The other three reconstructions were produced using the approximate expansions given in Eqs. (43)–(45). However, since the background cylinder for these reconstructions was strongly attenuating, reciprocal fields were used instead of conjugate fields on the right hand sides of these equations. In each case, reconstructions were computed for 17 frequencies ranging from 1.4 to 2.66 MHz. These potentials were averaged together to produce multifrequency potentials that do not exhibit the oscillatory artifacts that appear in monochromatic potentials.

Impulse responses for each of these reconstructions were computed for a point reflector located midway between the center and the outer radius of the cylinder. Averaging monochromatic responses over frequency, as described above, resulted in a single frequency response for each reconstruction. The real parts of these impulse responses are very nearly circularly symmetric and are essentially independent of the reflector location. The imaginary parts are roughly antisymmetric along the ray from the center of the cylinder to the reflector; they vary with the position of the reflector and are much smaller than the real parts. Ignoring the effects of the imaginary components of the impulse responses allows the reconstructed potentials to be viewed as convolutions of the medium variations with the real parts of the impulse responses.

Frequency responses corresponding to the real parts of the impulse responses from each of the four different reconstructions are plotted in Fig. 3. The frequency responses for the approximation reconstructions are all essentially the

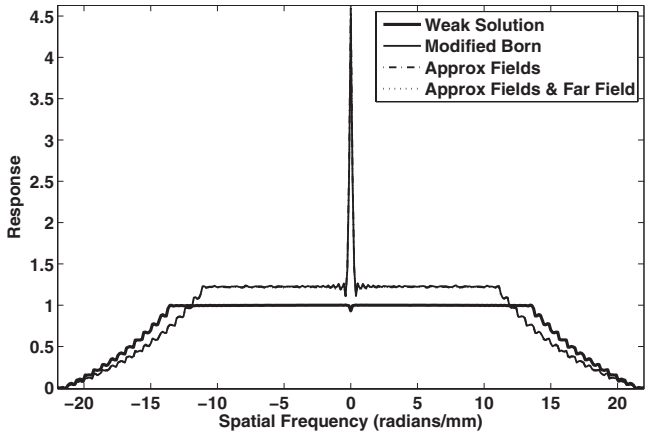


FIG. 3. Spatial-frequency responses corresponding to the real parts of frequency-averaged point-reflector reconstructions. (a) Complete inversion of the system of equations relating far-field measurements of weak scattering to medium variations. (b) Modified filtered backpropagation using Bessel expansions for the internal fields and the far-field pattern of the cylinder. (c) Modified filtered backpropagation using phase perturbation to approximate the internal fields. (d) Modified filtered backpropagation with the far-field pattern of the cylinder also approximated using the phase-perturbed plane waves.

same but differ from the response for the exact weak reconstruction. The large dc spikes in the frequency responses for the approximate reconstructions indicate that the asymptotic values of the impulse responses are nonzero. The flat region of the frequency responses for the approximate reconstructions only extends to $\pm 2\pi f_{\min}/c_1$, where $f_{\min}=1.4$ MHz and c_1 is the sound speed inside the cylinder, while the flat region of the response for the exact weak solution extends to $\pm 2\pi f_{\min}/c_0$, where c_0 is the speed of sound in water. However, all the responses taper off to 0 at $\pm 2\pi f_{\max}/c_0$, where $f_{\max}=2.66$ MHz.

Figure 4(a) is a gray-scale image of the exact weak reconstruction, and Figs. 4(b)–4(d) are gray-scale images, respectively, of the reconstructions $\Delta\tilde{\eta}_B(\mathbf{x})$, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, and $\Delta\tilde{\tilde{\tilde{\eta}}}_B(\mathbf{x})$ given by Eqs. (43)–(45), respectively. Filters were applied to the three approximate reconstructions to scale their frequency responses (as shown in Fig. 3) to match the frequency response for the weak solution. The filtered images differ from one another and also from the exact weak reconstruction because the reconstructions are not completely translation invariant and cannot be completely characterized as convolutions. The Cartesian plot in Fig. 5 provides a more quantitative comparison of the reconstructions along a cross section of the cylinder that intersects two of the filaments and one of the scatter-free regions.

Measurements for these calculations were obtained from 1024 incident plane waves with directions that were spaced evenly around the unit circle. All the reconstructions were expanded on a polar grid with 500 radii ranging from 0 and 24 mm and with an angular refinement of $2\pi/1024$ radians. Cubic splines were used to interpolate the potentials onto a rectangular grid with 1024 points in both dimensions. The gray scale for all the images show values of $\Delta\eta_B(\mathbf{x})$ between −0.04 and +0.04 that correspond to sound speed variations between 1.54 and 1.60 mm/μs.

IV. DISCUSSION

The studies described above validate the reconstructions given by Eqs. (43)–(45) under appropriate circumstances. The reconstructed points in the inhomogeneous object reconstructions shown in Figs. 1(b) and 1(c) have sidelobes that decay more quickly than those in the homogeneous reconstruction in Fig. 1(a). This improvement is due to the blurring of the Ewald disk caused by dispersion in the direction of plane waves that propagate across the cylinder boundary. The fourth point in the radial profile shown in Fig. 1(c) is slightly depressed, reflecting the increased error in phase-perturbation field approximations toward the edge of the cylinder. The plots in Fig. 1 may be interpreted as illustrations of the system responses for the $\Delta\tilde{\eta}_B(\mathbf{x})$ and $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$ reconstructions at different radii and are all very nearly the same as the $J_1(k\rho)/k\rho$ response for the homogeneous case.

Similar inferences are reached through an examination of the reconstructions of experimental data. The $\Delta\tilde{\eta}_B(\mathbf{x})$, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, and $\Delta\tilde{\tilde{\tilde{\eta}}}_B(\mathbf{x})$ reconstructions shown in Figs. 4(b)–4(d), respectively, are all similar toward the center of the cylinder but perform differently near the cylinder boundary. In each case, the medium variations in the interior of the cylinder are resolved nearly as well as in the image obtained by a full inversion of the inner product matrix shown in Fig. 4(a). This outcome can be investigated by expressing the expansion of Eq. (43) in the integral form as

$$\widetilde{\Delta\eta_B}(\mathbf{x}) = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} \sin(\theta-\theta')M^{\Delta}(\theta,\theta') \times \overline{u_B(\mathbf{x},\theta)u_B(\mathbf{x},\theta')}\,d\theta'\,d\theta, \tag{46}$$

where the far-field measurements $M^{\Delta}(\theta,\theta')$ with continuous angular parameters are

$$M^{\Delta}(\theta,\theta') = k^2\int_{\Omega} u_B(\mathbf{y},\theta)u_B(\mathbf{y},\theta')\Delta\eta_B(\mathbf{y})d^2\mathbf{y}. \tag{47}$$

Substituting Eq. (47) into Eq. (46) and interchanging the order of integration give

$$\widetilde{\Delta\eta_B}(\mathbf{x}) = \frac{k^2}{4\pi^2}\left[\int_0^{2\pi}\int_0^{2\pi} |\sin(\theta-\theta')|u_B(\mathbf{y},\theta)u_B(\mathbf{y},\theta') \times \overline{u_B(\mathbf{x},\theta)u_B(\mathbf{x},\theta')}\,d\theta'\,d\theta\right] \times \Delta\eta_B(\mathbf{y})d^2\mathbf{y}. \tag{48}$$

The bracketed double integral on the right side of Eq. (48) is the system response function for the reconstruction process and takes the form of a blurred delta function $\delta(\mathbf{y}-\mathbf{x})$. When the distance between $\mathbf{x}$ and $\mathbf{y}$ is large, rapid oscillations in the u_B fields cause cancellations that null the double integral, so attention can focus on the values of the system response function for $\mathbf{x}$ and $\mathbf{y}$ values that are close to one another. Substituting the perturbed-phase plane wave field approximation

$$u_B(\mathbf{x},\theta) = e^{jk[\mathbf{u}(\mathbf{x},\theta)\cdot\mathbf{x}+\delta(\mathbf{x},\theta)]} \tag{49}$$

for the u_B fields in Eq. (48) gives the expression

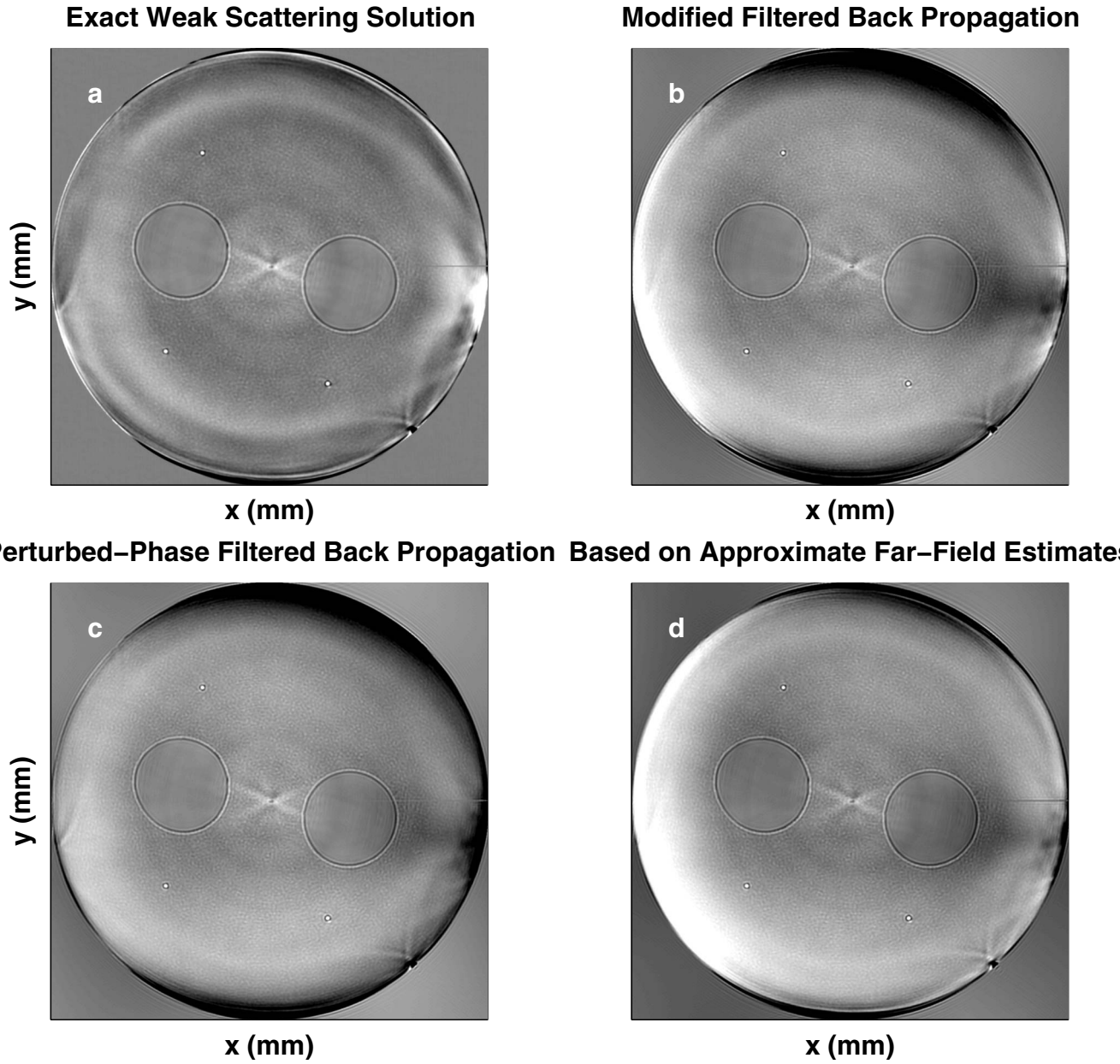


FIG. 4. Cross sections through the two filaments on the left side of reconstructions. (a) Complete inversion of the system of equations relating far-field measurements of weak scattering to medium variations. (b) Modified filtered backpropagation, $\Delta\tilde{\eta}_B(\mathbf{x})$, using Bessel expansions for the internal fields and the far-field pattern of the cylinder. (c) Modified filtered backpropagation, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, using phase perturbation to approximate the internal fields. (d) Modified filtered backpropagation, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, with the far-field pattern of the cylinder also approximated using the phase-perturbed plane waves.

$$\widetilde{\Delta\eta_B}(\mathbf{x}) = \frac{k^2}{4\pi^2}\left[\int_0^{2\pi}\int_0^{2\pi} |\sin(\theta-\theta')| e^{jk\{[\mathbf{u}(\theta)+\mathbf{u}(\theta')]\cdot(\mathbf{y}-\mathbf{x})+[\delta(\mathbf{y},\theta)-\delta(\mathbf{x},\theta)]+[\delta(\mathbf{y},\theta')-\delta(\mathbf{x},\theta')]\}} d\theta' d\theta\right]\Delta\eta_B(\mathbf{y}) d^2\mathbf{y}, \tag{50}$$

in which the phase perturbations only occur as differences between values at neighboring points. Because the perturbations are slowly varying, these differences can be neglected, yielding the same system response function as in the homogeneous reconstruction. However, this calculation is only valid locally, i.e., in regions where the propagated fields can be accurately estimated by phase perturbation.

For attenuating backgrounds, the spatial frequency k is complex. In that case, reciprocal rather than conjugate fields should be used in the expansions that represent the medium variations. Thus, reciprocal rather than conjugate $\mathbf{u}_B(\mathbf{x})$ terms should appear on the right hand sides of Eqs. (46) and (49). Use of reciprocal fields produces cancellation of the magnitude variations in the four field products in the integrand on the right hand side of Eq. (48). Equations (49) and (50) are valid for both attenuating and nonattenuating fields.

Equation (43) was derived earlier by making a diagonal approximation to the matrix of inner products in the linear

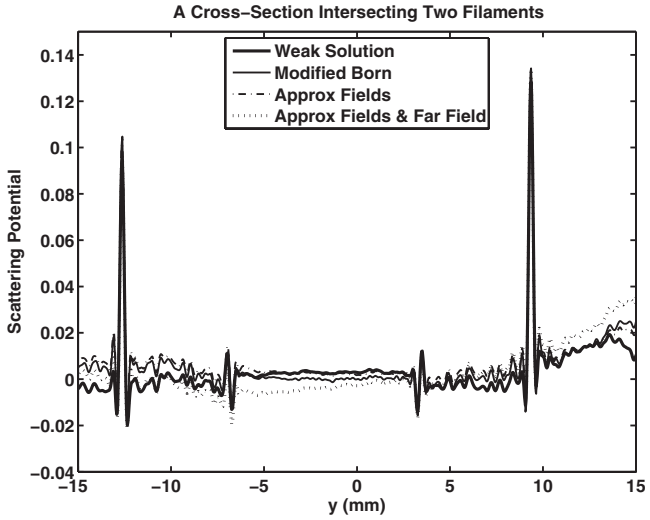


FIG. 5. Spatial-frequency responses corresponding to the real parts of frequency-averaged point-reflector reconstructions. (a) Complete inversion of the system of equations relating far-field measurements of weak scattering to medium variations. (b) Modified filtered backpropagation using Bessel expansions for the internal fields and the far-field pattern of the cylinder. (c) Modified filtered backpropagation using phase perturbation to approximate the internal fields. (d) Modified filtered backpropagation with the far-field pattern of the cylinder also approximated using the phase-perturbed plane waves.

system of equations that relate weak far-field scattering to medium variations. Thus, Eq. (43) should give a less accurate reconstruction than inversion of the original system of equations. However, the matrix of inner products in the more exact reconstruction produces small slowly varying oscillations that extend into the interior of the cylinder. This behavior is evidenced in Fig. 4(a) by the radial undulations that are superimposed on the medium variations. This effect is less prominent in the approximate reconstructions. However, reconstruction near the edge of the cylinder becomes progressively worse in Figs. 4(b)–4(d), which is consistent with the increased dependency on field approximations in the estimates $\Delta\tilde{\eta}_B(\mathbf{x})$, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, and $\Delta\tilde{\tilde{\tilde{\eta}}}_B(\mathbf{x})$. The eight-point star pattern in the center of these reconstructions is an artifact caused by crosstalk in the data acquisition apparatus.

Figure 4(d) is particularly noteworthy because this reconstruction relies on phase-perturbation approximations for both the u_B fields and the M^B_{nm} far-field measurements. In the case of a cylindrical background, there is no advantage in using these estimates since the u_B fields and the far-field measurements M^B_{nm} can be readily calculated from Bessel-function expansions. However, this is not the case if the boundary is irregular. Recent studies[13] have demonstrated that more complicated convex boundaries can be determined, *a priori*, from experimental measurements. The perturbed-phase filtered-backpropagation expansion $\Delta\tilde{\tilde{\tilde{\eta}}}_B(\mathbf{x})$ provides a way to make use of this information that avoids having to solve difficult and time consuming forward problems.

The estimates described above are incremental estimates of small-scale medium variations within a high-contrast background of large-scale objects. However, the monochromatic expressions $\Delta\tilde{\eta}_B(\mathbf{x})$, $\Delta\tilde{\tilde{\eta}}_B(\mathbf{x})$, and $\Delta\tilde{\tilde{\tilde{\eta}}}_B(\mathbf{x})$ are not effective for an iterative refinement of the background because small changes in the medium properties of large-scale objects can cause large errors in Born estimates as a result of phase deviations that accumulate in the incident and scattered waves as they propagate through the background. This difficulty can be overcome by forming estimates at multiple frequencies and obtaining phase corrections from the slopes of the phase at each reconstructed point with respect to frequency as described in Ref. 7. Use of these phase corrections led to a rapid convergence of the background parameters in Ref. 7. Another possibility that does not involve reconstruction is to select the background medium parameters to give the best least-squares fit of Eq. (42) to the measured far-field pattern. However, this procedure may find a local minimum if the initial guess is not sufficiently accurate.

The examples and discussion in this study have focused on an assumed background consisting of a homogeneous body, but the perturbed-phase filtered-backpropagation expansion is not limited to this application. In fact, better performance is expected from backgrounds that do not have abrupt discontinuities. The results presented here could be extended to include the use of assumed backgrounds that consist of homogeneous bodies with tapered edges and also the use of iterations that incorporate large-scale interior structures in the assumed background.

V. CONCLUSIONS

Image reconstruction from scattering measurements has the potential of achieving a better resolution than conventional *b*-scan images, but inverse scattering is an iterative process that can have high computational cost. The studies presented in this paper demonstrate that Born iterations can, under the right circumstances, be accomplished using an efficient variation of the filtered-backpropagation formula that does not require a solution to forward scattering problems or an inversion of large matrices. The conditions to which this approach is applicable are local. High-quality reconstruction can be realized in areas of the scattering volume where the fields of propagated plane waves are accurately approximated by phase perturbations. These requirements are satisfied, for example, when small medium variations are embedded in large-scale high-contrast background objects. The specific application that motivated this study is breast imaging by a ring transducer system, but the technique may also apply to other medical imaging applications that involve a resolution of detail within large anatomical features.

ACKNOWLEDGMENTS

David P. Duncan, Jin Jing, and Jason C. Tillett are thanked for their participation in discussions surrounding this study and also for their support in the development of the processing software. Michele M. Foster is thanked for her help in translating this article to a TeX format. This work was partially supported by NIH Grants Nos. HL 50855, CA 74050, and EB 00280 and by the University of Rochester Diagnostic Ultrasound Research Laboratory Industrial Associates.

APPENDIX

The approximate field given by Eq. (34) for plane-wave propagation through a cylinder with constant parameters is compared below with the exact field determined analytically. Let k_1 be the spatial-frequency magnitude that results from the medium parameters inside the cylinder, let k be the external spatial-frequency magnitude, and let the plane wave $u_i(\mathbf{x})=e^{jkr\cos\theta}$ be the incident field that is applied to the cylinder. An internal field $u_1(\mathbf{x})$ and a scattered field $u_s(\mathbf{x})$ are sought to satisfy

$$\Delta u_1(\mathbf{x}) + k_1^2 u_1(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega,$$

$$\Delta u_s(\mathbf{x}) + k^2 u_s(\mathbf{x}) = 0, \quad \mathbf{x} \in R^2 - \bar{\Omega} \text{ and } u_s \text{ radiating}, \tag{A1}$$

and also to satisfy the boundary conditions

$$\left.\begin{aligned} u_1(\mathbf{x}) &= u_s(\mathbf{x}) + u_i(\mathbf{x}) \\ (k_1/k)\partial_n u_1(\mathbf{x}) &= \partial_n u_s(\mathbf{x}) + \partial_n u_i(\mathbf{x}) \end{aligned}\right\} \quad \mathbf{x} \in \partial\Omega. \tag{A2}$$

Expansions that comply with these conditions are

$$u_i(r,\theta) = \sum_{n=-\infty}^{\infty} A_n J_n(kr) e^{jn\theta},$$

$$\partial_n u_i(r,\theta) = k \sum_{n=-\infty}^{\infty} A_n J_n'(kr) e^{jn\theta},$$

$$u_s(r,\theta) = \sum_{n=-\infty}^{\infty} D_n H_n^{(1)}(kr) e^{jn\theta},$$

$$\partial_n u_s(r,\theta) = k \sum_{n=-\infty}^{\infty} D_n H_n^{(1)\prime}(kr) e^{jn\theta},$$

$$u_1(r,\theta) = \sum_{n=-\infty}^{\infty} B_n J_n(k_1 r) e^{jn\theta}, \tag{A3}$$

$$\partial_n u_1(r,\theta) = k_1 \sum_{n=-\infty}^{\infty} B_n J_n'(k_1 r) e^{jn\theta}.$$

The assignments $A_n=j^n$ follow from the expansion

$$e^{jkr\cos\theta} = \sum_{n=-\infty}^{\infty} j^n e^{jn\theta} J_n(kr), \tag{A4}$$

and then values for the B_n and D_n coefficients follow from the application of the boundary conditions to the expansions in Eq. (A3), as described in Ref. 14.

The plots of the magnitude and phase of the interior field $u_1(\mathbf{x})$ are shown in Figs. 6(a) and 6(b). The radius of the cylinder for these computations is 24 mm, and the temporal frequency is 2.5 MHz. The sound speed inside the cylinder is 1.574 mm/μs, while the external sound speed is 1.509 mm/μs. The magnitude of the internal field is essentially 1 throughout most areas of the scattering volume and agrees with the magnitude of the phase-perturbed field given by Eq. (33). However, two small areas of significant fluctuations exist near the edges of the cylinder grazed by the incident field, and two crescent-shaped regions of smaller fluctuations exist near the far edge of the cylinder. These smaller fluctuations are attributed to the interference of the forward propagating field with reflection from the far edge. Both effects would diminish if the edges of the cylinder were smooth rather than abrupt.

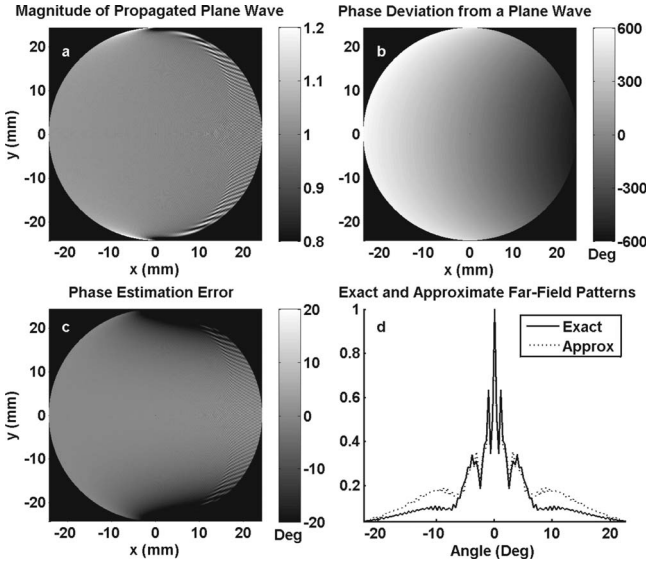


FIG. 6. (a) Magnitude of a unit-amplitude plane wave propagating through a homogeneous cylinder. (b) Phase deviation of a plane wave propagating through a cylinder from a plane wave propagating homogeneously. (c) Error in estimating the phase of a plane wave propagating through a cylinder using an acoustic path length along rays. (d) Comparison of the exact far-field pattern of the cylinder with the estimated pattern given by Eq. (41).

The plot in Fig. 6(c) is the residual phase fluctuation that remains after the phase perturbation of Eq. (34) has been deducted from the phase of $u_1(\mathbf{x})$ and demonstrates the close agreement in phase between the exact and approximate fields.

Figure 6(d) compares the far-field pattern of the cylinder obtained from the radial limit of the expansion for $u_s(\mathbf{x})$ given in Eq. (A3) with the approximate far-field pattern given in Eq. (41), which employs perturbed-phase field estimates. The absolute values of the patterns are normalized by the peak value of the pattern from the (A3) expansion and are plotted over a range of angles that extend 20° on either side of the forward-scattering direction.

[a]For strongly attenuating background potentials, the use of reciprocal fields rather than conjugate fields for the expansion in Eq. (19) can yield better results. The subsequent analysis applies to expansions of reciprocal fields without significant alteration.

[b]During the review process, the authors learned that the reconstruction given by Eq. (37) has been published in a somewhat different form in Ref. 15 where the expansion is a time-domain reconstruction in which phase perturbations appear as time delays. The perspective used in Ref. 15 is easily translated to the monochromatic case that is considered here. From the view of Ref. 15, the sums in Eq. (37) combine plane-wave transmitted fields and receiver sensitivities to form a single compound measurement. The exponential terms act as geometric phase factors that focus the measurement at point $\mathbf{x}$. The phase perturbations that appear in the modified expansion given in Eq. (37) are focusing corrections that

compensate for the aberration caused by propagation through the assumed background.

[1]A. I. Nachman, "Reconstructions from boundary measurements," Ann. Math. **128**, 531–576 (1988).
[2]A. I. Nachman, "Global uniqueness for a two-dimensional inverse boundary problem," Ann. Math. **143**, 71–96 (1996).
[3]C. Lu, J. Lin, W. Chew, and G. Otto, "Image reconstruction with acoustic measurement using distorted Born iteration method," Ultrason. Imaging **18**, 140–156 (1996).
[4]T. J. Cavicchi and W. D. O'Brien, "Numerical study of higher-order diffraction tomography via the sinc basis moment method," Ultrason. Imaging **11**, 42–74 (1989).
[5]T. D. Mast, A. I. Nachman, and R. C. Waag, "Focusing and imaging using eigenfunctions of the scattering operator," J. Acoust. Soc. Am. **102**, 715–725 (1997).
[6]F. Lin, A. I. Nachman, and R. C. Waag, "Quantitative imaging using a time-domain eigenfunction method," J. Acoust. Soc. Am. **108**, 899–912 (2000).
[7]R. C. Waag, F. Lin, T. K. Varslot, and J. P. Astheimer, "An eigenfunction method for reconstruction of large-scale and high-contrast objects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1316–1332 (2007).
[8]J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972), p. 270.
[9]K. Chadan, D. Colton, L. Päivärinta, and W. Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems* (SIAM, Philadelphia, 1997), p. 35.
[10]R. C. Waag and R. J. Fedewa, "A ring transducer system for medical ultrasound research," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1707–1718 (2006).
[11]A. J. Devaney and M. L. Oristaglio, "Inversion procedure for inverse scattering within the distorted-wave Born approximation," Phys. Rev. Lett. **51**, 237–240 (1983).
[12]A. J. Devaney, "A filtered back-propagation algorithm for diffraction tomography," Ultrason. Imaging **4**, 336–350 (1982).
[13]J. Jin, J. P. Astheimer, and R. C. Waag, "Estimation of scattering object characteristics for image reconstruction by using a nonzero background," Report No. DURL 08-1, Diagnostic Ultrasound Research Laboratory, University of Rochester, Rochester, NY, January 2008.
[14]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8.
[15]T. D. Mast, "Aberration correction for time-domain ultrasound diffraction tomography," J. Acoust. Soc. Am. **112**, 55–64 (2002).

Modeling the impact of soft tissue on axial transmission measurements of ultrasonic guided waves in human radius

Petro Moilanen[a)]
Department of Physics, University of Jyväskylä, Jyväskylä FI-40014, Finland

Maryline Talmant
Laboratoire d'Imagerie Paramétrique UMR7623, Université Pierre et Marie Curie - Paris 6, Paris F-75005, France and CNRS UMR7623, Paris F-75005, France

Vantte Kilappa, Patrick Nicholson, Sulin Cheng, and Jussi Timonen
Department of Physics and Department of Health Sciences, University of Jyväskylä, Jyväskylä FI-40014, Finland

Pascal Laugier
Laboratoire d'Imagerie Paramétrique UMR7623, Université Pierre et Marie Curie-Paris 6, Paris F-75005, France and CNRS UMR7623, Paris F-75005, France

(Received 26 February 2008; revised 18 July 2008; accepted 21 July 2008)

Recent *in vitro* and simulation studies have shown that guided waves measured at low ultrasound frequencies (f=200 kHz) can characterize both material properties and geometry of the cortical bone wall. In particular, a method for an accurate cortical thickness estimation from ultrasound velocity data has been presented. The clinical application remains, however, a challenge as the impact of a layer of soft tissue on top of the bone is not yet well established, and this layer is expected to affect the dispersion and relative intensities of guided modes. The present study is focused on the theoretical modeling of the impact of an overlying soft tissue. A semianalytical method and finite-difference time domain simulations were used. The models developed were shown to predict consistently real *in vivo* data on human radii. As a conclusion, clinical guided wave data are not consistent with *in vitro* data or related *in vitro* models, but use of an adequate *in vivo* model, such as the one introduced here, is necessary. A theoretical model that accounts for the impact of an overlying soft tissue could thus be used in clinical applications.
 [DOI: 10.1121/1.2973228]

PACS number(s): 43.80.Ev, 43.80.Vj, 43.80.Jz, 43.80.Qf [CCC] Pages: 2364–2373

I. INTRODUCTION

Interest in developing novel methods of noninvasive bone characterization has increased in recent years because osteoporosis has become a significant health issue, but it remains a largely underdiagnosed disease due to the lack of methods suitable for everyday clinical practice. The standard bone densitometry dual-energy x-ray absorption is expensive and nonportable and cannot accurately identify individuals who will later sustain a fracture (Bolotin and Sievänen, 2001; Stone *et al.*, 2003). Part of this inaccuracy can be explained by the lack of sensitivity to bone treats other than mass and geometry in a two-dimensional (2D) projection mode. On the contrary, quantitative ultrasound (QUS) enables manufacturing of inexpensive and convenient devices suitable for screening the fracture risk. While calcaneal QUS is already being used in clinical practice, new methods are being developed to enable multisite QUS assessment (Glüer, 2007; Laugier, 2006; Siffert and Kaufman, 2007). In this context, axial transmission is one among the novel promising techniques under development for assessing long cortical bones, such as the radius and tibia (Barkmann *et al.*, 2000; Njeh *et al.*, 2001; Sievänen *et al.*, 2001).

It has been shown that by using the axial transmission method, several modes of guided waves can be excited and detected in bones (Nicholson *et al.*, 2002; Bossy, 2002; Lefebvre *et al.*, 2002; Bossy, 2004a, 2004b; Lee and Yoon, 2004, Protopappas *et al.*, 2006; Moilanen *et al.*, 2007a), each mode having individualized characteristic sensitivities to different bone properties such as the cortical thickness (CTh), Young's modulus, and bone mineral density. It has thus been suggested that a multiparametric measurement could yield characterization of both geometry and material properties of bone (Muller *et al.*, 2005; Tatarinov *et al.*, 2005; Moilanen, 2008).

We focus here on CTh estimation based on the fundamental flexural type of guided wave, as reported previously in a series of *in vitro* and simulation studies on radius (Moilanen *et al.*, 2007a, 2007b, 2007c). It was shown that the phase velocity measured with a low-frequency axial transmission setup (f_c=200 kHz) is consistent with that of the fundamental flexural mode (F_{11}) of an axisymmetric solid tube, a mode with strong characteristic sensitivity to the wall thickness of the waveguide. Consequently, fitting the F_{11}

a)Electronic mail: pemoilan@cc.jyu.fi

tube model, thickness being the fitting parameter, with the recorded phase velocity yielded an accurate ultrasonic estimate (UTh) for the cortical bone thickness.

A previous *in vivo* study on the tibia of growing children (Moilanen *et al.*, 2003) suggested that measurement of a signal contribution consistent with the fundamental flexural mode is possible also in clinical measurements. However, later *in vivo* studies on elderly subjects (unpublished, University of Jyväskylä) have highlighted challenges in the identification of the fundamental flexural mode, and this difficulty has been associated with the presence of soft tissues on top of the bone. This could have occurred due to a thicker layer of overlying soft tissue in the older subjects and the impact of aging on elastic and absorptive properties of the soft tissue.

In order to model the impact of overlying soft tissue on *in vivo* guided wave measurements in bone (Dodd *et al.*, 2006), we consider a fluid-solid bilayer model. The suitability of the 2D model of Yapura and Kinra (1995) on bone was considered in a phantom study of the authors (Moilanen *et al.*, 2006). However, on further investigations on real *in vivo* signals, we observed that the 2D bilayer model did not fit well with *in vivo* experiments (unpublished results). Therefore we expect that the three-dimensional (3D) geometry must be accounted for, and we propose that one should consider the *in vivo* waveguide system as a traction-free fluid-coated tube. Most papers related to coated waveguides consider viscoelastic coating [see, e.g., Barshinger and Rose (2004)]. We do neglect here absorption in the media constituting the bilayered waveguide and leave for further study the analysis of this additional parameter.

The objectives of the present paper are thus to model the impact of the thickness of overlying soft tissue on guided wave phase velocities and the relative amplitudes in three dimensions, focusing on the slow energetic signal contribution (we refer to it as Wave2) typically associated with the fundamental flexural mode. To this end, methods of semianalytical (Sec. II) and numerical modeling (Sec. III) were used. In addition, *in vivo* measurements were mimicked by numerical simulation, and the model developed was also compared to real *in vivo* data on human radii (Sec. IV).

II. FLUID-SOLID BILAYER TUBE MODEL

The impact of a fluid overlayer of a given thickness on the phase velocity dispersion was modeled analytically in three dimensions by a hollow axisymmetric elastic cylinder (tube) coated with a thin fluid shell. This system attempted to model respectively the long bone and a layer of soft tissue on top of (or around) it. The model used is a direct extension of the 2D fluid-solid bilayer plate model presented by Yapura and Kinra (1995). The well known guided wave theory for tubes was first presented by Gazis (1959) and can be found in several textbooks. For introducing the fluid shell, we used a multilayer tube solution of Pavlakovic (1998). The characteristic equation used is given in the Appendix.

A compression velocity of $c_F=1500$ m/s was used for the ideal fluid. The bone material was considered isotropic with compression and shear velocities of $c_L=4000$ m/s and

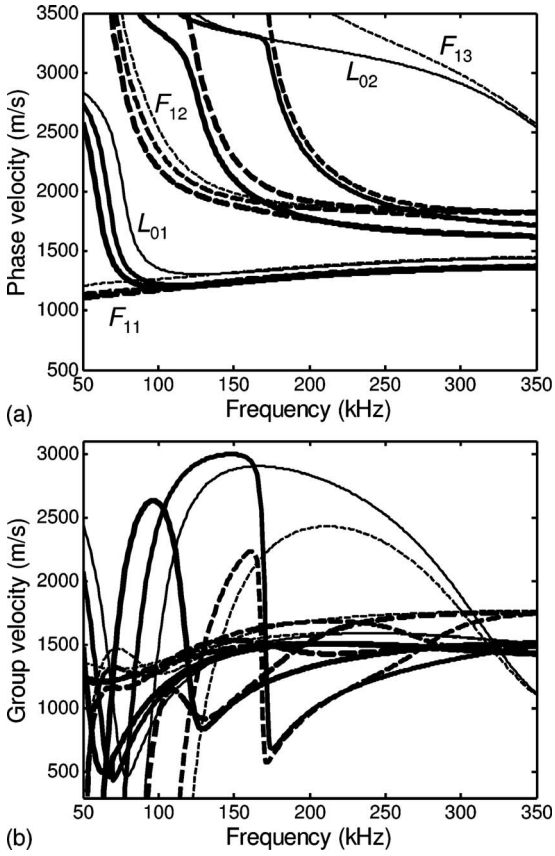


FIG. 1. (a) Phase velocity of fluid-solid bilayer tube modes (solid lines: longitudinal modes L_{01} and L_{02}; dashed lines: flexural modes F_{11}, F_{12}, and F_{13}) for three different soft overlayer thicknesses, $s=1$ mm (thin lines), 3 mm (intermediate lines), and 5 mm (thick lines). The solid tube core had a fixed wall thickness of $e=3$ mm and an outer radius of curvature of $b=7$ mm. Curves are shown as functions of frequency within the range typical of low-frequency axial guided wave measurements in bones. (b) Corresponding group velocities are difficult to distinguish, which suggests that the observation of a given region of temporal signal (time-of-flight) may yield ambiguous mode identification.

$c_T=1800$ m/s, respectively. The ratio of the fluid over the solid density used was $\tau=\rho_{\text{fluid}}/\rho_{\text{solid}}=0.54$. The wall thickness ($e=3$ mm) and outer radius of curvature ($b=7$ mm) of the solid tube were fixed according to average dimensions of the human radius bone, and the thickness (s) of the fluid coating was varied to investigate its effect.

Phase and group velocities of fluid-solid bilayer tube modes were computed as functions of frequency for various values of thickness of the soft coating layer (Fig. 1). The labeling scheme of the free solid tube was adopted for longitudinal (L_{0i}) and flexural (F_{1i}) modes, with $i=1,2,3,\ldots$, even though the corresponding branches of a bilayer tube may evolve differently from those of a free tube. Alterations in the thickness of the fluid shell affect differently the phase velocities of different modes [Fig. 1(a)]. Within the frequency range of Fig. 1, for certain modes (such as L_{01}, F_{11}, and F_{12}) the impact is relatively small, although clearly visible, whereas some other modes (such as L_{02} and F_{13}) are profoundly affected by the overlayer thickness. The thickness of soft coating thus affects the density of modes (number of modes) possible for a given frequency. On the other

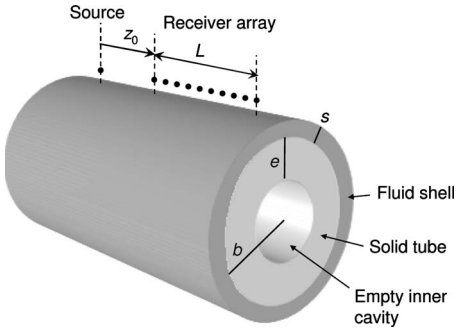


FIG. 2. Geometry of the bilayer tube system in numerical simulations: a solid empty tube (thickness e and outer radius of curvature b) surrounded with a fluid shell of thickness s. Positions of a point source and receivers are shown.

hand, group-velocity dispersion curves [Fig. 1(b)] suggest that the main parts of most of the possible modes arrive almost simultaneously at the receiver.

Phase and group velocities, characteristic of a fluid overlayer of a given thickness on top of the bone, should thus be accounted for when modes are identified in measured signals. To overcome the difficulty with the identification of modes, which overlap or are close to each other when signals are observed, further information is necessary. To this end, details on relative amplitudes of the different modes possible on top of the soft coating must also be modeled.

Yapura and Kinra (1995) analyzed displacement profiles of modes for the 2D bilayer system, and although we did not have corresponding results available for the 3D geometry, we could expect that it is challenging to measure the first flexural mode (also) in the 3D bilayer tube system.

Relative intensities of the tube bilayer modes (as detected at the receiver on top of the coating) were evaluated by means of numerical simulation (Sec. III).

III. SIMULATIONS ON A TUBULAR FLUID-SOLID BILAYER WAVEGUIDE

The question of how the soft overlayer of a given thickness on top of the bone affects the relative intensities of different guided modes was not answered by the analytical modeling in Sec. II. This, however, is a relevant question in the identification of modes and will be addressed by numerical modeling in the following.

A. Simulation methods

Finite-difference time domain method (FDTD) was used to reconstruct the response of a transient impact of a waveguide. Simulations were accomplished by a custom-made code (SIMSONIC) developed previously in LIP (Bossy *et al.*, 2004a) and based on the so-called Virieux scheme for discretization of linear elastic wave equations (Virieux, 1986; Graves, 1996). In addition, absorbing boundaries [perfectly matched layers (PMLs)] are implemented in order to eliminate spurious reflections at the boundaries of the simulation box. The grid step of $h=0.1$ mm was used in the simulations, satisfying the condition $h<\lambda_{\min}/15$.

The geometry of simulations is illustrated in Fig. 2. A point source and an array of point receivers ($z_0=20$ mm, $L=90$ mm, and intertransducer distance of 0.7 mm) were placed on top of the waveguide. The source was excited by a Gaussian-distributed 200 kHz pulse, with a −16 dB bandwidth of 50–350 kHz. The receiver array was called the extended array as distinguished from the shorter array of receivers used in the physical *in vitro* measurements (Sec. IV). Such an extended array enabled an enhanced reconstruction of the response of a transient excitation as compared to a short array.

The bone material was considered isotropic and the fluid shell an ideal fluid. The material and geometrical parameters used in the simulations were consistent with those used in analytical modeling (Sec. II). The simulation volume was rectangular in the FDTD simulations (volume boundaries not illustrated in Fig. 2). Thus the excess spaces around and inside the bilayer tube system were filled with a light ambient fluid (with its acoustic properties and density much lower than those of the waveguide) to mimic the air. The volume boundaries were conditioned by a perfectly matched layer (PML). See Moilanen *et al.* (2007b) for further details on a description of the general simulation parameters used also in this study.

The response signal (particle velocity in the radial direction) recorded at the receivers was analyzed in two dimensions as a function of time and source-receiver distance. Two-dimensional fast Fourier transform (2D-FFT) (Alleyne and Cawley, 1991) was used to determine the phase velocity/frequency spectra, and Hanning windowing was used to reduce Fourier artifacts.

B. Results

Phase velocity/frequency spectra [Figs. 3(a) and 3(c)] show that a significant portion of the excited energy was coupled to the F_{11} mode when the soft overlayer was thin (e.g., $s=2$ mm), but it tended to couple to other higher modes when the overlayer was thick (e.g., $s=6$ mm). These figures also show [consistent with Fig. 1(a)] the impact of overlayer thickness on superimposed dispersion curves; i.e., the density of modes appearing within a given range of frequency depends strongly on the value of parameter s.

Figures 3(b) and 3(d) represent corresponding phase velocity spectra at $f=200$ kHz and illustrate the relative intensities of the modes. It is evident that the extended array did not yield a sufficient resolution to completely distinguish between all propagating modes, but a yet longer array would have been needed. Extending the array length was not, however, possible because the consequent extension of the simulation volume would have exceeded the limit of the random access memory (RAM) available (2 Gbytes). The increasing thickness of the overlayer decreased the ability to identify modes. The results for a reduced array length, also shown in Fig. 3, will be discussed in more detail in Sec. IV.

The intensity of a 2D spectrum [such as the ones in Figs. 3(a) and 3(c)] was determined along the locus of the F_{11} dispersion curve (semianalytical result). Thus a simulated intensity consistent with that of the F_{11} mode was obtained. Figure 4 shows the resulting maximum intensity of a simulated F_{11} mode in the 50–350 kHz frequency range as a

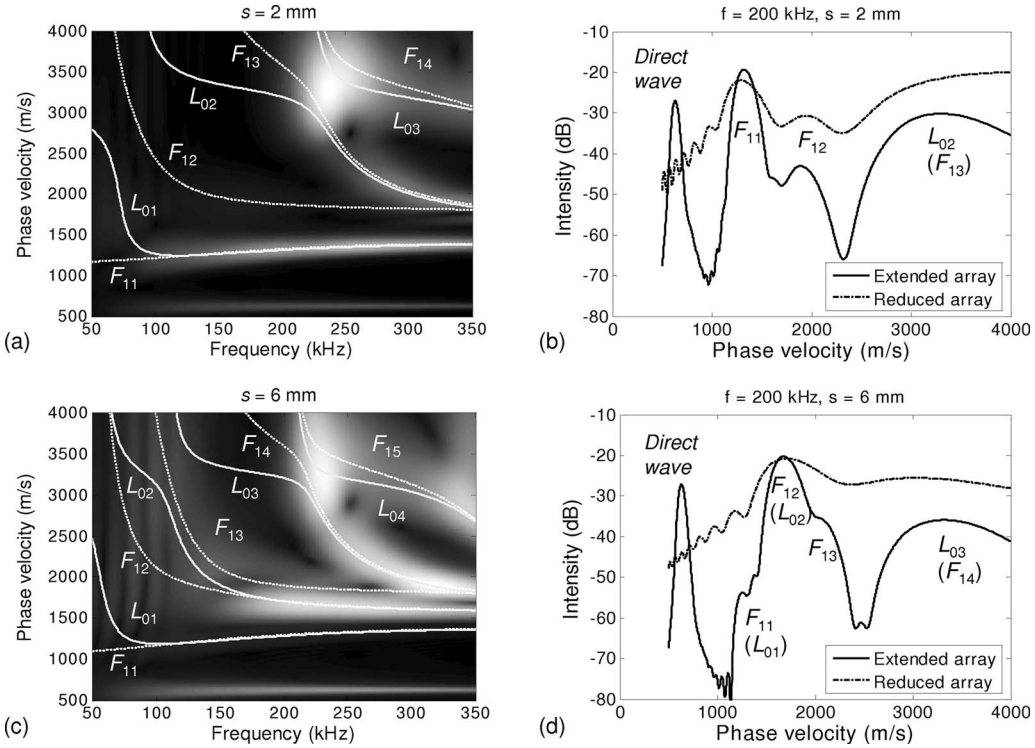


FIG. 3. 2D-FFT as functions of phase velocity and frequency for signals recorded on fluid-solid bilayer tubes with fluid overlayer thicknesses of $s=2$ and 6 mm ($e=3$ mm, $b=7$ mm) [(a) and (c)]. Loci of intensity maxima (bright intensities) represent recorded modes, and curves illustrate the corresponding semianalytical solutions (longitudinal modes are shown by solid lines and first order flexural modes are represented by dashed lines). Relative intensities of the modes are illustrated by phase velocity spectra [(b) and (d)], extracted respectively from the phase velocity/frequency spectra at $f=200$ kHz [solid lines at (b) and (d)]. Corresponding phase velocity spectra by using a reduced array are shown for comparison (dashed lines).

function of overlayer thickness. It can be seen that the intensity of F_{11} decreases rapidly and linearly (-7.6 ± 0.8 dB/mm) with increasing overlayer thickness.

C. Discussion

The method of thickness evaluation reported in our previous papers was based on the F_{11} mode. It is shown here, however, by the simulation results that the intensity of F_{11} seems to decrease rapidly when the thickness of the fluid shell is increased, suggesting that the normal displacement associated with F_{11} decreases. Thus, F_{11} can only be observed (i.e., distinguished from other modes and noise) when the fluid shell is very thin $s<3$ mm. Therefore, thickness evaluation based on F_{11} is expected not to be possible for the majority of subjects (see Table I and Sec. IV for a typical soft tissue thickness (STh) on top of human radius).

It was observed that other modes with higher phase velocities than that of F_{11} take over when the thickness of the fluid shell is increased. Thus, in principle, the suitability of such higher order modes for developing a method of *in vivo* thickness evaluation should be investigated instead of using a method based on F_{11}.

Another complication of the *in vivo* problem predicted by this simulation model is the increasing density of modes as a function of increasing soft overlayer thickness. Greater mode density as compared to the *in vitro* system makes the identification of modes challenging. This concerns the higher order modes in particular.

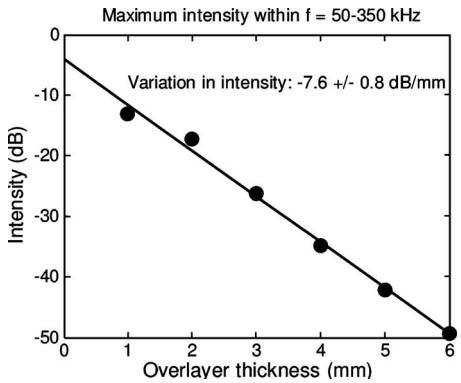


FIG. 4. Intensity of the measured F_{11} mode as a function of soft overlayer thickness ($e=3$ mm and $b=7$ mm). The maximum intensity of the recorded 2D spectra was read along the loci of semianalytical solution for the F_{11} mode. Result for an extended array.

TABLE I. Descriptive statistics of *in vivo* ultrasound and computed tomography data for a subset of "Wave2 velocity consistent with F_{12} mode velocity" ($n=87$).

	Mean	SD	Range
V2 (m/s)	1864	91	1712–2064
PE_{V2} (m/s)	45	48	3–309
DE_{V2} (m/s)	48	38	3–153
CBMD (mg/cm^3)	1189	44	1051–1242
CTh (mm)	3.1	0.5	2–4
STh (mm)	11	3.2	6.3–21

Among the higher order modes, F_{12} has its phase velocity closest to that of F_{11}. In fact, it appears that the F_{12} phase velocity is close to that of the plate mode A_0 (results not shown here). It can thus be speculated that attempts to identify an A_0 like of mode (A_0 or F_{11}) in the past studies on *in vivo* signals may not have been able to distinguish it from the F_{12} mode. On the other hand the modeling results of the present study may also be characteristic of the geometry of the radius so that other bone sites, such as tibia, which have a thinner layer of soft tissue on top of a thicker wall of cortical bone, may behave differently.

IV. SIMULATED AND REAL *IN VIVO* MEASUREMENTS

We have described in the previous sections the impact of the thickness of a fluid coating layer on dispersion curves and relative amplitudes of guided waves in the bilayer tube. We will now compare the results of this model to those of *in vivo* measurements on human radii focusing on the slow energetic signal contribution, typically associated in the *in vitro* case with the fundamental flexural mode (F_{11}). We refer to the slow energetic signal contribution as Wave2, and its phase velocity is denoted as v_2, indicating a velocity determined from a recorded signal (experimental or simulated) as distinguished from c, which is used to represent bulk wave velocities.

A. Methods

1. Numerical simulations

Numerical simulations were made by using the FDTD method. The setup was identical to that described in Sec. III A, with an exception of the receiver array length. Here a reduced array (z_0=20 mm, L=28 mm, and an intertransducer distance of 0.7 mm) was used to mimic the response recorded by a typical physical array suitable for clinical measurements (see Sec. IV A 2).

2. In vivo study

(a) *Subject population.* The left arm radius was measured on 97 voluntary subjects participating in the calcium and vitamin *D* supplementation and intervention study (CALEX) at the Department of Health Sciences (University of Jyväskylä, Jyväskylä, Finland). The age range of the subjects was 22–83 yr, with a mean age (standard deviation) of 51 (±14) yr. The population consisted mostly of female subjects. Cross-sectional data corresponding to one follow-up measurement point were analyzed.

(b) *Ultrasonic assessments.* A low-frequency axial scanner [f_c=200 kHz; see Nicholson *et al.*, (2002)] was used to perform measurements at the posterolateral site, 1/3 from the distal end of the radius. Phase velocity v_2 of Wave2 was recorded. In this device a progressive scan of one receiver element was used to produce a spatial array of receiver points. The transducers were compressed against the skin by an approximately 300 g loading, which has been found essential for a proper coupling and creation of the slow energetic signal contribution (Wave2). The contact load was not released during the scan, and the skin and soft tissue were slightly dragged by the receiver, enabling slight compression of the tissue throughout the scanning region. The emitter was excited for each emitter-receiver distance separately by using a half-period square pulse of 2.5 μs length, which resulted in an impulse excitation at a center frequency of approximately 200 kHz. The signal was excited by turns at both ends of the pseudoreceiver array, and v_2 was defined as the harmonic mean of the unidirectional velocities as in Bossy *et al.* (2004b). The purpose of the bidirectional method of Bossy *et al.*. (2004b) is to account for varying STh along the scanning range. However, due to uncertainty in the repositioning of the unidirectional measurements, we did not obtain here a pure bidirectional correction. Nevertheless, repeating the measurement in two directions enabled an estimation of the uncertainty related to the scanning direction. To this end, the unidirectional velocities were compared with the v_2 velocity determined, and this comparison gave a root-mean-square direction error (DE_{V2}) for the unidirectional components of v_2. The precision error PE_{V2} of v_2 was evaluated as a root-mean-square standard deviation (RMS SD) of two repeated measurements (where one measurement is a measurement performed in both directions) for each subject.

(c) *Reference bone property assessments*. Peripheral quantitative computed tomography (XCT 2000, Stratec Medizintechnik GmbH, Pforzheim, Germany) was used to assess the mineral density and geometry of bones by a voxel size of 590 μm. One cross-sectional slice was recorded for each subject at the middle of the ultrasound scan. Cortical bone mineral density (CBMD) was determined from the cross-sectional image. In addition, circumferentially local CTh and STh were assessed manually from the image to correspond to the circumferential location where the ultrasonic measurements were made.

Computed tomography yielded an assessment for the unloaded STh, whereas in ultrasonic measurements the soft tissue was compressed by an approximately 300 g loading. An assumption was made that the compressed thickness was 1/3 STh.

3. Signal analysis

2D-FFT (Alleyne and Cawley, 1991) was used to determine the phase velocity v_2 of Wave2. Prior to 2D-FFT, a method called group-velocity filtering was used to isolate Wave2 and to decrease the interference of other components (Moilanen *et al.*, 2006). Group-velocity filtering is based on Hanning windowing of the temporal signal and includes two parameters, width and temporal position. The half-amplitude width of the Hanning window was fixed to a value of 5 μs, which is around one period of the associated center frequency of excitation. The temporal position can be determined from the supposed propagation path and corresponds to the alignment of the window, with Wave2 observed in distance-time diagrams [shown by dashed lines in Figs. 5(a) and 6(a)]. The shift in temporal position at two successive

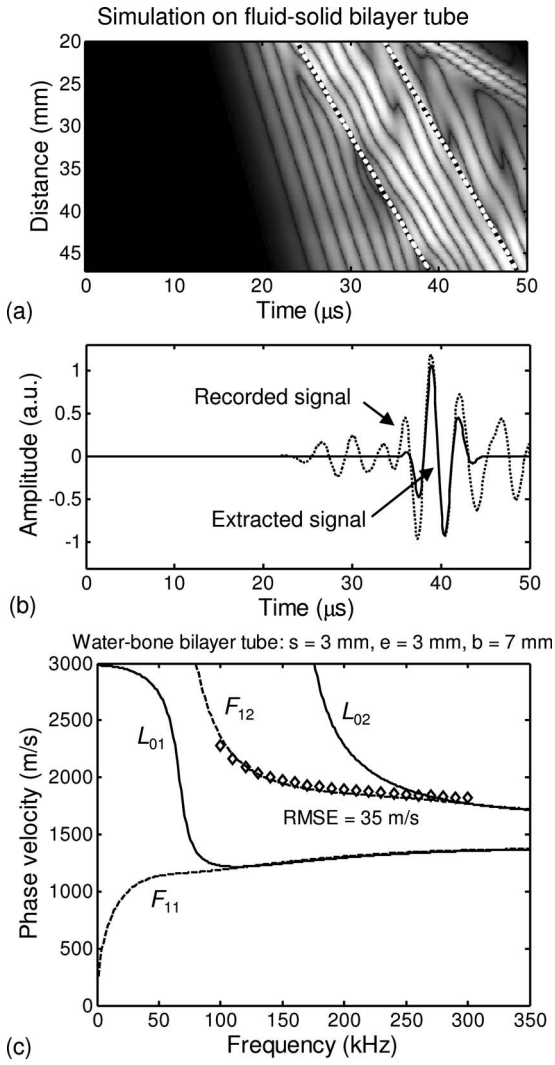


FIG. 5. Simulation results for a fluid-solid bilayer tube (s=3 mm and e =3 mm): (a) rt diagram with range markers showing a region isolated by group-velocity filtering and (b) recorded (dashed line) and group-velocity-filtered (solid line) signals. (c) Phase velocity v_2 (diamond markers) and phase velocities of bilayer modes L_{01}, L_{02}, F_{11}, and F_{12} as functions of frequency. The RMSE between v_2 and F_{12} phase velocity is shown.

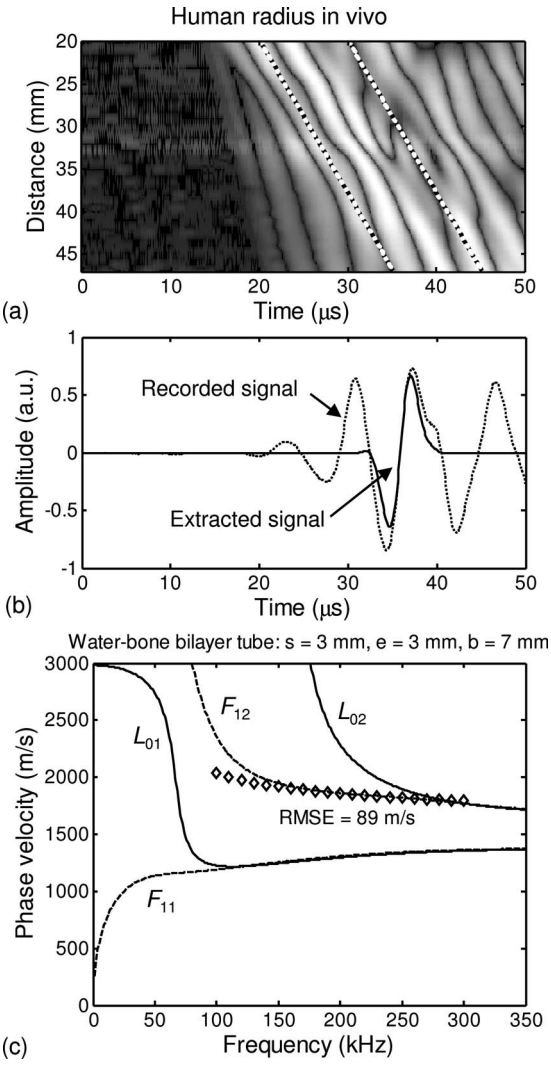


FIG. 6. *In vivo* results for human radius (STh=10 mm and CTh=3 mm): (a) rt diagram with range markers showing a region isolated by group-velocity filtering and (b) recorded (dashed line) and group-velocity-filtered (solid line) signals at a source-receiver distance of 40 mm. (c) Phase velocity v_2 (diamond markers) and phase velocities of bilayer modes L_{01}, L_{02}, F_{11}, and F_{12} as functions of frequency. Compressed soft tissue thickness (i.e., the parameter value to compute the semianalytical results) was approximated by $s=1/3$ STh$\approx$3 mm.

receivers is associated with the group velocity of Wave2. In practice a fixed value of 1800 m/s was used for the group velocity. The absolute value of the temporal position was, in addition to the ratio of source-receiver distance and group velocity, determined by an additional delay parameter due to passing through the overlying soft tissue and was determined manually so as to align the filter window with Wave2. Examples of proper alignment of the filter on Wave2 are shown in Figs. 5(a) and 6(a).

The phase velocity $v_2(f)$ of the windowed signal was determined as a function of frequency f by using the 2D-FFT within the range of f=50–350 kHz. Velocity $v_2=v_2(f$ =200 kHz) was used consistently with the excitation frequency to represent the recorded (simulated or experimental) phase velocity if the frequency is not explicitly specified.

B. Results

1. Analysis of recorded signals

Simulation results in Figs. 3(b) and 3(d) show the adverse impact of the reduction in the receiver array length to 3 cm. The signals shown for the reduced array were processed with group-velocity filtering, which has been found to be a good compromise between the peak resolution and the intensity of Fourier artifacts and noise. It appears that by using this best method, which is suitable and as yet available to us in clinical measurements, one can really identify the F_{11} mode only through a very thin soft tissue (of the order of a few millimeters). It will be seen below that an unquestionable identification of any other mode is challenging as well.

Distance-time (rt) diagrams recorded at the receiver array and the corresponding ultrasonic radio-frequency (rf) signals for one source-receiver distance are shown for a simulation in Figs. 5(a) and 5(b) and for a real *in vivo* measurement in Figs. 6(a) and 6(b). The overlayer thickness was (approximately) 3 mm. It can be observed that the *in vivo* signal had a lower fundamental frequency than the simulated signal despite the similar frequency of excitation

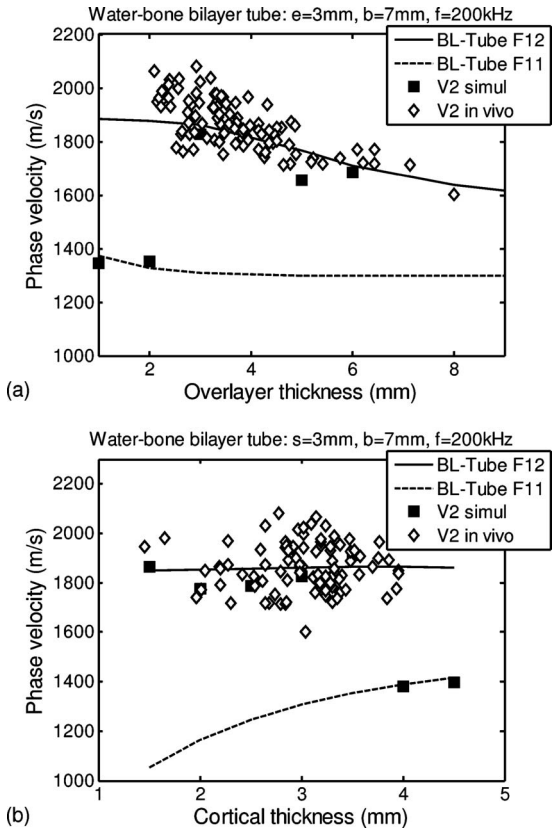


FIG. 7. Simulated v_2 (solid square markers) and *in vivo* v_2 (diamond markers) across simulated samples and a subset population of $n=87$ (a) as functions of overlayer thickness (fixed $e=3$ mm in simulations) and (b) as functions of CTh (fixed $s=3$ mm in simulations). The compressed (by ultrasonic probes) STh on *in vivo* measurements was evaluated by 1/3 STh. The analytical results for F_{11} (dashed line) and F_{12} (solid line) modes of a fluid-solid bilayer tube are shown for comparison.

used. This could be understood by frequency dependent absorption in bone and soft tissue, which influenced the *in vivo* measurements but was not modeled by the simulation code used.

Wave2 can be observed in simulations and *in vivo* measurements despite the presence of a fluid overlayer or soft tissue of typical thickness. A typical STh on the radius is STh=10 mm when measured as unloaded. When an ultrasonic probe compresses the soft tissue by a reasonable load, the measurement time thickness (as seen by ultrasound) could be approximated by 1/3 STh=10/3≈3 mm.

When the signals were analyzed by using group-velocity filtering and 2D-FFT within the phase velocity range of 500–3000 m/s, the phase velocities thus obtained were in good accordance with the F_{12} (i.e., the second flexural) mode both for simulations [Fig. 5(c)] and *in vivo* experiments [Fig. 6(c)]. The root-mean-square errors (RMSEs) between the recorded phase velocity and the F_{12} phase velocity for the above cases were 35 m/s (2%) in the simulation result and 89 m/s (5%) in the *in vivo* result.

2. Statistical analysis of recorded velocity data

A comparison between simulated v_2 (filled square markers) and phase velocities of the semianalytical model (lines curves) is illustrated in Fig. 7. Velocity v_2 was consistent with that of F_{11} only when the fluid shell was thin and the bone wall was thick. Otherwise v_2 was consistent with the F_{12} phase velocity. If the fluid shell was thick (e.g., $s=5$ mm), mode identification became ambiguous.

For a subset of 90% of the measured subjects ($n=87$) the *in vivo* v_2 (diamond markers) followed consistently the phase velocity of the F_{12} mode (Fig. 7). Descriptive statistics for the 90% subset are shown in Table I.

For the remaining 10% of the subjects, Wave2 could not be identified as F_{12}, and v_2 could rather be identified as the velocity of the F_{11} mode or a mode with a higher phase velocity than F_{12}. For most of these cases, mode identification was not consistent throughout the four repeated measurements; thus these data were discarded from further analysis.

For the 90% subset ($n=87$), the *in vivo* v_2 was correlated with STh ($r=-0.70$, $p<0.001$). It was not correlated with CTh but was weakly correlated with CBMD ($r=0.28$, $p<0.01$).

3. Discussion

One of the issues addressed here was related to what can be observed by using a transducer array of limited length in clinical measurements on a human radius. We showed by simulations that mode identification is then problematic for typical STHs. As a solution we suggest that methods of signal analysis should be developed so as to provide, at least, an unambiguous mode identification. Preferably such a mode (or modes) should be identified whose phase velocity is sensitive to the variability in bone thickness after compensation for the STh. This would lead to an *in vivo* revision of the thickness inversion scheme that we have introduced previously (Moilanen *et al.*, 2007a, 2007b, 2007c).

In the present work we did not, however, consider a strategy for estimating bone thickness *in vivo*. Instead, we showed by simulations that another mode, namely, F_{12}, was identified in the majority of cases with structural dimensions of a human radius. This mode cannot be likely used for bone thickness evaluation due to its low sensitivity to that thickness but preferably for the evaluation of soft tissue properties or, potentially, to the evaluation of the elastic properties of the bone wall. On the other hand, the simulation results of the present study predict that the fundamental flexural wave is observable in a bilayer tube of the lowest fluid overlayer thickness (2 mm) considered here. This suggests that the procedure of CTh determination, already tested in an *in vitro* case, could be extended to *in vivo* measurements on anatomic sites such as the tibia for which the overlying soft tissue is generally thinner than in the radius. This assumption is also supported by a study of *in vivo* measurements on the tibia of growing children (Moilanen *et al.*, 2003), already mentioned in Sec. 1. Furthermore, using optimized excitation and detection of the fundamental flexural mode could help in the determination of the *in vivo* thickness.

Another issue addressed was the practical relevance of the models introduced: their suitability to represent clinical measurements. It is evident that the *in vivo* results for radius demonstrate that an F_{12} mode can be identified consistently in a significant majority of the cases, very much in agree-

ment with the modeling results. We can thus conclude that the models used help to understand the challenges of *in vivo* measurements and that the fluid-solid bilayer tube model could be used to develop measurement strategies for *in vivo* situations.

One of the limitations related to the evaluation of the clinical applicability of the ultrasound measurement considered was the assumption of compressibility of soft tissue during an *in vivo* measurement. As such compression was not uniform throughout the measurement range, and we did not have an accurate estimation of it. We assumed that STh was compressed to 1/3 of its uncompressed value, as determined by computed tomography. Although this was a rough estimate, we assume that it was realistic enough as it yielded in general a good accordance between the recorded v_2 and phase velocity of F_{12}.

Another limitation with respect to the *in vivo* results was the exclusion of a minor subset from comparison due to ambiguous mode identification. We observed that the average age (standard deviation), 56 (12) yr, in the discarded subset was slightly higher than the average age of the whole data set and that the average STh of the discarded subjects, STh (standard deviation) =9.0 (2.8) mm, was smaller than the overall average. Thinner soft tissues thus did not make mode identification easier, as could have been expected. On the other hand, aging typically brings with it changes in the properties of soft tissues (elasticity, dehydration, etc.), which affect the coupling and propagation of ultrasonic signals. Future work should thus pay attention to this problem as these subjects have a high risk of having osteoporosis.

The reproducibility error of the v_2 measurement (for the 90% subset) was approximately 2.4% (PE_{V2}/v_2; Table I), which is substantially higher than the approximately 1% reported previously for *in vitro* measurements (Moilanen *et al.*, 2007a, 2007c). This is an obvious indication of the impact of an overlying soft tissue and poses a limitation to the technique used. The precision error PE_{V2} covered approximately 13% of the v_2 data range, which suggests that v_2 data can, regardless of the repeatability error, represent true variability in bone parameters.

The value of the directional error parameter (DE_{V2}; Table I) also poses a limitation to the techniques used. Namely, a particular limitation of the present study was that the ultrasound scanning technique did not completely compensate for this impact of variation in STh. Therefore, the DE_{V2} potentially explains part of the precision error PE_{V2}.

V. CONCLUSION

We demonstrated in this paper that a bilayered fluid-solid tube model provides useful insight into understanding the challenges of *in vivo* v_2 measurements and of a related CTh evaluation. The F_{11} mode, based on which the method of thickness evaluation was developed in previous *in vitro* and modeling studies (Moilanen *et al.*, 2007a, 2007c), is difficult to identify when the soft tissue on top of the bone is thicker than a few millimeters. What also happens is that mode density increases with increasing thickness of the soft tissue. This can cause ambiguity in mode identification in the *in vivo* measurements.

Despite the challenges related to the current *in vivo* v_2 measurements, we showed here for the first time that *in vivo* measurements can be modeled semianalytically and numerically. These models provide tools to further optimize the measurement setup and methods of signal analysis, so as to achieve an unambiguous identification of mode(s). Paying attention to proper signal coupling and to the investigation of more powerful methods of signal analysis than the FFT could provide further enhancements in the techniques.

Given that an improved measurement can provide unambiguous velocity data for a mode sensitive to CTh (such as F_{11}), the semianalytical bilayer model introduced here provides a way to compensate for the soft tissue effect and thus enables an estimation of cortical bone thickness also in clinical (*in vivo*) measurements. Compensation for soft tissue effects and evaluation of CTh could include (a) an independent assessment of the STh, e.g., by using pulse-echo ultrasound measurement, and fitting for CTh alone, or (b) a simultaneous fitting for soft tissue and CTh.

As the methods used so far (Nicholson *et al.*, 2002; Moilanen *et al.*, 2003) for detecting slow guided waves in *in vivo* low-frequency axial transmission measurements may not be suitable for CTh evaluation in a human radius (based on A_0 or F_{11}), further method development is called for to properly take into account the effects that arise from an overlying layer of soft tissue. The present method, being sensitive to the F_{12} mode when applied to a human radius, could be optimized so as to yield (a) a method to assess the quality of soft tissue and (b) a method to evaluate the material properties of a bone independent of the CTh. It should be noted, finally, that the present work was focused only on the human radius with its typical geometrical dimensions. An interpretation of v_2 measurements on other bone sites, such as the tibia, may differ significantly due to different relative thicknesses of bone and overlying soft tissue. We expect, however, that inclusion of the impact of an overlying soft tissue would also, in this case, improve the assessment of bone properties.

ACKNOWLEDGMENTS

This work was supported by the French Ministry of Education and the Academy of Finland (Project No 115621).

APPENDIX: CHARACTERISTIC EQUATION

The derivation of the characteristic equation used for an empty solid tube (compression bulk wave velocity c_L, shear bulk wave velocity c_T, and mass density ρ) coated by a fluid layer (compression bulk wave velocity c_F and mass density ρ_F) will be summarized in the following. Using cylindrical coordinates (r,θ,z), displacement in a solid tube can be expressed via the scalar potential,

$$\varphi = [A_1 J_n(\alpha r) + A_2 Y_n(\alpha r)] \cdot \cos(n\theta) \cdot e^{i(k_z z - \omega t)}, \qquad \text{(A1)}$$

and the vector potential,

TABLE II. Organization of matrix M.

	Solid tube						*Fluid coating*	
	A_1	B_1	C_1	A_2	B_1	C_1	D_1	D_2
$\sigma_{rr}\|r=a$							0	0
$\sigma_{rq}\|r=a$							0	0
$\sigma_{rz}\|r=a$			Terms associated with the characteristic equation of a free solid tube				0	0
$\sigma_{rr}\|r=b$							m_{47}	m_{48}
$\sigma_{rq}\|r=b$							0	0
$\sigma_{rz}\|r=b$							0	0
$u_r\|r=b$	m_{71}	m_{72}	m_{73}	m_{74}	m_{75}	m_{76}	m_{77}	m_{78}
$\sigma_{rr}\|r=R$	0	0	0	0	0	0	m_{87}	m_{88}

$$\Psi_r = [B_1 J_{n+1}(\beta r) + B_2 Y_{n+1}(\beta r)] \cdot \cos(n\theta) \cdot e^{i(k_z z - \omega t)},$$

$$\Psi_\theta = -[B_1 J_{n+1}(\beta r) + B_2 Y_{n+1}(\beta r)] \cdot \cos(n\theta) \cdot e^{i(k_z z - \omega t)},$$

$$\Psi_z = [C_1 J_n(\beta r) + C_2 Y_n(\beta r)] \cdot \sin(n\theta) \cdot e^{i(k_z z - \omega t)}. \quad \text{(A2)}$$

Correspondingly, the displacement in the fluid layer can be expressed via the scalar potential,

$$\varphi_F = [D_1 J_n(\alpha_F r) + D_2 Y_n(\alpha_F r)] \cdot \cos(n\theta) \cdot e^{i(k_z z - \omega t)}. \quad \text{(A3)}$$

In Eqs. (A1)–(A3) the symbol $\alpha^2 = w^2/c_L^2 - k^2$, $\beta^2 = w^2/c_T^2 - k^2$, $\alpha_F^2 = w^2/c_F^2 - k^2$, J_n, and Y_n are Bessel functions of order n.

The boundary conditions imposed in a free fluid-solid bilayer tube are

$$\sigma_{rr} = \sigma_{r\theta} = \sigma_{rz} = 0 \quad \text{at} \quad r = a,$$

$$u_r = u_{rF}, \quad \sigma_{rr} = \sigma_{rrF} \quad \text{and} \quad \sigma_{r\theta}\sigma_{rz} = 0 \quad \text{at} \quad r = b,$$

and

$$\sigma_{rrF} = 0 \quad \text{at} \quad r = R \quad \text{(A4)}$$

where a is the inner radius and b the external radius of the solid tube, and radius R represents that of the free surface of the coating fluid layer. These boundary conditions [Eq. (A4)] yield the characteristic equation formed by the determinant of coefficients $A_1, \ldots, D_2$

$$\mathrm{Det}(M) = 0, \quad \text{(A5)}$$

where the matrix M is organized according to Table II.

Elements m_{ij} $(i=7, 8; j=7, 8)$ of matrix M due to the coating fluid layer are

$$m_{71} = \alpha b J_n'(\alpha b), \quad m_{72} = k b J_{n+1}(\beta b), \quad m_{73} = n J_n(\beta b),$$

$$m_{74} = \alpha b Y_n'(\alpha b), \quad m_{75} = k b Y_{n+1}(\beta b), \quad m_{76} = n Y_n(\beta b),$$

$$m_{77} = -\alpha_F b J_n'(\alpha_F b), \quad m_{78} = -\alpha_F b Y_n'(\alpha_F b),$$

$$m_{47} = \frac{1}{2}\frac{\rho_F}{\rho} \cdot k_T^2 b^2 J_n(\alpha_F b), \quad m_{48} = \frac{1}{2}\frac{\rho_F}{\rho} \cdot k_T^2 b^2 Y_n(\alpha_F b),$$

$$m_{87} = \frac{1}{2}\frac{\rho_F}{\rho} \cdot k_T^2 R^2 J_n(\alpha_F R), \quad m_{88} = \frac{1}{2}\frac{\rho_F}{\rho} \cdot k_T^2 R^2 Y_n(\alpha_F R), \quad \text{(A6)}$$

where $k_T = w/c_T$. The matrix elements associated with the characteristic equation of a free solid tube are well known and can be found in various references (e.g., Gazis, 1959; Pavlakovic, 1998). Notice, however, that when implementing these terms from the literature, one may need to adjust the scaling. We used a scaling factor of $r^2/(2\ \mu)$ for all the stress terms.

Alleyne, D., and Cawley, P. (**1991**). "A two-dimensional Fourier transform method for the measurement of propagating multimode signals," J. Acoust. Soc. Am. **89**, 1159–1168.

Barkmann, R., Kantorovich, E., Singal, C., Hans, D., Genant, H. K., Heller, M., and Glüer, C. C. (**2000**). "A new method for quantitative ultrasound measurements at multiple skeletal sites: First results of precision and fracture discrimination," J. Clin. Dent. **3**, 1–7.

Barshinger, J. N., and Rose, J. L. (**2004**). "Guided wave propagation in an elastic hollow cylinder coated with a viscoelastic material," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1547–1556.

Bolotin, H. H., and Sievänen, H. (**2001**). "Inaccuracies inherent in dual-energy x-ray absorptiometry *in vivo* bone mineral density can seriously mislead diagnostic/prognostic interpretations of patient-specific bone fragility," J. Bone Miner. Res. **16**, 799–805.

Bossy, E., Talmant, M., and Laugier, P. (**2002**). "Effect of bone cortical thickness on velocity measurements using ultrasonic axial transmission: A 2D simulation study," J. Acoust. Soc. Am. **112**, 297–307.

Bossy, E., Talmant, M., Defontaine, M., Patat, F., and Laugier, P. (**2004b**). "Bidirectional axial transmission can improve accuracy and precision of ultrasonic velocity measurement in cortical bone: A validation on test materials," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 71–79.

Bossy, E., Talmant, M., and Laugier, P. (**2004a**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Dodd, S. P., Cunningham, J. L., Miles, A. W., Gheduzzi, S., and Humphrey, V. F. (**2006**). "Ultrasonic propagation in cortical bone mimics," Phys. Med. Biol. **51**, 4635–4647.

Gazis, D. C. (**1959**). "Three dimensional investigation of the propagation of waves in hollow circular cylinders I: Analytical foundation," J. Acoust. Soc. Am. **31**, 568–573.

Glüer, C. C. (**2007**). "Quantitative ultrasound—it is time to focus research efforts," Bone **40**, 9–13.

Graves, R. W. (**1996**). "Simulating seismic wave propagation in 3D elastic Media using staggered-grid finite differences," Bull. Seismol. Soc. Am. **86**, 1091–1106.

Laugier, P. (**2006**). "Quantitative ultrasound of bone: Looking ahead," Joint Bone Spine **73**, 125–128.

Lee, K. I., and Yoon, S. W. (**2004**). "Feasibility of bone assessment with leaky Lamb waves in bone phantoms and a bovine tibia," J. Acoust. Soc. Am. **115**, 3210–3217.

Lefebvre, F., Deblock, Y., Campistron, P., Ahite, D., and Fabre, J. J. (**2002**). "Development of a new ultrasonic technique for bone and biomaterials *in*

vitro characterization," J. Biomed. Mater. Res. **63**, 441–446.
Moilanen, P. (**2008**). "Ultrasonic guided waves in bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1277–1286.
Moilanen, P., Nicholson, P. H., Kärkkäinen, T., Wang, Q., Timonen, J., and Cheng, S. (**2003**). "Assessment of the tibia using ultrasonic guided waves in pubertal girls," Osteoporosis Int. **14**, 1020–1027.
Moilanen, P., Nicholson, P. H. F., Kilappa, V., Cheng, S., and Timonen, J. (**2006**). "Measuring guided waves in long bones: Modelling and experiments in free and immersed plates," Ultrasound Med. Biol. **32**, 709–719.
Moilanen, P., Nicholson, P. H. F., Kilappa, V., Cheng, S., and Timonen, J. (**2007a**). "Assessment of the cortical bone thickness using ultrasonic guided waves: Modelling and *in vitro* study," Ultrasound Med. Biol. **33**, 254–262.
Moilanen, P., Talmant, M., Bousson, V., Nicholson, P. H. F., Cheng, S., and Laugier, P. (**2007b**). "Ultrasonically determined thickness of long cortical bones: Two-dimensional simulations of *in vitro* experiments," J. Acoust. Soc. Am. **122**, 1818–1826.
Moilanen, P., Talmant, M., Nicholson, P. H. F., Cheng, S., and Laugier, P. (**2007c**). "Ultrasonically determined thickness of long cortical bones: Three-dimensional simulations of *in vitro* experiments," J. Acoust. Soc. Am. **122**, 2439–2445.
Muller, M., Moilanen, P., Bossy, E., Nicholson, P., Kilappa, V., Timonen, J., Talmant, M., Cheng, S., and Laugier, P. (**2005**). "Comparison of three ultrasonic axial transmission methods for bone assessment," Ultrasound Med. Biol. **31**, 633–642.
Nicholson, P. H. F., Moilanen, P., Kärkkäinen, T., Timonen, J., and Cheng, S. (**2002**). "Guided ultrasonic waves in long bones: Modelling, experiment and *in vivo* application," Physiol. Meas **23**, 755–768.
Njeh, C. F., Saeed, I., Grigorian, M., Kendler, D. L., Fan, B., Shepherd, J., McClung, M., Drake, W. M., and Genant, H. K. (**2001**). "Assessment of bone status using speed of sound at multiple anatomical sites," Ultrasound Med. Biol. **27**, 1337–1345.
Pavlakovic, B. N. (**1998**). "Leaky guided ultrasonic waves in NDT," Ph.D. thesis, Imperial College of Science, Technology, and Medicine, University of London, London; URL: http://www.imperial.ac.uk/ndt/public/publications/abstracts/pavlakovic.htm (Last viewed July 18, 2008).
Protopappas, V. C., Fotiadis, D. I., and Malizos, K. N. (**2006**). "Guided ultrasound wave propagation in intact and healing long bones," Ultrasound Med. Biol. **32**, 693–708.
Sievänen, H., Cheng, S., Ollikainen, S., and Uusi-Rasi, K. (**2001**). "Ultrasound velocity and cortical bone characteristics *in vivo*," Osteoporosis Int. **12**, 399–405.
Siffert, R. S., and Kaufman, J. J. (**2007**). "Ultrasonic bone assessment: The time has come," Bone **40**, 5–8.
Stone, K., Seeley, D., Lui, L.-Y., Cauley, J. A., Ensrud, K., Browner, W. S., Nevitt, M. C., and Cummings, S. R. (**2003**). "BMD at multiple sites and risk of fracture of multiple types: Long-term results from the study of osteoporotic fractures," J. Bone Miner. Res. **18**, 1947–1954.
Tatarinov, A., Sarvazyan, N., and Sarvazyan, A. (**2005**). "Use of multiple acoustic wave modes for assessment of long bones: Model study," Ultrasonics **43**, 672–80.
Virieux, J. (**1986**). "P-SV wave propagation in heterogenous media: Velocity-stress finite-difference method," Geophysics **51**, 889–901.
Yapura, C. L., and Kinra, V. K. (**1995**). "Guided waves in a fluid-solid bilayer," Wave Motion **21**, 35–46.

Ultrasonic excitation of a bubble inside a deformable tube: Implications for ultrasonically induced hemorrhage

Hongyu Miao
Mechanical Engineering, University of Rochester, Rochester, New York 14627

Sheryl M. Gracewski[a)]
Department of Mechanical Engineering, Department of Biomedical Engineering, and Rochester Center for Biomedical Ultrasound, University of Rochester, Rochester, New York 14627

Diane Dalecki
Department of Biomedical Engineering and Rochester Center for Biomedical Ultrasound, University of Rochester, Rochester, New York 14627

(Received 31 December 2007; revised 23 June 2008; accepted 24 June 2008)

Various independent investigations indicate that the presence of microbubbles within blood vessels may increase the likelihood of ultrasound-induced hemorrhage. To explore potential damage mechanisms, an axisymmetric coupled finite element and boundary element code was developed and employed to simulate the response of an acoustically excited bubble centered within a deformable tube. As expected, the tube mitigates the expansion of the bubble. The maximum tube dilation and maximum hoop stress were found to occur well before the bubble reached its maximum radius. Therefore, it is not likely that the expanding low pressure bubble pushes the tube wall outward. Instead, simulation results indicate that the tensile portion of the acoustic excitation plays a major role in tube dilation and thus tube rupture. The effects of tube dimensions (tube wall thickness 1–5 μm), material properties (Young's modulus 1–10 MPa), ultrasound frequency (1–10 MHz), and pressure amplitude (0.2–1.0 MPa) on bubble response and tube dilation were investigated. As the tube thickness, tube radius, and acoustic frequency decreased, the maximum hoop stress increased, indicating a higher potential for tube rupture and hemorrhage.
 [DOI: 10.1121/1.2967488]

PACS number(s): 43.80.Gx, 43.35.Ei, 43.35.Wa, 43.25.Yw [DLM] Pages: 2374–2384

I. INTRODUCTION

Biomedical ultrasound is widely used for diagnostic imaging because it is inexpensive, relatively portable, and can generate real time dynamic images. In some cases, echo contrast agents, which are essentially stabilized microbubbles, are injected into the vasculature to improve the image contrast of blood flowing through organs such as the heart. However, a number of studies also indicate that ultrasound-induced hemorrhage could be enhanced by the presence of contrast agents (e.g., Miller and Gies, 1998; Miller and Quddus, 2000; Miller and Gies, 2000; Wible *et al.*, 2002; Li *et al.*, 2003, 2004).

Two main hypotheses have been proposed to explain how acoustically excited bubbles can cause hemorrhage (e.g., Zhong *et al.*, 2001; Hwang *et al.*, 2006; Hu *et al.*, 2005). One hypothesis is that vascular damage is caused by the local high pressures and temperatures or by high-speed liquid jets that are generated by the inertial collapse of the bubbles. The second hypothesis is that vessel rupture occurs as the bubble expands. To understand more precisely potential hemorrhage mechanisms, Zhong *et al.* (2001) used a high-speed camera to observe the behavior of contrast agent bubbles flowing within a submerged hollow fiber subject to lithotripter shock waves. Fiber rupture was observed when a microbubble was present inside the fiber for certain lithotripter shock waves, and no rupture was found in the absence of such bubbles under the same excitation. Zhong *et al.* (2001) concluded that lithotripter shock waves can cause the large expansion of intravascular bubbles that can lead to vessel dilation and eventually vessel rupture. In this paper, a simulation model is used to investigate the interaction between an acoustically excited bubble and a deformable vessel.

Numerous mathematical models have been developed to investigate bubble dynamics. Classical cavitation models, such as the Rayleigh–Plesset equation (Leighton, 1994; Young, 1989) and the Gilmore equation (Gilmore, 1952), describe well the spherically symmetric bubble responses to acoustic excitation in an infinite liquid domain. However, if the bubble is constrained by neighboring objects such as a blood vessel, the bubble motion will become asymmetric and a high-speed liquid jet may form. Therefore, an understanding of asymmetric bubble dynamics is needed to investigate possible mechanisms of ultrasound-induced hemorrhage in tissues containing ultrasound contrast agents.

A number of studies have been reported on asymmetric bubble expansion and collapse near rigid boundaries. Most of them investigated laser- or spark-generated bubbles (e.g., Shima *et al.*, 1981; Vogel *et al.*, 1989; Tipton *et al.*, 1992; Harris *et al.*, 1999; Brujan *et al.*, 2002; Tomita *et al.*, 2002).

[a)]Author to whom correspondence should be addressed; electronic mail: grace@me.rochester.edu

Only a few studies have been reported on the bubble behavior near rigid boundaries in an oscillating pressure field (Sato *et al.*, 1994; Krasovitski and Kimmel, 2001; Brujan, 2004). For these numerical investigations, the boundary element method has been successfully employed to simulate asymmetric bubble motion, assuming that the bubble is in an incompressible and inviscid liquid. For ultrasonically excited bubbles, the work of Sato *et al.* (1994) and Krasovitski and Kimmel (2001) indicated that as the ultrasound frequency increases or the distance from the bubble to the rigid boundary decreases, the bubble collapse is mitigated.

The effect of the deformation of nearby structures on bubble behavior has also been investigated. The behavior of a laser-induced bubble near a deformable gel sample was observed using high-speed photography by Brujan *et al.* (2001a, 2001b). Also, experimental observations of interactions between ultrasonically excited bubbles and cells (Wolfrum *et al.*, 2002; Kudo *et al.*, 2002) or tubes (Zheng *et al.*, 2007) have been reported. A numerical investigation of bubble interaction with nearby deformable structures was conducted by Chahine and Kalumuck (1998), Duncan *et al.* (1996), and Klaseboer *et al.* (2005). Numerical techniques, such as the fluid-structure coupling procedures in Chahine and Kalumuck (1998) and Duncan *et al.* (1996), were used to develop the simulation models used in the current study.

Recently, a number of papers that present numerical investigations specifically developed to investigate the asymmetric oscillations of an intravascular bubble have been published. Hu *et al.* (2005) developed a theoretical model, including effects to second order, for asymmetric bubble oscillations inside a rigid tube with a radius much larger than the bubble radius. The asymmetric oscillations of a bubble inside a deformable tube excited by a short lithotripter pulse simulated using a finite element technique and a finite volume technique were described in Qin *et al.* (2006) and Gao *et al.* (2007), respectively. In both of these papers, the shock wave pressure was applied at the bubble surface. Qin and Ferrara (2006) used the finite element program COMSOL to model the response of a bubble in an incompressible, viscous liquid to a single sinusoidal cycle acoustic excitation. The effect of the tube was approximated by a nonlinear boundary condition. These papers presented predictions of the resulting pressure on the inner tube surface or the transmural pressure across the vessel wall. Qin and Ferrara (2006) predicted that an acoustic excitation with a frequency of 1 MHz and an amplitude of 0.5 MPa would be sufficient to rupture small vessels with a diameter less than 15 μm. In the current paper, a harmonic acoustic pressure excitation is applied at infinity, similar to the Rayleigh–Plesset model, and the surrounding vessel is modeled with finite elements so the internal stresses can be determined.

An understanding of interactions between ultrasonically excited contrast agents and nearby tissues and vessels is needed to determine damage mechanisms and provide guidelines for safe usage of diagnostic ultrasound. The goal of this investigation was to further explore the possible damage mechanisms of ultrasound-induced hemorrhage when contrast agents are present. A coupled finite element and boundary element code was employed to investigate the axisymmetric ultrasonically excited bubble-vessel interactions. In particular, simulations were used to investigate the hypothesis that tube rupture occurs due to bubble expansion. The finite element model of the vessel can predict not only vessel deformation but also stresses induced in the vessel, and therefore the potential for vessel rupture can be assessed. The effects of a range of vessel material properties (Young's moduli from 1 to 10 MPa), vessel dimensions (vessel lengths from 30 to 60 μm; vessel wall thicknesses from 1 to 5 μm; inner vessel radii from 2 to 8 μm), and acoustic excitations (frequencies from 1 to 10 MHz; pressure amplitudes from 0.2 to 1.0 MPa) on bubble behavior and tube dilation and internal stresses were investigated. Implications of the numerical results for ultrasound bioeffects are discussed.

II. COUPLED BOUNDARY ELEMENT–FINITE ELEMENT MODEL

A coupled finite element and boundary element code has been developed to solve axisymmetric gas-liquid-solid interaction problems to predict the response of acoustically excited bubbles. The gas within the bubble is assumed to be spatially uniform and to obey the polytropic gas law (Prosperetti, 1991),

$$p_b = p_v + p_g = p_v + p_{g0}\left(\frac{R_0}{R}\right)^{3\Gamma}, \tag{1}$$

where p_g is the gas pressure inside the bubble, R is the equivalent instantaneous bubble radius, R_0 is the equilibrium bubble radius, p_{g0} is the equilibrium gas pressure inside the bubble, Γ is the polytropic exponent (equal to the ratio of the specific heats of the gas for an isentropic process), p_v is the constant vapor pressure, and p_b is the sum of the gas pressure and the vapor pressure within the bubble. In addition, a surface tension σ at the bubble liquid interface is considered in our model. The liquid is assumed to be incompressible, irrotational, and inviscid, and therefore modeled in terms of the potential flow equations using the boundary element method. The first two assumptions are consistent with the Rayleigh–Plesset model, which adequately models the bubble expansion and early collapse phases of the bubble considered in this paper. The main effects of compressibility and viscosity occur during inertial collapse, which is not the focus of this paper. Therefore these effects have not been included for simplicity. Viscosity may also reduce the maximum bubble expansion and will therefore be investigated in future work. Since velocity potential in the liquid ϕ_L satisfies the Laplace equation, the governing equation of liquid is given in the cylindrical coordinates as

$$\frac{\partial^2 \phi_L}{\partial r^2} + \frac{1}{r}\frac{\partial \phi_L}{\partial r} + \frac{\partial^2 \phi_L}{\partial z^2} = 0. \tag{2}$$

The acoustic pressure is applied at infinity as in the Rayleigh–Plesset equation for a spherical bubble response to acoustic excitation (Leighton, 1994). The vessel is modeled as a linear elastic solid using a finite element method. Tractions and normal velocity are assumed continuous across the fluid-solid boundary to couple the finite element and bound-

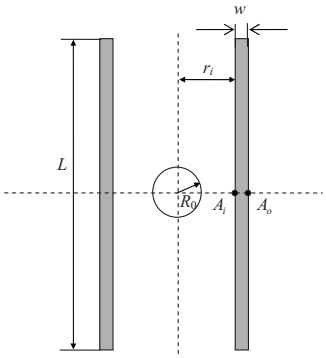


FIG. 1. Schematic illustration of the axisymmetric geometry of a bubble within a finite tube length.

ary element domains. For details of theoretical issues, numerical techniques, and validation of this code, the reader is referred to Miao and Gracewski (2008).

The following investigation focuses on interactions between a tube and a bubble located at the center of the tube. A schematic illustration of the problem geometry is given in Fig. 1, where R_0 is the initial bubble radius, r_i is the inner radius of the tube, w is the tube wall thickness, A_i is the midpoint on the inner tube surface, and A_o is the midpoint on the outer tube surface.

For numerical simulations, the equilibrium bubble radius, $R_0=1.5$ μm, was chosen to represent the average size of typical ultrasound contrast agents (Bouakaz and De Jong, 2007). The average inner radius of capillaries is about 4 μm. However, to investigate the influence of the vessel inner radius on bubble oscillations, simulation results were obtained for $r_i=2$, 4, and 8 μm. Considering the average dimensions of microvessels such as capillaries, the length of the elastic tube should be much greater than the size of the bubble. However, due to limited computing resources, a length of 30 μm, ten times the initial bubble diameter, was actually employed. The implications of this limitation will be discussed later in the paper. The average thickness of a capillary wall is about 1 μm (Megerman and Abbott, 1983). Since capillaries are generally surrounded by other tissues such as muscle, the effects of larger wall thicknesses were also investigated as a simplified model of a vessel surrounded by tissue. The related computing costs increase dramatically with increases in vessel wall thickness. Therefore simulations were done mainly with 1 and 5 μm wall thicknesses. Increasing the wall thickness further does not significantly affect the bubble and tube responses for the cases with an acoustic pressure amplitude equal to 0.2 MPa considered in this paper, as suggested by our preliminary results (not shown).

Blood vessel walls generally consist of multiple layers with different mechanical properties, and experimental measurements indicate that vessel materials exhibit viscoelastic characteristics (Fung, 1981). However, to obtain the general characteristics of the effects of vessel deformation on the bubble response, a simplified homogeneous, isotropic, and linear elastic model was used. Specifically, the linear elastic constitutive equation is given as

$$\sigma_{ij} = \frac{E}{1+\nu}\left(\varepsilon_{ij} + \frac{\nu}{1-2\nu}\varepsilon_{kk}\delta_{ij}\right), \tag{3}$$

where δ_{ij} is the Kronecker delta ($\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ if $i \neq j$) and σ_{ij} and ε_{ij} are the ijth components of the stress and strain tensors, respectively. The material parameters are Young's modulus E and Poisson's ratio ν.

The Young's modulus, the Poisson's ratio, and the density of the vessel wall material need to be specified. From previous experimental measurements vascular materials are known to be nearly incompressible (Fung, 1981), so a Poisson's ratio of 0.49 is used (Melbin and Noordergraaf, 1971). The value of Young's modulus of vascular materials may vary between 0.98 MPa (or lower) and 9.6 MPa (or higher) (Yamada, 1970; Duck, 1990; Rowe *et al.*, 2003; Snowhill and Frederick, 2005) for vessels within different organs. In the numerical simulations, Young's moduli of 1, 5, and 10 MPa were employed to span a range of Young's moduli relevant to vascular materials. A vessel wall density of 1100 kg/m^3 was used in all the simulations (Megerman and Abbott, 1983).

The tensile strength of vessels also varies between 0.46 MPa (or lower) and 3.6 MPa (or higher) (Yamada, 1970; Rowe *et al.*, 2003; Snowhill and Frederick, 2005). The longitudinal tensile strength is generally greater than the circumferential tensile strength. For example, Rowe *et al.* (2003) measured cerebral arteries and obtained a longitudinal strength of 3.6 MPa, which is greater than the circumferential strength of 2.4 MPa. Rowe *et al.* (2003) explained that this difference is due to the multidirectional layered structure of vessel wall. In addition, for the fiber rupture observed by Zhong *et al.* (2001), the cleft was along the axial direction of the fiber, which indicated that the failure of the fiber was due to the circumferential or hoop stress. Therefore, in this investigation of ultrasonically excited bubble-vessel interactions, the hoop stresses inside vessel walls were calculated and compared to the circumferential tensile strength of vessels to predict the damage threshold.

The bubble and the tube were excited by an acoustic source with a frequency of 1, 5, or 10 MHz and a pressure amplitude of 0.2, 0.5, or 1.0 MPa to investigate the effect of the excitation parameters on the bubble response. The continuous sinusoidal acoustic source was applied at infinity using the same assumptions as those of the Rayleigh–Plesset model. This range of pressure amplitudes was chosen because 0.2 MPa is just above the inertial cavitation threshold predicted by the Rayleigh–Plesset model (Young, 1989) for an excitation frequency of 1 MHz. The acoustic excitation begins with the tensile phase, which causes the bubble to initially expand. Gravity was not considered in this model, and other constant parameters used for numerical simulations were $\rho_l=1000$ kg/m^3, $p_v=2300$ Pa, $p_0=101\,230$ Pa, $\sigma=0.0717$ N/m, and $\Gamma=1.4$, where ρ_l is the liquid density, p_v is the vapor pressure, p_0 is the ambient pressure, σ is the surface tension, and Γ is the polytropic exponent of the gas set equal to the ratio of specific heat capacities for air.

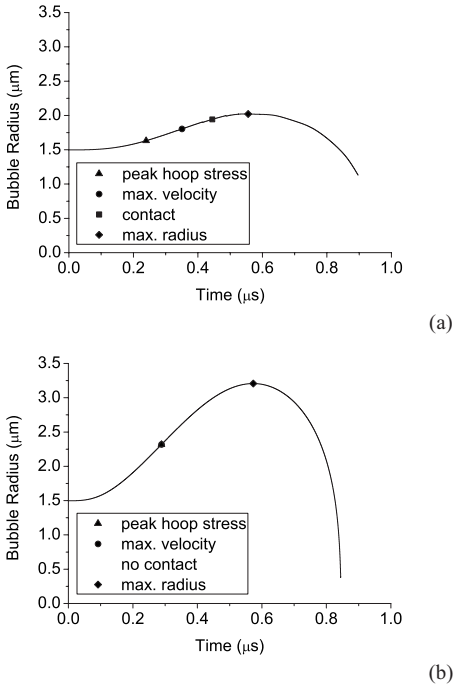


FIG. 2. Comparison of equivalent bubble radius vs time for (a) $r^*=4/3$ with bubble-tube contact during bubble expansion and (b) $r^*=16/3$ without bubble-tube contact during bubble expansion ($w=1$ μm, $E=10$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa).

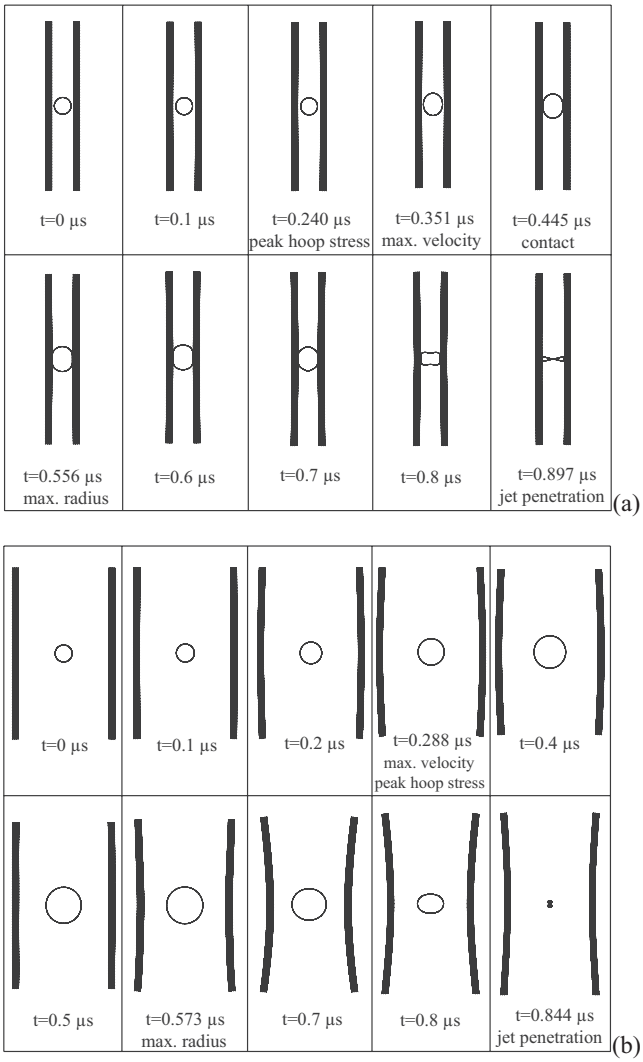


FIG. 3. Time sequence of bubble and tube shapes for (a) $r^*=4/3$ and (b) $r^*=16/3$ ($w=1$ μm, $E=10$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa). Note that tube ends may be truncated to fit into the frames. In addition, jet penetration denotes the moment when the water jet penetrates the opposite bubble boundary.

III. BUBBLE-TUBE INTERACTIONS

In this section, some general characteristics of bubble-tube interactions are described by considering typical simulation results. Results are compared for vessel wall thicknesses $w=1$ and 5 μm and inner tube radii $r_i=2$, 4, and 8 μm with ratios $r^*=r_i/R_0=4/3$, 8/3, and 16 3, respectively. The ultrasonic excitation used has a frequency of 1, 5, or 10 MHz and an amplitude of 0.2 MPa. More details about the effects of vessel wall thickness, Young's modulus, and acoustic frequency and amplitude on the interactions between the vessel and the bubble will be discussed in the next section.

General characteristics of the bubble-tube interaction depend on whether or not the bubble comes in to contact with the tube during the bubble expansion phase. The bubble is more likely to touch the inner tube surface for (1) r^* closer to 1, (2) stiffer vessels, and (3) higher amplitude acoustic excitations. Characteristics of bubble responses with and without touching are discussed by comparing the response of a bubble in a stiff tube with Young's modulus equal to 10 MPa and a thickness of 1 μm for two different inner radii, corresponding to $r^*=4/3$ and $r^*=16/3$. For both simulations, the acoustic excitation used has a frequency equal to 1 MHz and an amplitude equal to 0.2 MPa.

In Figs. 2(a) and 2(b), typical equivalent bubble radius versus time curves are depicted for $r^*=4/3$ and $r^*=16/3$, respectively. The equivalent bubble radius is the radius of a sphere with the same volume as the bubble and is used here to allow comparisons to results obtained from the Rayleigh–Plesset model. For the oscillation of a single bubble in an infinite liquid domain subject to the same acoustic excitation, the maximum bubble radius R_{max} is equal to 3.7 μm. For $r^*=16/3$, the bubble does not touch the tube wall, and $R_{max}=3.2$ μm is ~15% smaller than that for a bubble in an infinite liquid domain. For $r^*=4/3$, the bubble touches the tube during expansion and $R_{max}=2.0$ μm, almost 50% less than the free field value. Compared to the bubble collapse for $r^*=16/3$ indicated in Fig. 2(b), the bubble collapse for $r^*=4/3$ is mitigated due to a smaller value of R_{max}, as indicated in Fig. 2(a). In addition, the bubble radius versus time curve in Fig. 2(a) is still smooth (first-order differentiable) although the bubble touches the inner tube surface during bubble expansion.

In Fig. 3, a series of bubble and tube shapes are depicted for the same parameters used in Fig. 2. Figures 3(a) and 3(b) are for $r^*=4/3$ and $r^*=16/3$, respectively. The time points are spaced evenly with an increment of 0.1 μs except for some important moments, such as the time points of maximum bubble wall velocity, maximum tube dilation, bubble-vessel contact, if it occurs, and maximum bubble radius. For

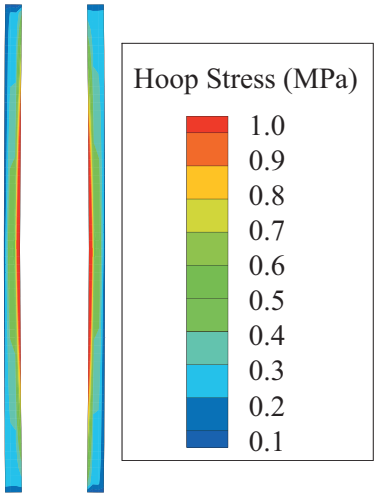


FIG. 4. Hoop stress distribution in the tube wall during bubble expansion at the moment of peak hoop stress for $r^*=4/3$ ($w=1$ μm, $E=10$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa). The maximum hoop stress of 1 MPa occurs on the inner tube wall at the location of the bubble.

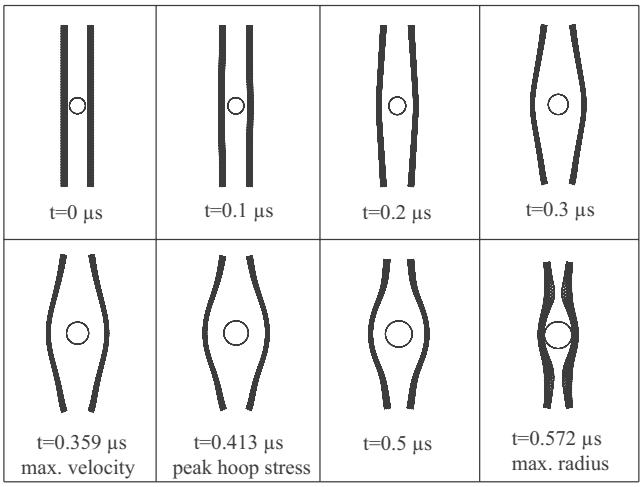


FIG. 5. Time sequence of bubble and tube shapes for a more compliant tube during bubble expansion for $r^*=4/3$ and $E=1$ MPa ($w=1$ μm, $r_i=2$ μm, $f=1$ MHz, and $p_a=0.2$ MPa).

the smaller ratio of $r^*=4/3$, the bubble touches the inner tube surface at $t=0.445$ μs. No contact occurs for the larger ratio of $r^*=16/3$.

Note that the maximum hoop stress occurs during the tensile portion of the acoustic pulse, well before the bubble reaches its maximum radius. In Fig. 3(a), between $t=0.24$ μs and $t=0.445$ μs, the tube contracts as the bubble expands. The gas pressure p_g in the expanding bubble is nearly zero. Therefore the bubble nearing maximum expansion could not "push" on the vessel wall if the stress outside the vessel were not tensile. Figure 4 gives the hoop stress distribution within the tube wall at $t=0.24$ μs, the moment of peak hoop stress during bubble expansion, for $r^*=4/3$. The parameters used in Fig. 4 are the same as the values in Figs. 2(a) and 3(a). The higher hoop stresses occur at the inner tube surface near the middle of the tube (this is the area that deforms the most and is closest to the bubble). The hoop stress distribution in the tube wall at the moment of maximum tube dilation during bubble expansion is similar for all cases considered in this paper.

The ratio r^* is not the only factor that determines whether the bubble will touch the inner tube surface. For instance, if Young's modulus is changed from $E=10$ MPa to $E=1$ MPa for the ratio $r^*=4/3$, with all other parameters the same, the bubble will not touch the wall upon bubble expansion due to the dilation of the compliant tube wall, as indicated in Fig. 5. For this more compliant tube, the maximum radius $R_{\max}=2.5$ μm is greater than the original tube radius. However, the tube expands appreciably as the bubble expands, so they never touch. For this case, results are only shown for the bubble expansion phase because the vessel wall maximum strain is on the order of 0.1. This exceeds the small deformation assumption and leads to an accumulation of numerical errors such that the numerical results for the collapse stage become unreliable. Figure 6 shows the equivalent bubble radius versus time curve during bubble expansion for the same parameters used in Fig. 5.

For all three cases, as shown in Figs. 2(a), 2(b), and 6, the peak hoop stress occurs well before the bubble reaches its maximum radius. Even though the peak hoop stress occurs at the same time as the maximum bubble wall velocity in Fig. 2(b), this is not true in general. For example, the peak hoop stress occurs before and after the moment of maximum bubble wall velocity in Figs. 2(a) and 6, respectively.

During the bubble expansion phase, the maximum principal stress in the vessel wall is the hoop stress. In Fig. 7, the maximum hoop stress $(\sigma_{\theta\theta})_{\max}$ is plotted versus time during the bubble expansion stage for different inner tube radii r_i and acoustic frequencies f. Here, $(\sigma_{\theta\theta})_{\max}$ is the maximum value of $(\sigma_{\theta\theta})_i$, where $(\sigma_{\theta\theta})_i$ is the hoop stress of element i. For all cases shown, $(\sigma_{\theta\theta})_{\max}$ occurs at point A_i. The parameters of these six cases in Fig. 7 are intentionally chosen to be as diversified as possible. For example, the tube wall thickness for Fig. 7(a) is $w=1$ μm, which is intentionally different from the wall thickness $w=5$ μm for Fig. 7(b). Figure 7 indicates that, as the bubble expands, $(\sigma_{\theta\theta})_{\max}$ increases from zero to a maximum value and then decreases down to almost zero. In Fig. 7(a), as the inner tube radius increases from 2 to 8 μm, the peak value of the maximum hoop stress decreases from 1.7 to 1.1 MPa. In Fig. 7(b), as the ultra-

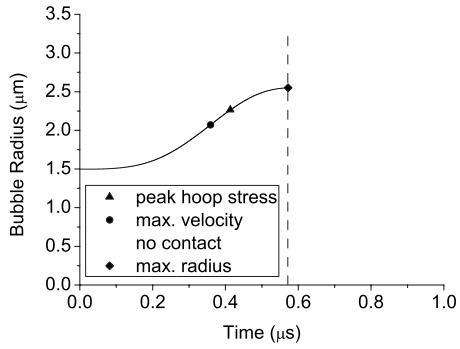


FIG. 6. Equivalent bubble radius vs time during bubble expansion for $r^*=4/3$ and $E=1$ MPa ($w=1$ μm, $r_i=2$ μm, $f=1$ MHz, and $p_a=0.2$ MPa), corresponding to the time sequence of Fig. 5.

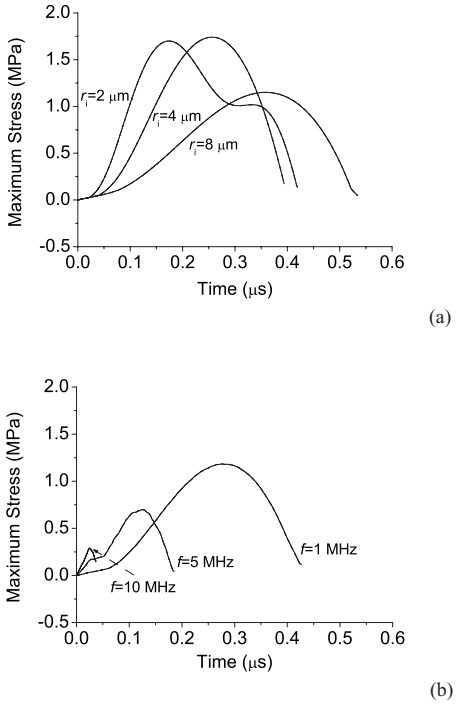


FIG. 7. Maximum hoop stress $(\sigma_{\theta\theta})_{\max}$ vs time for (a) different inner tube radii ($w=1$ μm, $E=5$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa) and (b) different ultrasound frequencies ($r_i=4$ μm, $w=5$ μm, $E=5$ MPa, and $p_a=0.2$ MPa).

sound frequency increases from 1 to 10 MHz, the peak value of the maximum hoop stress decreases from 1.2 to 0.3 MPa.

Since the hoop stress is the maximum stress component during bubble expansion, vessel rupture is predicted when the hoop stress exceeds the circumferential strength of the vessel wall material. This numerical prediction is consistent with the experimental observation of tube rupture patterns by Zhong *et al.* (2001). In their experiments, a cleft was found along the axis of the hollow fiber, which indicated that the tube was ripped open as the hoop stress exceeded the circumferential tensile strength.

An important question is what causes the increase in hoop stress during bubble expansion shown in Fig. 7, or, in other words, what causes tube dilation as the bubble expands. The answer to this question is critical to determine the damage mechanisms of vessel rupture. Zhong *et al.* (2001) observed the response of echo contrast agent bubbles confined within tubes subject to lithotripter shock waves and claimed that the large bubble expansion causes vessel dilation. However, in their experiment, the bubble with an equilibrium radius of 1 μm expands to almost 200 times its initial radius, which results in a low pressure inside the bubble at the maximum bubble radius (about $1/10^7$ of the initial equilibrium pressure inside the bubble). Such low pressure is unlikely to push the vessel wall and cause rupture. Instead, Gracewski and Miao (2006) reported that the bubble can expand to meet the vessel wall during the tensile portion of the lithotripter shock wave. At this time, the pressure in the bubble is almost zero so the external radial tensile stress is not balanced by an internal radial tensile stress as it would be for a liquid-filled vessel. Therefore, it was hypothesized that vessel rupture in lithotripsy occurs because the tensile force of the shock wave pulse stretches the vessel, producing a large hoop strain and a hoop stress that exceeds the strength of the vessel. This simple model neglects inertia effects. Therefore, the current simulations, which account for inertia effects, can add more insight into the mechanisms of tube rupture.

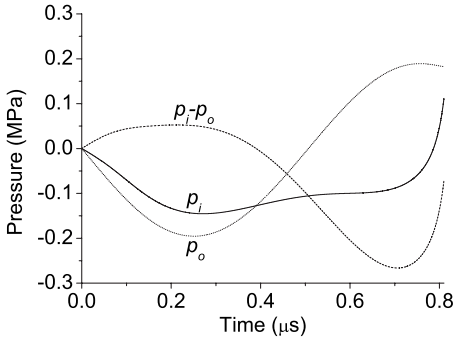


FIG. 8. Time traces of the gauge pressure p_i at point A_i (solid), the gauge pressure p_o at point A_o (dotted), and the pressure drop across the tube wall (p_i-p_o) (dashed) for $w=1$ μm, $r_i=8$ μm, $E=10$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa. These are the same parameters as in Figs. 2(b) and 3(b). A positive pressure drop will cause the tube to dilate. Note that the pressure drop becomes negative at ~0.4 μs, well before the maximum bubble radius at 0.57 μs.

For this investigation, the acoustic excitation amplitude is not as high as for lithotripter shock waves, so the bubble may not expand to touch the tube wall. However, during the time when the negative acoustic pressure is being applied, the bubble could expand rapidly and create a velocity field in the liquid around the bubble. The liquid directed to the inner tube wall surface can generate a positive dynamic pressure on the inner wall surface. Much of the bubble expansion occurs during the tensile portion of the acoustic excitation; therefore, the pressure on the outer tube surface could be negative, increasing the pressure drop across the tube wall. This pressure drop would cause the vessel to dilate, thereby generating a hoop stress in the vessel wall.

Figure 8 shows typical plots of the gauge pressures p_i at point A_i on the inner surface and p_o at point A_o on the outer surface of the tube wall (as indicated in Fig. 1). These pressures and the pressure drop (p_i-p_o) across the tube wall are plotted versus time for $w=1$ μm, $r_i=8$ μm, $E=10$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa, the same parameters as in Figs. 2(b) and 3(b). As indicated in Fig. 8, the time trace of the gauge pressure p_o at point A_o is very close to the time-dependent acoustic excitation pressure, indicating a negligible influence of bubble motion on the pressure on the outer tube surface. However, the gauge pressure p_i at point A_i deviates significantly from the acoustic excitation pressure due to the close proximity of the bubble. The curve for p_i qualitatively matches the pressure curve (not shown) predicted by the Rayleigh–Plesset model for a point near a spherically oscillating single bubble in an infinite liquid domain.

The positive peak of the pressure drop curve occurs at $t=0.21$ μs, close to the time of the peak hoop stress at $t=0.28$ μs. The pressure drop across the wall then decreases, becoming negative at $t=0.39$ μs before the bubble reaches its maximum radius at $t=0.57$ μs. The nadir of the pressure drop curve occurs at $t=0.71$ μs, and then the pressure drop

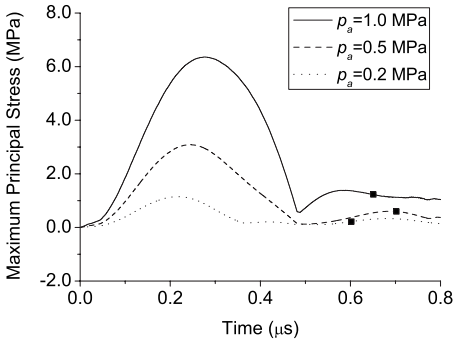


FIG. 9. Maximum principal stress σ_{max} in the vessel wall vs time for different acoustic pressure amplitudes (w=5 μm, r_i=4 μm, E=10 MPa, and f=1 MHz). The black square on each curve indicates the moment of maximum bubble radius.

rapidly increases as the bubble collapses. This plot is consistent with the tube behavior shown in Fig. 3(b).

Usually the collapse stage attracts more attention from researchers due to the formation of a high-speed water jet that can damage nearby structures. However, in this investigation, the collapse stage is less interesting since the hoop stress in the tube wall during the collapse stage is generally significantly less than the stress during the expansion stage, as indicated in Fig. 9. Some typical plots of the maximum principal stress versus time are presented in Fig. 9 for different pressure amplitudes. The results indicate that even though the acoustic pressure increases from 0.2 to 1.0 MPa, which leads to a more violent expansion and collapse of the bubble, the maximum principal stress σ_{max} at the collapse stage is less than 20% of the value at the expansion stage. Although water jets form during bubble collapse, the current model is axisymmetric such that two opposing water jets form and mitigate the impact of each other, as indicated in Figs. 3(a) and 3(b). The jet direction could be along the axis of symmetry or perpendicular to the tube wall, depending mainly on whether the bubble touches the inner tube surface during the bubble expansion phase. To investigate the influence of water jets, a three-dimensional model needs to be developed to allow the bubble to be located off the axis of symmetry. Therefore, the following numerical investigation and discussion will mainly focus on the bubble expansion stage.

IV. EFFECTS OF PARAMETERS ON TUBE DILATION

In this section, the influence of parameters, such as tube geometry, Young's modulus, and ultrasound excitation frequency and amplitude, on tube dilation are investigated. The peak hoop stress $(\sigma_{\theta\theta})_{pk}$ values, defined as the maximum value of the $(\sigma_{\theta\theta})_{max}$ versus time curve during bubble expansion, are compared for cases with different parameters, as an indicator of tube dilation.

A. Effects of tube geometry on peak hoop stress $(\sigma_{\theta\theta})_{pk}$

First, the influence of different tube geometries and material properties on tube dilation is investigated. In particular, tube wall thicknesses w=1 and 5 μm and inner tube radii r_i=2, 4, and 8 μm are considered for Young's moduli E=1, 5, and 10 MPa. A fixed ultrasonic excitation with a frequency f=1 MHz and a pressure amplitude p_a=0.2 MPa was used in the simulations.

TABLE I. Effects of tube geometry on peak hoop stress $(\sigma_{\theta\theta})_{pk}$ (f=1 MHz, p_a=0.2 MPa). Values given in the table for $(\sigma_{\theta\theta})_{pk}$ are in MPa.

E (MPa)	w (μm)	r_i (μm) 2	4	8
1	1	1.3	1.0	0.5
	5	1.3	0.8	0.3
5	1	1.7	1.7	1.1
	5	1.5	1.2	0.6
10	1	1.7	1.6	1.2
	5	1.2	1.2	0.6

In Table I, the values of $(\sigma_{\theta\theta})_{pk}$ are listed for all combinations of the parameters mentioned above. In general, as the wall thickness increases from 1 to 5 μm, the value of $(\sigma_{\theta\theta})_{pk}$ decreases, even as much as 50% for a stiffer tube (E=10 MPa) with a larger inner radius (r_i=8 μm). As the inner tube radius increases from 2 to 8 μm, the decrease in $(\sigma_{\theta\theta})_{pk}$ is more noticeable and can be as much as 80%. Zhong *et al.* (2001) experimentally observed that as the size of hollow fibers becomes smaller, the likelihood of ultrasound-induced fiber rupture is substantially increased, consistent with the results obtained here. The numerical results indicate that if all the conditions remain the same except the tube geometry, ultrasound-induced tube rupture during bubble expansion is less likely to occur for thicker and larger tubes.

B. Effects of tube geometry on peak hoop strain $(\varepsilon_{\theta\theta})_{pk}$ and pressure drop across the tube wall $(p_i-p_o)_{pk}$

The peak values of the hoop strain $(\varepsilon_{\theta\theta})_{pk}$ and the pressure drop across the tube wall $(p_i-p_o)_{pk}$ are listed in Tables II and III, respectively. To understand these results, it is instructive to compare them with the theoretical solution for a

TABLE II. Effects of tube geometry on peak hoop strain $(\varepsilon_{\theta\theta})_{pk}$ (f=1 MHz, p_a=0.2 MPa).

E (MPa)	w (μm)	r_i (μm) 2	4	8
1	1	0.60	0.50	0.27
	5	0.28	0.20	0.08
5	1	0.11	0.16	0.13
	5	0.05	0.06	0.04
10	1	0.05	0.07	0.07
	5	0.02	0.03	0.02

TABLE III. Effects of tube geometry on pressure drop across the tube wall $(p_i-p_o)_{pk}$ ($f=1$ MHz, $p_a=0.2$ MPa). Values given in the table for $(p_i-p_o)_{pk}$ are in MPa.

E (MPa)	w (μm)	r_i (μm) 2	4	8
1	1	0.143	0.099	0.052
	5	0.154	0.106	0.055
5	1	0.153	0.100	0.052
	5	0.157	0.106	0.055
10	1	0.157	0.102	0.053
	5	0.158	0.106	0.056

circular cylinder with constant internal and external pressures. The equations for the hoop stress and hoop strain are (Ansel *et al.*, 1995)

$$\sigma_{\theta\theta}=\frac{p_i-(r_o/r_i)^2p_o}{(r_o/r_i)^2-1}+\frac{(p_i-p_o)}{[(r_o/r_i)^2-1](r/r_o)^2}, \tag{4}$$

$$\varepsilon_{\theta\theta}=\frac{1-\nu}{E}\frac{p_i-(r_o/r_i)^2p_o}{(r_o/r_i)^2-1}+\frac{1+\nu}{E}\frac{(p_i-p_o)}{[(r_o/r_i)^2-1](r/r_o)^2}, \tag{5}$$

where r_i is the inner cylinder radius, r_o is the outer cylinder radius, p_i is the internal pressure, p_o is the external pressure, and r is the radius ($r_i \leqslant r \leqslant r_o$). Note that these equations neglect the inertial effects and the variation of pressure along the cylinder axis. The hoop strain predicted by Eq. (5) decreases from its maximum at the inner wall, consistent with Fig. 4.

For a given inner radius and applied pressures, as the thickness of the tube increases, the hoop strain and tube dilation decreases. Numerical results for the bubble-tube interaction are consistent with this prediction. Table II shows a decrease in $(\varepsilon_{\theta\theta})_{pk}$ by more than a factor of 2 as w increases from 1 to 5 μm. In addition, the hoop force is supported by a larger area. Therefore, the maximum hoop stress decreases as the wall thickness increases. However, the decrease in hoop stress is not as pronounced for the numerical simulations because the smaller dilation of a thicker tube will affect the bubble expansion and the resulting internal pressure on the tube.

For a given tube thickness and applied pressure, the hoop stress increases as the inner radius increases. However, the bubble has a smaller effect on the internal tube wall pressure in larger tubes, producing a smaller pressure drop across the tube wall, as shown in Table III. Therefore, for the current simulations, hoop stress decreases as the inner radius increases, as shown in Table I.

Finally, for a pressurized cylinder the hoop strain and dilation are inversely proportional to Young's modulus, while the hoop stress is independent of Young's modulus. The maximum hoop strains in Table II show a decrease with increasing Young's modulus that approximately follows this inverse relationship. The bubble response will be affected by a decrease in dilation. Therefore the numerical results in Table II show that the hoop stress is affected by Young's modulus. If the bubble does not touch the tube, in general, a higher pressure and thus a higher hoop stress will be generated for a tube with a larger Young's modulus. Contact with the tube will mitigate the bubble expansion further. The effect of Young's modulus on hoop stress becomes negligible if the Young's modulus is high enough such that the dilation is negligible.

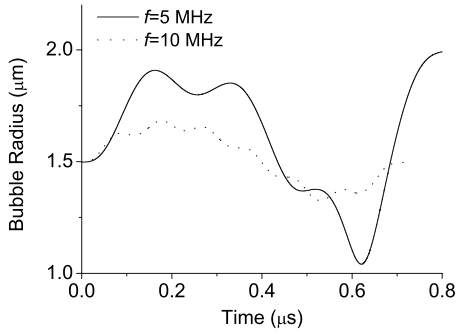


FIG. 10. Comparison of equivalent bubble radius vs time curves for $f=5$ MHz and $f=10$ MHz ($w=5$ μm, $r_i=4$ μm, $E=5$ MPa, and $p_a=0.5$ MPa).

C. Effects of excitation parameters and tube stiffness on bubble and tube response

To determine the influence of acoustic frequency and pressure amplitude on tube dilation, bubble responses in a tube with wall thickness $w=5$ μm, inner radius $r_i=4$ μm, and Young's modulus $E=1$, 5, or 10 MPa are compared. The ultrasound frequencies considered are $f=1$, 5, and 10 MHz, and the pressure amplitudes are $p_a=0.2$, 0.5, and 1.0 MPa.

In Fig. 10, typical curves of equivalent bubble radius and maximum hoop stress $(\sigma_{\theta\theta})_{max}$ versus time for ultrasonic excitation with higher frequencies ($f=5$ and 10 MHz) are plotted. Since the driving frequency is higher than the natural frequency of the bubble, the forced oscillation of the bubble is modulated by the natural frequency. The first peak value is not necessarily the highest one. Instead, the highest peak value can be found after a few cycles of the ultrasonic excitation. However, for this research, the peak value within the first cycle of the ultrasonic excitation is considered since the investigation of the effects of pulse duration on tube dilation is beyond the scope of this research. However, the response to a single pulse is typically a good indicator of the magnitude of the bubble response to multiple pulses at the same frequency, except when the bubble is excited near resonance. As indicated in Fig. 10, as the frequency increases, the maximum bubble radius during expansion decreases, which mitigates the bubble collapse. Therefore, the peak value of $(\sigma_{\theta\theta})_{max}$ within the first period of the acoustic excitation for $f=5$ MHz is greater than the value for $f=10$ MHz.

To better understand the influence of frequency and amplitude on tube dilation, in Table IV, the peak value of $(\sigma_{\theta\theta})_{max}$ is listed during the first cycle of the acoustic excitation for all combinations of the parameters mentioned above. The numerical results indicate that if all the conditions remain the same except the acoustic excitation parameters, ultrasound-induced vessel rupture during bubble ex-

TABLE IV. Effects of ultrasound frequency (f), pressure amplitude (p_a), and Young's modulus (E) on the peak hoop stress $(\sigma_{\theta\theta})_{pk}$ during the first cycle of ultrasound waves ($w=5$ μm, $r_i=4$ μm). Values given in the table for $(\sigma_{\theta\theta})_{pk}$ are in MPa.

p_a (MPa)	E (MPa)	f (MHz) 1	5	10
0.2	1	0.8	0.2	0.1
	5	1.2	0.7	0.3
	10	1.2	1.0	0.5
0.5	1	1.4	0.5	0.3
	5	3.1	1.6	0.9
	10	3.1	2.5	1.2
1.0	1	1.7	0.7	0.5
	5	5.3	2.7	1.3
	10	6.3	4.5	2.2

pansion is less likely to occur for higher ultrasound frequencies and lower pressure amplitudes. Note that this investigation does not consider ultrasound waves with frequencies equal to the natural frequency of the bubble, so further work is required to investigate resonance effects.

Tables I and IV indicate that if all the conditions remain the same except Young's modulus, the peak value of $(\sigma_{\theta\theta})_{max}$ generally increases as Young's modulus increases. Qin *et al.* (2006) investigated the dynamic pressure on the inner tube surface for bubble-tube interactions exposed to a very short shock wave pulse. They found that the maximum value of the dynamic pressure increased as the tube became stiffer. This observation is consistent with the conclusion drawn from Table IV. However, it should not be concluded that stiffer vessels are more likely to rupture if all other conditions remain the same since generally stiffer vessels have higher circumferential tensile strengths (Yamada, 1970; Rowe *et al.*, 2003; Snowhill and Frederick, 2005).

D. Effects of tube length on the maximum bubble radius and the peak value of $(\sigma_{\theta\theta})_{max}$

Finally, our preliminary results indicated that as the tube length increases, the maximum bubble radius decreases but the peak value of $(\sigma_{\theta\theta})_{max}$ increases. For example, for the case of $w=1$ μm, $r_i=4$ μm, $E=5$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa, the results are compared for $L=30$ μm and $L=60$ μm in Fig. 11. The equivalent bubble radius versus time curves are plotted in Fig. 11(a). As the tube length increases from 30 to 60 μm, the maximum bubble radius decreases from 2.7 to 2.4 μm. The curves of $(\sigma_{\theta\theta})_{max}$ versus time during bubble expansion in Fig. 11(b) indicate that the peak value of $(\sigma_{\theta\theta})_{max}$ increases from 1.7 to 2.5 MPa as the tube length increases. Because the liquid is assumed to be incompressible, the tube dilates and liquid flows from the tube ends as the bubble expands. Longer tubes provide more resistance to liquid motion along the tube length. Therefore, these results should provide a conservative estimate for predicting bubble-capillary response *in vivo*. Due to the limited computing resources, the effects of tube length on bubble-tube interactions were not thoroughly investigated here but should be addressed in the future.

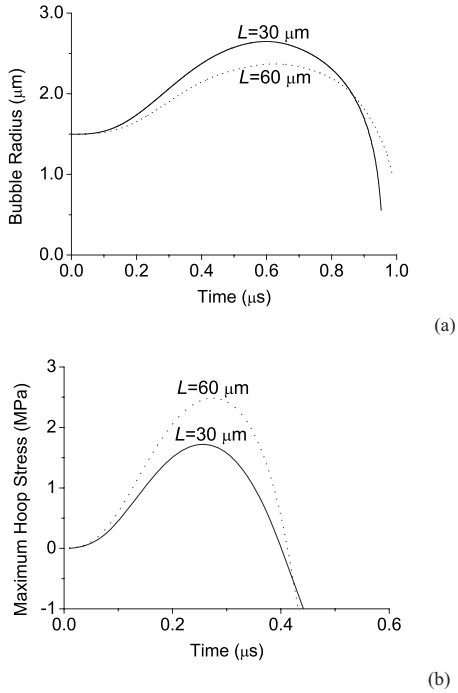


FIG. 11. Comparison of (a) bubble radius vs time and (b) maximum hoop stress vs time for $L=30$ μm and $L=60$ μm ($w=1$ μm, $r_i=4$ μm, $E=5$ MPa, $f=1$ MHz, and $p_a=0.2$ MPa). The longer tube inhibits bubble expansion.

V. SUMMARY

In this paper, results were presented for a computational model developed to simulate the interactions of a bubble and a surrounding deformable tube subject to acoustic excitation. An axisymmetric geometry was used with the bubble centered within the tube. The effects of tube dimensions and stiffness and the effects of acoustic excitation frequency and amplitude on the maximum bubble radius and tube hoop stresses and strains were investigated. The simulations provide new insight into the mechanism of tube rupture.

The numerical results indicated that the hoop stress is the maximum stress component in the tube during bubble expansion. Therefore, tube rupture is predicted when the hoop stress exceeds the hoop strength of the tube. As shown in Table IV, a peak hoop stress of 3.1 MPa was predicted for an acoustic excitation of 0.5 MPa and 1 MHz for a tube with an elastic modulus of 5 or 10 MPa. This exceeds the published value of vessel wall circumferential strength of 2.4 MPa (Rowe *et al.*, 2003) and therefore indicates that vessel damage may occur at this acoustic excitation level. This value is close to the experimentally measured threshold for petechiae and hemorrhage in a mouse intestine of 0.85 MPa at 1.09 MHz when microbubble contrast agents are present in the vasculature (Miller and Gies, 2000; Miller *et al.* 2008).

It was shown that the maximum hoop stress occurs during the tensile phase of the acoustic excitation as the bubble expands and that the tube begins to contract, well before the maximum bubble radius is reached. During the tensile phase of the acoustic excitation, the external pressure is lower than

the pressure inside the vessel. This results in vessel dilation and a positive hoop stress that could eventually lead to vessel rupture. Due to inertial effects, the bubble maximum occurs during the compressive phase of the acoustic cycle. At this point, the pressure inside the tube surrounding the bubble is low due to the low gas pressure inside the expanded bubble. In addition, the pressure outside the tube increases due to the increasing pressure of the acoustic source. Therefore, the pressure drop across the tube wall causes the tube to contract. For this axisymmetric geometry, the peak hoop stress during the bubble collapse stage is less than the peak value during the expansion stage. Therefore, at least for this axisymmetric geometry, vessel rupture is more likely to occur during bubble expansion.

In addition, the numerical results indicate that ultrasound-induced tube rupture is less likely to occur for thicker and larger tubes and for ultrasound waves with a higher frequency and a lower pressure amplitude. This may indicate that ultrasonically induced hemorrhage is also less likely to occur for these conditions. These trends are consistent with experimental observations. For example, when contrast agents are present, the ultrasound threshold for the production of petechiae and hemorrhages in a mouse intestine increased with increasing ultrasound frequency (Miller and Gies, 2000), and the severity and area of renal hemorrhage in rats increased with increasing transducer power or decreasing ultrasound frequency (Wible *et al.*, 2002). Also, Skyba *et al.* (1998) observed rupture of smaller ($\leqslant 7$ μm) microvessels when exposing the spinotrapezius muscle in rats to ultrasound after echo contrast agent infusion. Finally, during the expansion stage, the peak hoop stress increased as the vessel material became stiffer. However, since a stiffer vessel usually has a higher circumferential tensile strength, implications for hemorrhage should be drawn with caution.

The tube length for most of the results presented in this manuscript was equal to ten times the initial bubble diameter. As shown in Fig. 11, extending the tube length will further inhibit bubble expansion. For short tubes, liquid can flow along the tube axis as the bubble expands. The longer the tube is, the more resistant it is to this flow. Due to a limitation in computational resources, the effect of the tube length on the bubble response has not yet been investigated in detail. However, in the limit of an infinite tube, tube dilation must accompany bubble expansion if the liquid is incompressible. Therefore, in this limit, tube dilation and bubble expansion must be in phase.

ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful discussions with Edwin L. Carstensen and many members of the Rochester Center for Biomedical Ultrasound at the University of Rochester. This work was supported in part by NIH Grant No. R01HL69824 and NSF Grant No. CMMI-0652947.

Ansel, C. U., and Fenster, S. K. (**1995**). *Advanced Strength and Applied Elasticity* (Prentice-Hall, Englewood Cliffs).

Bouakaz, A., and De Jong, N. (**2007**). "WFUMB safety symposium on echo-contrast agents: Nature and types of ultrasound contrast agents," Ultrasound Med. Biol. **33**, 187–196.

Brujan, E. A. (**2004**). "The role of cavitation microjets in the therapeutic applications of ultrasound," Ultrasound Med. Biol. **30**, 381–387.

Brujan, E. A., Keen, G. S., Vogel, A., and Blake, J. R. (**2002**). "The final stage of the collapse of a cavitation bubble close to a rigid boundary," Phys. Fluids **14**, 85–92.

Brujan, E. A., Nahen, K., Schmidt, P., and Vogel, A. (**2001a**). "Dynamics of laser-induced cavitation bubbles near an elastic boundary," J. Fluid Mech. **433**, 251–281.

Brujan, E. A., Nahen, K., Schmidt, P., and Vogel, A. (**2001b**). "Dynamics of laser-induced cavitation bubbles near elastic boundaries: Influence of the elastic modulus," J. Fluid Mech. **433**, 283–314.

Chahine, G. L., and Kalumuck, K. M. (**1998**). "BEM software for free surface flow simulation including fluid-structure interaction effects," Int. J. Comput. Appl. Technol. **11**, 177–198.

Duck, F. A. (**1990**). *Physical Properties of Tissue* (Academic, London).

Duncan, J. H., Milligan, C. D., and Zhang, S. (**1996**). "On the interaction between a bubble and a submerged compliant structure," J. Sound Vib. **197**, 17–44.

Fung, Y. C. (**1981**). *Biomechanics: Mechanical Properties of Living Tissues* (Springer-Verlag, New York).

Gao, F., Hu, Y., and Hu, H. (**2007**). "Asymmetrical oscillation of a bubble confined inside a micro pseudoelastic blood vessel and the corresponding vessel wall stresses," Int. J. Solids Struct. **44**, 7197–7212.

Gilmore, F. R. (**1952**). "The growth or collapse of a spherical bubble in a viscous compressible liquid," Hydrodynamics Laboratory Report No. 26-4, California Institute of Technology, Pasadena, CA.

Gracewski, S. M., and Miao, H. (**2006**). "Vessel damage mechanisms by ultrasound or shock wave pulses." J. Acoust. Soc. Am. **119**, 3406.

Harris, P. J., Verma, A., and Chakrabarti, R. (**1999**). "Interaction of an explosion bubble with a fixed rigid structure," Int. J. Numer. Methods Fluids **29**, 389–396.

Hu, Y. T., Qin, S., Hu, T., Ferrara, K. W., and Jiang, Q. (**2005**). "Asymmetric oscillation of cavitation bubbles in a microvessel and its implications upon mechanisms of clinical vessel injury in shock-wave lithotripsy," Int. J. Non-Linear Mech. **40**, 341–350.

Hwang, J. H., Tu, J., Brayman, A. A., Matula, T. J., and Crum, L. A. (**2006**). "Correlation between inertial cavitation dose and endothelial cell damage in vivo," Ultrasound Med. Biol. **32**, 1611–1619.

Klaseboer, E., Hung, K. C., Wang, C., Wang, C. W., Khoo, B. C., Boyce, P., Debono, S., and Charlier, H. (**2005**). "Experimental and numerical investigation of the dynamics of an underwater explosion bubble near a resilient/rigid structure," J. Fluid Mech. **537**, 387–413.

Krasovitski, B., and Kimmel, E. (**2001**). "Gas bubble pulsation in a semi-confined space subjected to ultrasound," J. Acoust. Soc. Am. **109**, 891–898.

Kudo, N., Miyaoka, T., Okada, K., and Yamamoto, K. (**2002**). "Study on mechanism of cell damage caused by microbubbles exposed to ultrasound," Proc.-IEEE Ultrason. Symp. **2**, 1383–1386.

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

Li, P., Armstrong, W. F., and Miller, D. L. (**2004**). "Impact of myocardial contrast echocardiography on vascular permeability: Comparison of three different contrast agents," Ultrasound Med. Biol. **30**, 83–91.

Li, P., Cao, T., Cou, C., Armstrong, W. F., and Miller, D. (**2003**). "Impact of myocardial contrast echocardiography on vascular permeability: An in vivo dose response study of delivery mode, pressure amplitude, and contrast dose," Ultrasound Med. Biol. **29**, 1341–1349.

Megerman, J., and Abbott, W. M. (**1983**). "Compliance in vascular grafts," in *Vascular Grafting: Clinical Applications and Techniques*, edited by C. B. Wright (John Wright, Boston).

Melbin, J., and Noordergraaf, A. (**1971**). "Elastic deformation in orthotropic vessels: Theoretical and experimental results," Circ. Res. **28**, 680–692.

Miao, H., and Gracewski, S. M. (**2008**). "Coupled FEM and BEM code for simulating acoustically excited bubbles near deformable structures," Comput. Mech. **42**, 95–106.

Miller, D. L., Averkiou, M. A., Brayman, A. A., Everbach, E. C., Holland, C. K., Wible, J. H., and Wu, J. (**2008**). "Bioeffects considerations for diagnostic ultrasound contrast agents," J. Ultrasound Med. **27**, 611–632.

Miller, D. L., and Gies, R. A. (**1998**). "Gas-body-based contrast agent enhances vascular bioeffects of 1.09 MHz ultrasound on mouse intestine," Ultrasound Med. Biol. **24**, 1201–1208.

Miller, D. L., and Gies, R. A. (**2000**). "The influence of ultrasound frequency and gas-body composition on the contrast agent-mediated enhancement of vascular bioeffects in mouse intestine," Ultrasound Med. Biol. **26**, 307–313.

Miller, D. L., and Quddus, J. (**2000**). "Diagnostic ultrasound activation of

contrast agent gas bodies induces capillary rupture in mice," Proc. Natl. Acad. Sci. U.S.A. **97**, 10179–10184.
Prosperetti, A. (**1991**). "The thermal behavior of oscillating gas bubble," J. Fluid Mech. **222**, 587–616.
Qin, S., and Ferrara, K. W. (**2006**). "Acoustic response of compliable microvessels containing ultrasound contrast agents," Phys. Med. Biol. **51**, 5065–5088.
Qin, S., Hu, Y., and Jiang, Q. (**2006**). "Oscillatory interaction between bubbles and confining microvessels and its implications on clinical vascular injuries of shock-wave lithotripsy," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1322–1329.
Rowe, A. J., Finlay, H. M., and Canham, P. B. (**2003**). "Collagen biomechanics in cerebral arteries and bifurcations assessed by polarizing microscopy," J. Vasc. Res. **40**, 406–415.
Sato, K., Tomita, Y., and Shima, A. (**1994**). "Numerical analysis of a gas bubble near a rigid boundary in an oscillatory pressure field," J. Acoust. Soc. Am. **95**, 2416–2424.
Shima, A., Takayama, K., Tomita, Y., and Miura, N. (**1981**). "An experimental study on effects of a solid wall on the motion of bubbles and shock waves in bubble collapse," Acustica **48**, 293–301.
Skyba, D. E., Price, R. J., Linka, A. Z., Skalak, T. C., and Kaul, S. (**1998**). "Direct in vivo visualization of intravascular destruction of microbubbles by ultrasound and its local effects on tissue," Circulation **98**, 290–293.
Snowhill, P., and Frederick, S. (**2005**). "A mechanical model of porcine vascular tissues-Part II: Stress-strain and mechanical properties of juvenile porcine blood vessels," Cardiovasc. Eng. **5**, 157–169.
Tipton, R. E., Steinberg, D. J., and Tomita, Y. (**1992**). "Bubble expansion and collapse near a rigid wall," JSME Int. J., Ser. II **35**, 67–75.
Tomita, Y., Robinson, P. B., Tong, R. P., and Blake, J. R. (**2002**). "Growth and collapse of cavitation bubbles near a curved rigid boundary," J. Fluid Mech. **466**, 259–283.
Vogel, A., Lauterborn, W., and Timm, R. (**1989**). "Optical and acoustic investigations of the dynamics of laser-produced cavitation bubbles near a solid boundary," J. Fluid Mech. **206**, 299–338.
Wible, J. H., Galen, K. P., Wojdyla, J. K., Hughes, M. S., Klibanov, A. L., and Brandenburger, G. H. (**2002**). "Microbubbles induce renal hemorrhage when exposed to diagnostic ultrasound in anesthetized rats," Ultrasound Med. Biol. **28**, 1535–1546.
Wolfrum, B., Mettin, R., Kurz, T., and Lauterborn, W. (**2002**). "Observations of pressure-wave-excited contrast agent bubbles in the vicinity of cells," Appl. Phys. Lett. **81**, 5060–5062.
Yamada, H. (**1970**). *Strength of Biological Materials* (Williams and Wilkins, Baltimore).
Young, F. R. (**1989**). *Cavitation* (McGraw-Hill, New York).
Zheng, H., Dayton, P. A., Caskey, C., Zhao, S., Qin, S., and Ferrara, K. W. (**2007**). "Ultrasound-driven microbubble oscillation and translation within small phantom vessels," Ultrasound Med. Biol. **33**, 1978–1987.
Zhong, P., Zhou, Y. F., and Zhu, S. L. (**2001**). "Dynamics of bubble oscillation in constrained media and mechanisms of vessel rupture in SWL," Ultrasound Med. Biol. **27**, 119–134.

Information theory analysis of Australian humpback whale song

Jennifer L. Miksis-Olds[a)]
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

John R. Buck
University of Massachusetts Dartmouth, North Dartmouth, Massachusetts 02747

Michael J. Noad
The University of Queensland, St. Lucia, Queensland 4072, Australia

Douglas H. Cato
Defence Science and Technology Organisation and University of Sydney, P.O. Box 44, Pyrmont, New South Wales 2009, Australia

M. Dale Stokes
Scripps Institution of Oceanography, La Jolla, California 92093

(Received 28 January 2008; revised 24 June 2008; accepted 8 July 2008)

Songs produced by migrating whales were recorded off the coast of Queensland, Australia, over six consecutive weeks in 2003. Forty-eight independent song sessions were analyzed using information theory techniques. The average length of the songs estimated by correlation analysis was approximately 100 units, with song sessions lasting from 300 to over 3100 units. Song entropy, a measure of structural constraints, was estimated using three different methodologies: (1) the independently identically distributed model, (2) a first-order Markov model, and (3) the nonparametric sliding window match length (SWML) method, as described by Suzuki *et al.* [(2006). "Information entropy of humpback whale song," J. Acoust. Soc. Am. **119**, 1849–1866]. The analysis finds that the song sequences of migrating Australian whales are consistent with the hierarchical structure proposed by Payne and McVay [(1971). "Songs of humpback whales," Science **173**, 587–597], and recently supported mathematically by Suzuki *et al.* (2006) for singers on the Hawaiian breeding grounds. Both the SWML entropy estimates and the song lengths for the Australian singers in 2003 were lower than that reported by Suzuki *et al.* (2006) for Hawaiian whales in 1976–1978; however, song redundancy did not differ between these two populations separated spatially and temporally. The average total information in the sequence of units in Australian song was approximately 35 bits/song. Aberrant songs (8%) yielded entropies similar to the typical songs. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2967863]

PACS number(s): 43.80.Ka [WWA] Pages: 2385–2393

I. INTRODUCTION

Male humpback whales (*Megaptera novaeangliae*) sing complex songs that are believed to be an important element of the humpback mating system (Winn and Winn, 1978; Tyack, 1981; Darling and Bérubé, 2001). Songs are sung predominantly on the winter breeding grounds and during migration to and from the breeding grounds, but may sometimes be heard on high-latitude feeding grounds (Clapham and Mattila, 1990; Cato, 1991; Clark and Clapham, 2004). Analyses of song from humpbacks of the Northern and Southern Hemispheres reveal a hierarchical structure of repeating themes in a largely stereotyped order (Payne and McVay, 1971; Guinee and Payne, 1988; Cato, 1991; Suzuki *et al.*, 2006). Themes consist of repeated phrases, which are composed of discrete sound elements called units. Unit durations can range from less than 1 s to longer than 5 s with an average of approximately 1 s. Spacing between the units is typically on the order of 2–5 s. Sometimes there are minor variations in frequency and timing between phrases and units within the same theme. The whales may repeat the song several times in the course of a song session, usually surfacing to breathe or "blow" in between consecutive songs. Song duration has been reported to range from 6 to 35 mins, whereas song sessions can last for up to 22 h (Payne and McVay, 1971; Winn and Winn, 1978).

The song structure of the Northern Hemisphere humpbacks has been studied more extensively than that of their Southern relatives. This is attributed to the difference in the geographic distribution of whales on breeding grounds between the two hemispheres (Cato, 1991). Whales on the Northern Hemisphere breeding grounds are relatively concentrated within close proximity to shore, whereas they are more dispersed during offshore migrations. The opposite is true in many parts of the Southern Hemisphere. Whales are more widely distributed on broad continental shelf breeding grounds (e.g., in Australia or Africa) or around scattered island groups (e.g., in Oceania), and known concentrations of

[a)]Electronic mail: jlm91@psu.edu

whales in Australia, in particular, are more easily observed on the migration route that closely follows the coastline in some places (Dawbin, 1966; Simmons and Marsh, 1986; Cato, 1991; Cato *et al.*, 2001; Noad and Cato, 2007; Noad *et al.*, 2004).

Males within the same breeding stock or population typically sing the same song; however, the songs are not sung synchronously. Males from different stocks and different oceans sing different songs (Payne, 1978; Winn and Winn, 1978) although songs used by populations within the same ocean basin are often similar (Payne and Guinee, 1983; Helweg *et al.*, 1990, 1998). Song exchange between the hemispheres has not been observed due to geographic separation and out of phase migration patterns (Cato, 1991). Information theory provides a mechanism to investigate and relate the sequence of song units from different humpback stocks where individual song units, phrases, and themes can be drastically different.

For the purposes of this study, song structure specifically refers to the sequence of units. It does not capture any information on the timing differences or subtle variations between occurrences of the same units. While we do not dispute the possibility that information may be contained in other song parameters such as differences in the duration of the same units, timing between units, frequency differences, etc., humpback song has evolved in an acoustically complex environment where signals are subjected to frequency dependent reverberation and attenuation. The limited results available analyzing the timing and frequency characteristics of song units indicate that there is relatively little variation in these features, making it unlikely that these features contain substantial information (Macknight *et al.*, 2001; Au *et al.*, 2005). Songs are only viable for conveying information about individuality or fitness if they exhibit reliably perceivable interindividual differences (Tyack, 1981; Macknight *et al.*, 2001). Signals traveling long distances underwater are likely to experience both distortions in timing and relative energies at different frequencies due to heterogeneities along the transmission path making these parameters unattractive for discriminating between individuals over long distances (Michalopoulou, 1997). Consequently, the sequence of units is the feature of the song most robust to transmission through the underwater acoustic channel as singer position, listener position, and oceanographic conditions vary. This does not negate the potential for other avenues of information transfer through humpback song. It merely acknowledges the sequence of units as the most robust channel for conveying information.

The information theory techniques in this study focus on the estimation of the *information entropy*, or simply entropy in the sequel (Shannon, 1948; Sec. 2.1 of Cover and Thomas, 1991). Intuitively, the entropy of a source is a measure of how much uncertainty the output sequence of the source contains. For the case when the source is a humpback whale singing a sequence of units, the entropy measures how unpredictable the next unit in the song is. The structure of humpback song implies constraints on the order of the units produced. Because these constraints make the next unit in the song predictable, songs that are highly structured have a reduced entropy. In contrast, songs composed of the same units but with less structure are more uncertain and less predictable, resulting in higher entropies for the sequence of units. The definition of mutual information (see Cover and Thomas, 1991, Theorem 2.4.1) implies that the entropy of a source establishes an upper bound on the amount of information conveyed by the source's output. A detailed analysis of humpback song using entropy estimates and correlation techniques showed that for whales singing on the Hawaiian breeding grounds, songs had strong structural constraints indicative of a hierarchical structure and conveyed a maximum amount of information of less than 1 bit/s (Suzuki *et al.*, 2006).

This study uses the same information theory techniques as Suzuki *et al.* (2006) in the analysis of song sequences produced by migrating Australian humpbacks. Entropy estimates were calculated using an independent identically distributed (iid) model, first-order Markov model, and nonparametric sliding window match length (SWML) method. In addition, both long-term and short-term correlations were computed for each song session. The objectives of the study were (1) to determine an upper bound on the amount of information potentially conveyed in the unit sequence of Australian whale song, (2) to determine if there is quantitative evidence that the songs are consistent with a hierarchical structure, and (3) to compare the entropy in the sequence of units produced by migrating Australian males to the sequence of units produced by males on the breeding grounds of Hawaii.

II. METHODS

Acoustic recordings of migrating humpback whales were made for six consecutive weeks off the coast of eastern Australia at Peregian Beach as part of the Humpback Whale Acoustic Research Collaboration (HARC) (Noad *et al.*, 2004; Thode *et al.*, 2006; Dunlop *et al.*, 2007). A distributed horizontal array of five moored hydrophone buoys provided 228 h of humpback song recordings from September to October 2003. The buoys were configured in a T-shape with buoys 1–3 oriented parallel to shore at a distance of 1.5 km. The buoys were spaced approximately 700 m apart. Buoys 4 and 5 were positioned in a straight line perpendicular to the shore running offshore to the east of buoy 2 and were separated by approximately 600 m. A VHF radio transmitter in each buoy transmitted the signal from a High Tech MIN96 hydrophone with a built-in +40 dB preamplifier to a Royal Australian Navy type 8101 four-channel sonobuoy receiver or a Winradio receiver. In the absence of a song at a reasonable signal-to-noise ratio (SNR), acoustic sampling occurred for 2 min every 15 mins. When whale song was detected at a SNR that allowed the sequence of units be clearly discernible, continuous recording commenced via a National Instruments E-series data acquisition card and was stored to a computer hard drive. All channels were recorded simultaneously by ISHMAEL software (D. Mellinger, NOAA) at 22 kHz. ISHMAEL also performed localizations of singing whales using standard hyperbolic techniques (Noad *et al.*, 2004).

A human classifier manually sequenced the units in each song session using COOL EDIT PRO software that provided simultaneous playback of the acoustic signal and visual spectrogram. Song units were assigned letters as symbols (A, B, C, …), and a dictionary of units specific to the eastern Australia humpback whale song in 2003 was created so that consistency was maintained over the course of the analysis. Biases associated with human classifiers were not strong enough to warrant the use of automated classifiers, as the work by Suzuki *et al.* (2006) showed that song structure analyzed by entropy estimation was robust to the choice of classifier. There was an approximate 5% disagreement between human classifiers in Suzuki *et al.* (2006), which indicated that classification was not always clear. Based on the low percentage of disagreement between human observers in the previous study, in addition to the substantial time and effort of employing multiple human classifiers, only one human classifier was utilized in this study. Units that stood out from the typical sequence of a phase or theme were double-checked for any error in classification or transcription.

Song sessions meeting the following criteria were considered for analysis: (1) uninterrupted, (2) a minimum of 300 consecutive units (approximately 20 min), and (3) independent with no overlap of other songs. If multiple song sessions were identified on the same day, acoustic localization and visual observations via theodolite were used to determine if multiple song sessions could be attributed to the same singer. In many instances song sessions from the same singer were interrupted or masked at times by passing ships. If multiple song sessions were identified as potentially coming from the same singer, only the longest song session of this singer was used in the analysis. A total of 48 song sessions ranging in length from 300 to over 3100 units were included in the analysis. This corresponded to song session durations of approximately 20 min–3.5 h.

The first phase of the analysis involved quantifying basic song session characteristics. Song alphabet size, statistical stationarity, and period were determined. A statistically stationary source has the same probability distribution at any point in time. Note that the statistical stationarity of the song should not be confused with the singing whale's physical movement, or lack thereof. This study focuses on the stationarity of singers in the statistical sense, not the physical sense, so references to stationarity in the remainder of the paper are understood to be in the statistical sense. The size of each song alphabet ($|A|$), or number of different discrete sound units, was tallied. Alphabet size limited the maximum entropy ($H_{\max}$) to $\log_2|A|$, which was attainable only if each classified unit was independently and uniformly distributed. The time invariance of the probability distribution of a stationary source implies that the autocorrelation function of a stationary source also does not vary with time. To check for stationarity, short-time (or partial) discrete autocorrelations were performed on the unit sequences at different points in the song session (Suzuki *et al.*, 2006) to determine if each song source (a singing humpback whale) was stationary.

The autocorrelations also provide a perspective on the ability of Markov sources to model humpback whale song. The probability distribution of a stationary source is the same

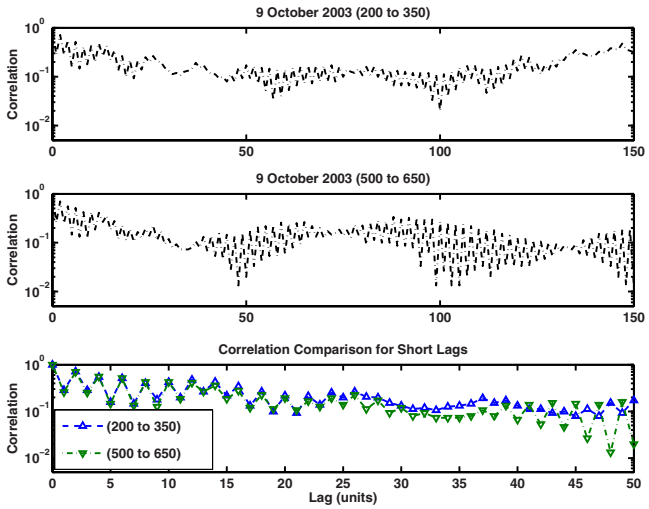


FIG. 1. (Color online) Two short-time autocorrelation estimates from song session 36. The top two panels plot the correlations for units in the ranges [200, 350] and [500, 650], respectively. Note that the correlation functions diverge substantially for lags greater than 100 units, indicating that the song is nonstationary. The third panel shows both correlations on the same axes for small lags. The values generally agree closely to each other with lags smaller than 30, indicating that the song may be considered locally stationary for lags less than 30 units.

at any point in time; consequently, the autocorrelation of a stationary source is the same at any point in the song. Any irreducible Markov model is a stationary source (Feller, 1960). The correlation function estimated from a finite realization of a song may fluctuate somewhat about the true correlation, but if the correlation function varies substantially at different points in the song, the source is not stationary and thus cannot be a Markov model. Partial song autocorrelation estimates were calculated using two nonoverlapping 151-unit song sections (Fig. 1). There was an order of magnitude divergence between the correlation functions for each song section, which indicated a nonstationary source. This lack of stationarity precludes a Markov source for the sequence of units in the song and implies that the song includes constraints beyond those embodied by a first-order or higher-order Markov model. However, comparison of the correlations showed that the songs could be considered locally stationary for lags less than 30 units. Local stationarity is a perquisite for entropy estimation using the SWML estimator. A window length of 30 units was used as the maximum window size for the SWML estimator in the second phase of analysis. Entire song session autocorrelation estimates, or global autocorrelations, were computed to determine the periodicity of each song (Fig. 2).

In the second phase of analysis, three different entropy estimators were applied to each song session: (1) the iid model (H_{iid}), (2) the first-order Markov model (H_{M}), and (3) the SWML estimator (H_{SW}). The mathematical theory underlying each method and a comparison of the methods are in Suzuki *et al.* (2006). The iid model is based on the assumption that each song unit is statistically independent and identically distributed. There is no memory requirement associated with this model. In a first-order Markov model, the current unit depends on the previous song unit but not on any earlier units. This model presumes a structure that has a

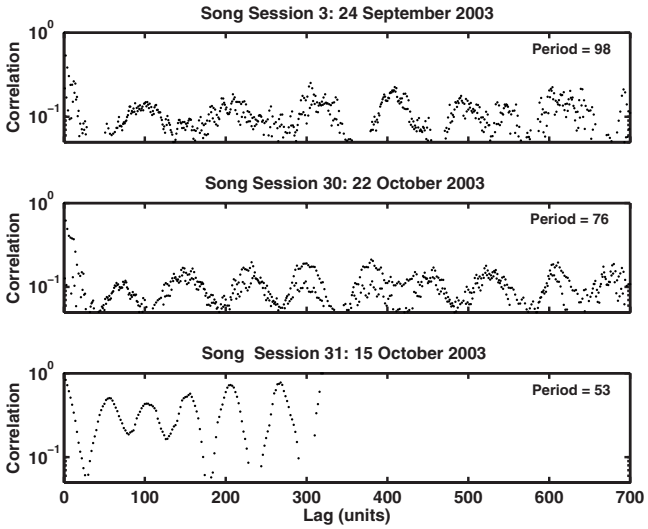


FIG. 2. Global autocorrelation estimates for three song sessions. Song session 3 had a song period of 98 units. Song session 30 had a song period of 76 units and song session 31 had a song period of 53 units.

memory of one unit. Most of our recordings lacked the length necessary to obtain accurate entropy estimates for higher-order Markov models (see Sec. I.B.5 of Suzuki *et al.*, 2006). Consequently, we limited our Markov model entropy estimation to first order and used the stationarity analysis discussed in the previous paragraph to rule out higher-order Markov models. The SWML entropy estimator relates the average length of a matching string of units within a fixed window to the entropy of the sequence (Wyner and Ziv, 1989). Based on the short-time correlation analysis, the range of window lengths used in this analysis was 10–40 units. The SWML estimator does not use a model and has the least restrictive assumptions, making it more applicable to a wider class of sources. The window size generating the minimum SWML entropy estimate for each song session is presented in parentheses after each value of H_{SW} in Table I. As noted in Sec. I.B.7 of Suzuki *et al.* (2006), the window size producing the minimum entropy estimate for the SWML estimator is a balance between a statistical bias if the window is too short and a violation of the local stationarity assumption if the window is too long.

We also used the SWML entropy estimate to estimate the redundancy (ρ) for each song session [Eq. (6) in Suzuki *et al.*, 2006]. Shannon (1948) defined redundancy to quantify the percentage that a message's length would be reduced when going from a coding using a naive representation, which requires H_{max} bits/symbol to the optimal coder, which compresses the sequence to the source entropy H bits/symbol (Theorem 2 in Suzuki *et al.*, 2006). For example, a plain ASCII file uses 8 bits for each character, but a compression program can often compress an ASCII file to be much smaller. The difference in the lengths of these files, normalized by the original length of the ASCII file, is the redundancy ρ. Because the SWML estimator has a positive bias, the resulting estimates of redundancy actually underestimate the redundancy and may be considered a lower bound on ρ.

The entropy estimates were used to make two hypothesis tests for each song. The first test compared H_{iid} with H_M. If H_M is significantly smaller than H_{iid} this indicates that the song source has sequential structure or memory, and knowing the current unit reduces uncertainty about the next unit. The second test compared H_M with H_{SW} for each song session. If H_{SW} is significantly smaller, this indicates the song source contains a structure whose memory extends beyond just one unit, as assumed by the first-order Markov model.

III. RESULTS

Table I summarizes the entropy data for all 48 song sessions analyzed. For each song session, the table presents the three entropy estimates H_{iid}, H_M, and H_{SW} measured in bits, as well as the song redundancy (ρ) for each song session. Song session duration is in units, the alphabet size is $|A|$, and the maximum possible entropy $H_{max}=\log|A|$. Table II reports the maximum total information that could be conveyed by the sequence of units in each song. The estimated period (length) of the song was obtained from the global autocorrelation function. The final column in Table II represents the maximum information that could be conveyed by the sequence of units in the song and is the product of the song period and H_{SW} reported in bits. The average of the maximum information that could be conveyed per song at the unit level equaled 34.5 bits/song.

Global autocorrelations showed oscillations relating to the periodicity, or song length, during each song session (Fig. 2). The average song length was 97 ($\pm$39) units. 4 of the 48 song sessions analyzed were considered to be composed of "aberrant" songs based on unit characteristics. In this study we defined aberrant songs as those substantially different from the typical song of the region at the unit level. This differs from the definition of aberrant song used by Frumhoff (1983), which discriminated at the level of song themes. Three of the songs we labeled as aberrant (song sessions 1, 16, and 45) used fewer than six of the units seen in the typical song. Song session 10 was considered aberrant because over 75% of the song units were not consistent with the units from the other songs recorded in the population and may have originated from another population, e.g., from the western Australian population. The songs of east and west coast males are typically very different due to differences in song units, but a small number of "foreign" singers have been previously cited as initiating a song exchange between the two populations (Noad *et al.*, 2000).

The results of the statistical analysis of the entropy estimates generated by the iid and Markov models are shown in Fig. 3. The confidence intervals were obtained by generating 1000 bootstrap sequences using an iid source whose distribution matched the observed distribution of symbols in the song (see Suzuki *et al.*, 2006 for details). The bound on the one-tailed 0.95 confidence interval associated with each circle reflects the 50th lowest of the first-order Markov model entropy estimates from the series of 1000 estimates obtained from the bootstrap sequences generated by the iid model. The confidence intervals do not overlap with any of the Markov entropy estimates for the actual songs leading to a rejection of the null hypothesis $H_{iid} \leqslant H_M$ for every song

TABLE I. Song session entropy estimates. $|A|$ is the song alphabet size. H_{max} is the limit of maximum entropy based on alphabet size. Window size for the SWML estimator is (n). The * denotes aberrant song sessions.

Song session	Date	Length (units)	A (units)	$\log_2 \|A\|$ H_{max}	iid H_{iid}	Markov H_M	SWML H_{SW} (n)	Redundancy ρ
1*	9/24/2003	506	4	2.00	1.57	0.64	0.38(10)	0.81
2	9/24/2003	392	11	3.46	2.80	1.21	0.62(10)	0.82
3	9/24/2003	854	12	3.58	2.95	1.12	0.57(15)	0.84
4	9/26/2003	972	12	3.58	2.03	1.08	0.65(10)	0.82
5	9/26/2003	1355	13	3.70	2.78	1.19	0.51(14)	0.86
6	9/26/2003	976	14	3.81	2.90	1.07	0.41(13)	0.89
7	9/28/2003	797	8	3.00	2.49	0.79	0.28(12)	0.91
8	9/30/2003	869	17	4.09	2.70	0.83	0.15(16)	0.96
9	9/30/2003	885	13	3.70	2.82	1.06	0.45(14)	0.88
10*	10/1/2003	434	10	3.32	2.47	1.18	0.50(14)	0.85
11	10/1/2003	1624	10	3.32	2.41	0.98	0.29(13)	0.91
12	10/1/2003	332	10	3.32	2.62	1.08	0.53(15)	0.84
13	10/1/2003	640	12	3.58	2.92	1.03	0.37(14)	0.90
14	10/4/2003	447	7	2.81	1.87	0.88	0.30(10)	0.89
15	10/4/2003	707	11	3.46	2.75	1.10	0.47(10)	0.86
16*	10/29/2003	286	3	1.58	1.48	0.48	0.21(10)	0.87
17	10/29/2003	594	9	3.17	2.44	0.94	0.29(13)	0.91
18	10/29/2003	547	6	2.58	2.23	0.93	0.28(10)	0.89
19	10/29/2003	798	13	3.70	2.72	1.13	0.34(13)	0.91
20	10/29/2003	3166	13	3.70	2.80	1.19	0.43(12)	0.88
21	10/28/2003	527	10	3.32	2.40	0.84	0.34(14)	0.90
22	10/7/2003	377	9	3.17	2.37	0.98	0.29(13)	0.91
23	10/7/2003	835	12	3.58	2.98	1.15	0.69(14)	0.81
24	10/27/2003	610	7	2.81	1.91	0.79	0.19(10)	0.93
25	10/30/2003	501	8	3.00	2.24	0.94	0.22(10)	0.93
26	10/10/2003	1400	9	3.17	2.24	0.81	0.23(13)	0.93
27	10/10/2003	954	9	3.17	2.64	1.05	0.35(10)	0.89
28	10/13/2003	352	12	3.58	2.82	0.98	0.44(15)	0.88
29	10/19/2003	649	11	3.46	2.62	0.88	0.35(13)	0.90
30	10/22/2003	1912	12	3.58	3.04	1.14	0.59(10)	0.84
31	10/15/2003	322	6	2.58	1.97	0.91	0.39(10)	0.85
32	10/15/2003	583	11	3.46	2.85	1.09	0.60(10)	0.82
33	10/14/2003	2311	12	3.58	2.59	0.98	0.36(14)	0.90
34	10/8/2003	646	12	3.58	2.61	1.01	0.28(10)	0.92
35	10/8/2003	2543	16	4.00	2.97	1.12	0.51(13)	0.87
36	10/9/2003	844	12	3.58	3.03	1.09	0.49(13)	0.86
37	10/9/2003	507	12	3.58	2.78	1.00	0.36(13)	0.90
38	10/9/2003	1026	12	3.58	2.56	0.91	0.30(13)	0.92
39	10/11/2003	628	11	3.46	2.43	0.92	0.28(14)	0.92
40	10/11/2003	333	11	3.46	2.90	1.21	0.47(10)	0.86
41	10/26/2003	685	6	2.58	1.76	0.65	0.20(10)	0.92
42	10/12/2003	1192	13	3.70	2.88	1.04	0.36(13)	0.9
43	10/20/2003	1112	11	3.46	2.63	0.93	0.45(14)	0.87
44	10/24/2003	446	13	3.70	3.06	1.13	0.51(10)	0.86
45*	10/24/2003	696	5	2.32	1.95	0.62	0.20(13)	0.91
46	10/21/2003	346	8	3.00	2.11	0.73	0.26(10)	0.91
47	10/23/2003	613	8	3.00	2.30	0.92	0.32(15)	0.89
48	10/23/2003	528	8	3.00	2.05	0.68	0.31(14)	0.90

session. This suggests that the iid model does not adequately represent the structure of the recorded humpback song.

Comparison of the entropy estimates generated by the Markov model for the actual song sequence (H_M) and SWML estimates (H_{SW}) is shown in Fig. 4. The 95% confidence interval on H_{SW} was calculated under the null hypothesis that the song was generated by a first-order Markov source. This confidence interval was obtained by generating 1000 bootstrap sequences using a first-order Markov source whose transition probabilities match the observed symbol transitions in the song (see Suzuki *et al.*, 2006 for details). The bound on the one-tailed 0.95 confidence interval associated with each square reflects the 50th lowest of the SWML entropy estimates from the series of 1000 estimates obtained from the bootstrap sequences generated by the Markov model. The confidence intervals overlap with only one of the

TABLE II. Total information in the unit sequence of song sessions. Bits is a measure of song information content calculated from the product of the song period and H_{SW}. The * denotes aberrant song sessions.

Song session	SWML H_{SW}	Period	Bits
1*	0.38	41	15.7
2	0.62	56	34.8
3	0.57	98	55.6
4	0.65	77	50.5
5	0.51	115	58.7
6	0.41	112	45.9
7	0.28	179	49.9
8	0.15	282	43.5
9	0.45	75	33.8
10*	0.50	120	60.4
11	0.29	119	34.9
12	0.53	62	32.7
13	0.37	90	33.3
14	0.30	72	21.6
15	0.47	79	37.1
16*	0.21	69	14.5
17	0.29	74	21.5
18	0.28	75	21.0
19	0.34	110	37.4
20	0.43	88	37.8
21	0.34	81	27.5
22	0.29	90	26.1
23	0.69	75	51.8
24	0.19	91	17.3
25	0.22	144	31.7
26	0.23	117	26.9
27	0.35	122	42.7
28	0.44	91	40.0
29	0.35	83	29.1
30	0.59	76	44.8
31	0.39	53	20.7
32	0.60	53	31.8
33	0.36	125	45.0
34	0.28	89	24.9
35	0.51	112	57.1
36	0.49	80	39.2
37	0.36	111	40.0
38	0.30	73	21.9
39	0.28	184	51.5
40	0.47	83	39.0
41	0.20	73	14.6
42	0.36	121	43.6
43	0.45	82	36.9
44	0.51	77	39.3
45*	0.20	110	22.0
46	0.26	100	26.0
47	0.32	79	25.3
48	0.31	73	22.6

SWML entropy estimates leading to a rejection of the null hypothesis $H_M \leqslant H_{SW}$ for almost every song session. This suggests that a first-order Markov model does not accurately model the structure of the recorded humpback song. This is consistent with our conclusion based on the short-term correlations that the song is not stationary and therefore cannot be modeled by a Markov model.

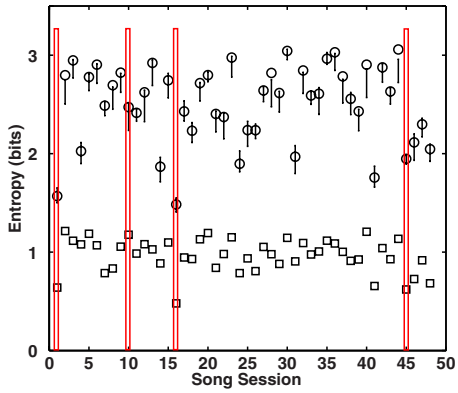


FIG. 3. (Color online) H_{iid} values (circles) and H_M values (squares) for all song sessions in Table I. The one-tailed 95% confidence intervals extending from each circle indicate the lower limit on H_M under the null hypothesis that the song was generated by an iid source. The confidence intervals do not overlap with any of the H_M values leading to a rejection of the null hypothesis $H_{iid} \leqslant H_M$ for every song. The vertical bars indicate aberrant song sessions.

Aberrant song sessions are indicated in Figs. 3 and 4 by bar markers. On average, the iid and Markov model entropy estimates of aberrant song sessions had entropy estimates less than most of the nonaberrant song sessions. However, the statistical comparison of the entropy estimates for the aberrant song sessions was consistent with that of the nonaberrant song sessions. This implies that the structure of the aberrant song sessions are not adequately modeled by either the iid or first-order Markov models, even though these sessions use a smaller set of units. Song session 10 showed no overall difference in entropy value or statistical significance compared to nonaberrant song sessions.

Figure 5 shows the relationship between H_{SW} and the song period. As a general trend, as the song period increased, the entropy decreased suggesting that longer songs are more predictable. The average amount of information per song (34.5 bits) is illustrated in Fig. 5 by the line $34.5=xy$.

IV. DISCUSSION

This work is consistent with the original proposal by Payne and McVay (1971), supported qualitatively by many

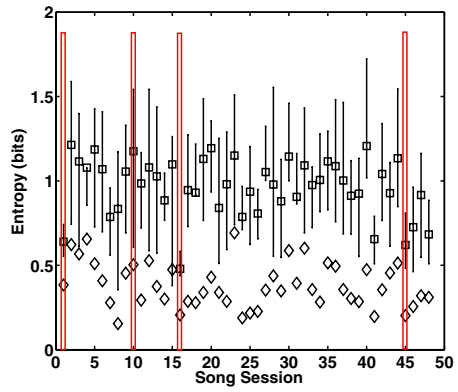


FIG. 4. (Color online) H_M values (squares) and H_{SW} values (diamonds) for all song sessions in Table I. The one-tailed 95% confidence intervals extending from each square indicate the lower limit on H_{SW} under the null hypothesis that the song was generated by a Markov source. The confidence intervals overlap with the H_{SW} value at song 15. The vertical bars indicate aberrant song sessions.

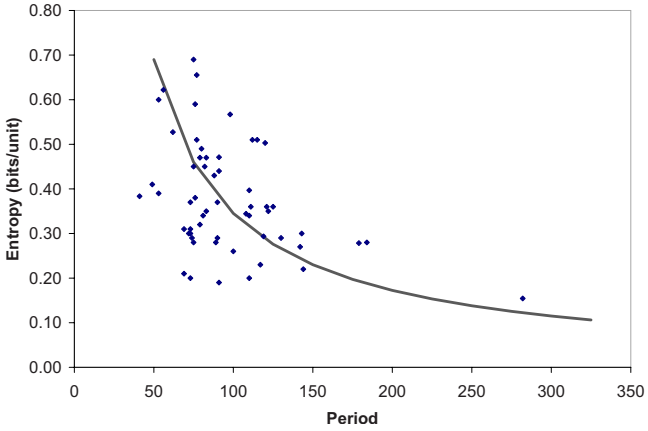


FIG. 5. (Color online) Source entropy as a function of song period. The solid line represents the best fit of a model where the average total information in a song (34.5 bits/song) is constant.

studies (e.g., Winn and Winn, 1978; Payne, 1983; Cato, 1991) and recently supported quantitatively by Suzuki *et al.* (2006), that the sequence of units of humpback song has a hierarchical structure. The periodicity of song within song sessions, ranging from 41 to 282 units, was illustrated in global autocorrelations (Fig. 2). Correlation analysis also indicated that humpback song is nonstationary; thus, there are more structural constraints in the whale song than the first-order or higher-order Markov models can accurately represent. Local stationarity was evident for lags less than 30 units, which supported the application of the SWML estimator. Bootstrap analysis of entropy values generated by the iid, first-order Markov model, and SWML estimator methods revealed that the SWML entropy estimates are significantly lower for the 2003 Australian humpback song produced during migration than both the iid and first-order Markov model estimates. The pattern of longer songs having lower entropy values and being more predictable is consistent with Guinee and Payne's (1988) suggestion that longer songs have a more redundant structure. This reduces the uncertainty about the next unit in any sequence and thus reduces the source entropy. Reducing the entropy while increasing the length of the song conserves the total information required to produce the song, at least in a loose sense. The proportion of aberrant song sessions composed approximately 8% of the song sessions analyzed; however, most of the aberrant song sessions yielded entropy estimates consistent with those of the typical song sessions recorded in the region.

Three of the four Australian migration songs classified as aberrant were due to low unit diversity. Although their entropy values did not differ from the values of the typical songs, their occurrence raises questions as to what the aberrant songs represent. It is possible that the aberrant songs are produced by juvenile males learning the song in which case they may be "age appropriate" rather than aberrant *per se*. Low unit diversity may also be a symptom of memory or language/speech disabilities similar to those occurring in humans (Cantwell and Baker, 1987; Gillam, 1998). As songs are believed to be an important element of the humpback mating system (Winn and Winn, 1978; Tyack, 1981; Darling and Bérubé, 2001), how the production of aberrant song affects reproductive fitness is unknown.

It is interesting to compare the results from this analysis with those reported for Hawaiian humpbacks on the breeding grounds in Suzuki *et al.* (2006), even though these populations are both spatially and temporally distant from each other. There are statistically significant differences between the entropies ($p<0.001$), song periods ($p<0.001$), and alphabet sizes ($p<0.001$) for the two populations (Wilcoxon rank-sum test for nonparametric data). The Australian songs had lower entropy values, shorter song periods, and smaller alphabet sizes. In addition, the average information content of the unit sequence in the 2003 eastern Australian humpback song produced during migration in this study (34.5 bits/song) was less than that reported by Suzuki *et al.* (2006) (130.3 bits/song) for humpback song produced on the breeding grounds of Hawaii in 1976–1978. In contrast, the redundancy ρ for the two populations was not significantly different (Wilcoxon rank-sum, $p=0.20$).

There are several potential explanations for the difference in information content between the two songs. First, it is possible that song content and structure differ extensively between different humpback populations. In general, the 1976–1978 Hawaiian songs analyzed in Suzuki *et al.* (2006) both were longer and had higher entropies than the 2003 Australian songs analyzed here. Given that both the entropies and song lengths are significantly different between the populations, it is not surprising that the total information content of the songs differ. Either increasing the length of the song or the entropy of the song will increase the total number of bits per song. Additionally, the Hawaiian songs used a larger number of units. What warrants attention is that in spite of all these changes, the redundancy did not vary much. This implies that the larger values of $H_{\max}$ for the Hawaiian songs also have proportionally larger entropies H. The entropy and correlation analyses indicate that both populations employ the same general structure for their songs, i.e., the hierarchical model proposed by Payne and McVay (1971). Given that both populations use the same general class of structure for their songs, the consistent redundancy estimates raise the interesting possibility that these temporally and spatially distinct groups have about the same level of structure constraints within their hierarchical songs. It would be interesting to compare with other populations to see if this is something fundamental in the song production centers of the humpback brain or simply coincidence. If song length and predictability are functions of population, while the redundancy remains roughly the same, the difference in selective pressures driving the difference in song structure between populations is of interest. Exploration of this would require quantitative analyses of songs from different populations over time and such studies, particularly including both Northern and Southern Hemisphere songs, have not yet been conducted.

A second explanation for these differences relating to possible song function is that differences in the unit sequence of song may be a result of seasonal behavior. The Hawaiian whales were recorded on the breeding grounds during winter, whereas the Australian whales were recorded during the

spring migration in transit to the feeding grounds. Change in song structure may reflect a shift in the seasonal behavioral state, which may or may not relate to differences in song function. That breeding ground song is innately more complex than migratory song, however, seems highly unlikely as there is no other evidence that the two are different. Indeed, in the South Pacific, east Australian migratory song and New Caledonia breeding ground song are usually almost identical despite being from different populations (Helweg *et al.*, 1998). If the breeding ground songs are longer, the song is often lengthened by repeating phrases more within a theme and not by adding units to phrases or new themes to the song. Migrating singers may sing shorter songs as a way of reducing the proportion of time singing compared with their counterparts on the breeding grounds if there is a greater cost in singing during migration. The increased energetic demands of actively swimming while singing during migration versus drifting on the breeding grounds (Noad and Cato, 2007) may have an effect, or migrating whales may minimize the time singing in order to minimize the delay in returning to the feeding grounds compared with the faster swimming nonsinging whales. However, a comparison of singing between the east Australian migration and the Hawaiian breeding grounds did not find a significant difference in the proportion of whales singing (Cato *et al.*, 2001), which suggests that there may be no significant difference in the proportion of time whales spent singing (assuming the proportions of singers in two populations were the same). Whales are also reported to sing on the feeding grounds (Clapham and Mattila, 1990; Clark and Clapham, 2004). Analysis of song from the same population during different seasons will be necessary to address the question of difference in song structure being attributed to seasonal behavior. As with innate population differences, however, song length itself is very unlikely to account for the full discrepancy between the east Australian and Hawaiian results.

A third explanation for the observed difference in entropy between the two songs is that it results from year by year variations in entropy within any population, and the Hawaiian song happened to be in a period of higher entropy and the Australian song in a period of lower entropy at the times of recording. The humpback whale song in any population changes with time, and this includes changes in pattern or structure. For example, recordings of migrating whales off the east coast of Australia indicated a period of relatively unstructured song in 1985 compared to songs in prior or following years, including a highly structured song in 1982 (Cato, 1991). Entropy analyses for a series of years for each of these populations would be required to determine the range in variation of entropy over a series of years and whether the differences between the two populations are significant compared to the temporal variation. This work illustrates the power of information entropy as a tool for objectively investigating the structure and information content of the sequence of units in humpback song. These techniques could be used to address temporal variation as suggested above and other interesting questions relating to humpback whale song.

Comparing the information and entropies between geographically and temporally distinct populations may be less informative than longitudinal studies of the same population. As reported above, less information was observed to be contained in the unit sequence of songs of migrating Australian whales in 2003 compared to Hawaiian whales on the breeding grounds in 1976–1978, while the redundancy of the songs was roughly constant. This indicates that there can be substantial variations in the information in the songs, which may be because of differences between populations, in behavior over the season, or as a result of year by year variation in the song. It raises questions associated with song function, selective pressures influencing whale song, and other potential avenues for information transfer within songs in addition to the unit sequence. Information theory techniques will be instrumental in quantitatively comparing song sequences produced among different populations, between whales during migration and at migration termini, and by whales in the same population in different years in order to address questions raised here. Exploration of song characteristics other than unit sequence and of the ability of these other characteristics to convey information through the underwater acoustic channel will further advance our understanding of humpback whale song.

ACKNOWLEDGMENTS

The authors thank Ryuji Suzuki for intellectual input and for sharing the entropy estimators. They also thank Rebecca Dunlop for providing theodolite data for singer verification, and the HARC volunteers for their long observations hours. The US Office of Naval Research (ONR) and the Australian Defence Science and Technology Organisation funded the HARC experiment. This analysis was supported by an American Association for University Women Fellowship to J.L.M.-O. and by grants from the Australian-American Fulbright Commission and ONR Grant No. N00014-00-1-0379 to J.R.B. This is Contribution No. 08-0101 from the UMass Dartmouth School for Marine Science and Technology.

Au, W. W. L., Lammers, M. O., Stimpert, A., and Schotten, M. (**2005**). "The temporal characteristics of humpback whale songs," J. Acoust. Soc. Am. **118**, p. 1940.

Cantwell, D. P., and Baker, L. (**1987**). *Developmental Speech and Language Disorders* (Guilford, New York, NY).

Cato, D. H. (**1991**). "Songs of humpback whales: The Australian perspective," Memoirs of the Queensland Museum **30**, 277–290.

Cato, D. H., Paterson, R., and Paterson, P. (**2001**). "Vocalization rates of migrating humpback whales over 14 years," Memoirs of the Queensland Museum **47**, 481–489, special issue on humpback whales.

Clapham, P. J., and Mattila, D. K. (**1990**). "Humpback whale songs as indicators of migration routes," Marine Mammal Sci. **6**, 155–160.

Clark, C. W., and Clapham, P. J. (**2004**). "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring," Proc. R. Soc. London, Ser. B **217**, 1051–1057.

Cover, T. M., and Thomas, J. A. (**1991**). *Elements of Information Theory* (Wiley, New York).

Darling, J. D., and Bérubé, M. (**2001**). "Interactions of singing humpback whales with other males," Marine Mammal Sci. **17**, 570–584.

Dawbin, W. H. (**1966**). "The seasonal migratory cycle of the humpback whale," in *Whales, dolphins, and porpoises*, edited by K. S. Norris (University of California Press, Berkeley, CA), pp. 145–170.

Dunlop, R. A., Noad, M. J., Cato, D. H., and Stokes, D. (**2007**). "The social vocalization repertoire of east Australian migrating humpback whales (*Megaptera novaeangliae*)," J. Acoust. Soc. Am. **122**(5), 2893–2905.

Feller, W. (**1960**). *An Introduction to Probability Theory and its Applications* (Wiley, New York, NY), Vol. **1**.
Frumhoff, P. (**1983**). "Aberrant songs of humpback whales (*Megaptera novaeangliae*): Clues to the structure of humpback songs," in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Boulder, CO), pp. 81–127.
Gillam, R. B. (**1998**). *Memory and Language Impairment in Children and Adults: New Perspectives* (Aspen, Gaithersburg, MD).
Guinee, L. N., and Payne, K. B. (**1988**). "Rhyme-like repetitions in songs of humpback whales," Ethology **79**, 295–306.
Helweg, D. A., Cato, D. H., Jenkins, P. F., Garrigue, C., and McCauley, R. D. (**1998**). "Geographic variation in South Pacific humpback whale songs," Behaviour **135**, 1–27.
Helweg, D. A., Herman, L. M., Yamamoto, S., and Forestell, P. H. (**1990**). "Comparison of songs of humpback whales (*Megaptera novaeangliae*) recorded in Japan, Hawaii, and Mexico during the winter of 1989," Sci. Rep. Cetacean Res. **1**, 1–20.
Macknight, F. L., Cato, D. H., Noad, M. J., and Grigg, G. C. (**2001**). "Qualitative and quantitative analyses of the song of the east Australian population of humpback whales," Memoirs of the Queensland Museum **47**, 525–537.
Michalopoulou, Z.-H. (**1997**). "Underwater transient signal processing: Marine mammal identification, localization and source signal deconvolution," International Conference on Acoustic, Speech, and Signal Processing, pp. 503–506.
Noad, M. J., and Cato, D. H. (**2007**). "Swimming speeds of singing and non-singing humpback whales during migration," Marine Mammal Sci. **23**, 481–495.
Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., and Jenner, K. C. S. (**2000**). "Cultural revolution in whale songs," Nature (London) **408**, p. 537.
Noad, M. J., Cato, D. H., and Stokes, D. (**2004**). "Acoustic tracking of humpback whales: Measuring interactions with the acoustic environment," Proceedings of ACOUSTICS 2004, Gold Coast, Australia, 3–5 November.
Payne, R. S. (**1978**). "Behavior and vocalizations of humpback whales (*Megaptera* sp)," Report No. MMC-77/03, pp. 56–77, U.S. Marine Mammal Commission, Washington, DC.
Payne, R. S. (**1983**). *Communication and Behavior of Whales* (Westview, Boulder, CO).
Payne, R. S., and Guinee, L. N. (**1983**). "Humpback whale (*Megaptera novaeangliae*) songs as an indicator of "stocks"," in *Communication and Behavior of Whales*, edited by R. Payne. (Westview, Boulder, CO), pp. 333–358.
Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 587–597.
Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.
Simmons, M. L., and Marsh, H. (**1986**). "Sightings of humpback whales in great barrier reef waters," Sci. Rep. Whales Res. Inst. **37**, 31–46.
Suzuki, R., Buck, J. R., and Tyack, P. L. (**2006**). "Information entropy of humpback whale song," J. Acoust. Soc. Am. **119**, 1849–1866.
Thode, A. M., Gerstoft, P., Burgess, W. C., Sabra, K., Guerra, M., Stokes, M. D., Noad, M., and Cato, D. H. (**2006**). "A portable matched-field processing system using passive acoustic time synchronization," IEEE J. Ocean. Eng. **31**, 696–710.
Tyack, P. L. (**1981**). "Interactions between singing Hawaiian humpback whales and conspecifics nearby," Behav. Ecol. Sociobiol. **8**, 105–116.
Winn, H. E., and Winn, L. K. (**1978**). "The song of the humpback whale *Megaptera novaeangliae* in the West Indies," Mar. Biol. (Berlin) **47**, 97–114.
Wyner, A. D., and Ziv., J. (**1989**). "Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression," IEEE Trans. Inf. Theory **35**, 1250–1258.

Three-dimensional transient and harmonic shear-wave scattering by a soft cylinder for dynamic vascular elastography

Anis Hadj Henni[a),b)]
Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center, Montreal, QC H2L 2W5, Canada

Cédric Schmitt
Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center, Montreal, QC H2L 2W5, Canada and Institute of Biomedical Engineering, University of Montreal, Montreal, QC H3C 3J7, Canada

Guy Cloutier[a),c)]
Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center, Montreal, QC H2L 2W5, Canada; Institute of Biomedical Engineering, University of Montreal, Montreal, QC H3C 3J7, Canada; and Department of Radiology, Radio-Oncology and Nuclear Medicine, University of Montreal, Montreal, QC H3T 1J4, Canada

(Received 17 April 2008; revised 18 July 2008; accepted 19 July 2008)

With the objective of characterizing biological soft tissues with dynamic elastography, a three-dimensional (3D) analytical model is proposed to simulate the scattering of plane shear waves by a soft cylinder embedded in an infinite soft medium. The 3D problem of harmonic plane shear-wave scattering is first formulated and solved, and the monochromatic solution is employed to simulate transient wave scattering. Both harmonic and transient simulations are compared with experimental 3D acquisitions. The good agreements obtained between measured and calculated displacement fields allowed to conclude on the validity of the proposed 3D harmonic and transient models. The spatial distribution of displacements (diffraction lobes, displacement oscillations, wave diffraction angles, etc.) and their relative amplitudes in both inclusion and surrounding materials depended on the contrast between the viscoelastic properties of the different media. The possibility of solving an inverse problem to assess soft heterogeneous medium viscoelasticity is discussed and some future theoretical and experimental developments are proposed.
 [DOI: 10.1121/1.2973194]

PACS number(s): 43.80.Qf, 43.80.Vj, 43.80.Ev, 43.35.Mr [FD] Pages: 2394–2405

I. INTRODUCTION

Detection and characterization of vascular pathologies are of great importance since they are responsible for more than one-third of global deaths caused by cardiovascular diseases in the world[1] (2002 World Health Organization statistics). In this context, the present work is dedicated to study, theoretically and experimentally, shear-wave (SW) scattering by mimicked venous thrombi[2] in order to enrich existing diagnostic methods, such as ultrasound imaging, x-ray angiography, and magnetic resonance imaging (MRI), by a new one based on dynamic elastography. Indeed, it has been proven that viscoelastic properties of coagulated blood depend on the clot age and composition.[3–5] This dependence is currently exploited to perform qualitative static elastography imaging (using a static or quasistatic loading to image the strain distribution in the medium)[6,7] but this technique is sensitive to mechanical boundary conditions and motion artifacts and does not allow to measure tissue viscosity. However, quantitative elasticity and viscosity imaging could become an interesting clinical indicator for pathology diagnosis and therapy planning.

Contrary to static methods, dynamic elastography[8,9] permits to perform viscoelastic characterization of living tissues by studying traveling of elastic SWs into the probed medium. Since SW speeds in soft tissues are very low (few meters per second) compared to compression waves, it is possible to track them with an ultrafast ultrasonic imaging system that can typically produce more than 5000 images per second. Compared to static methods, SW tracking allows to determine their velocity and attenuation independently of the mechanical boundary conditions. Moreover, the fast imaging system is not sensitive to typical motion artifacts since the propagation occurs during few milliseconds. Dynamic elastography allows to obtain the spatiotemporal displacement evolution and, consequently, the wave-medium interaction to perform tissue mechanical characterization. Here, one has to notice that contrary to elastic waves in solids, the total displacement field in tissues can be experimentally measured and imaged.

[a)]Authors to whom correspondence should be addressed.
[b)]Electronic mail: anis.hadjhenni@crchum.qc.ca
[c)]Electronic mail: guy.cloutier@umontreal.ca

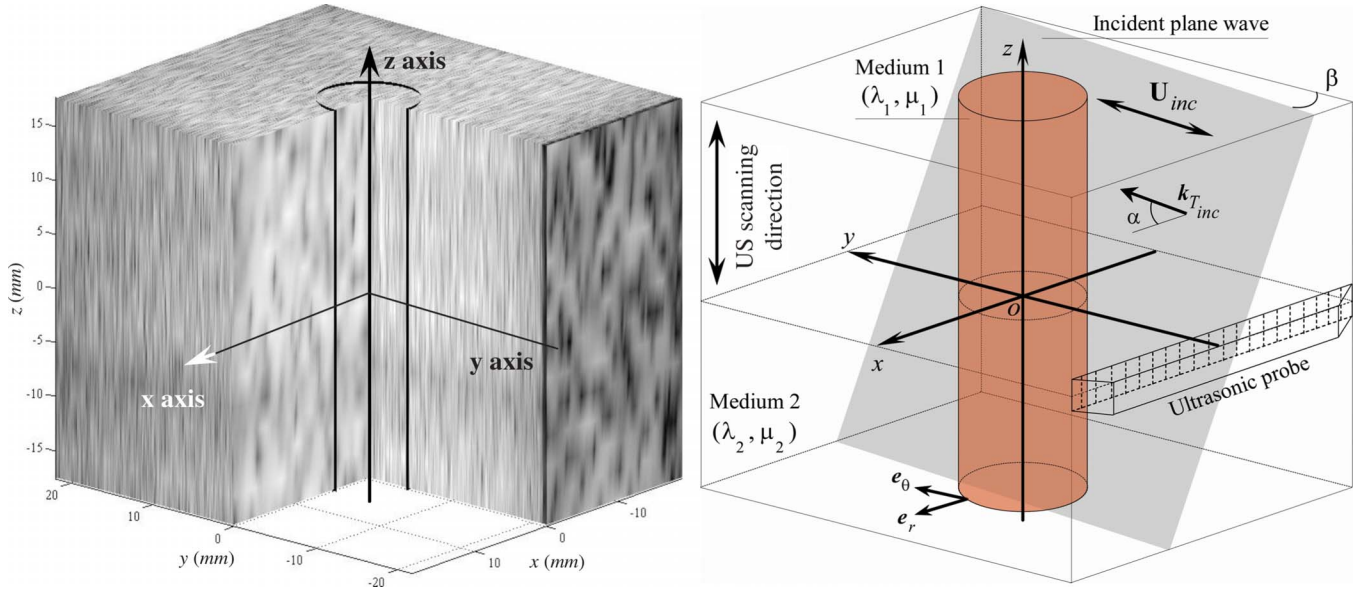


FIG. 1. (Color online) Left: 3D view of the reconstructed B-mode ultrasonic image of the experimentally tested heterogeneous medium. There is no echogenicity contrast between the agar-gelatin cylindrical inclusion (represented in the image) and its surrounding soft medium. Right: 3D representation of the plane SW scattering problem. The incident plane wave makes an angle of α with its propagation direction and is inclined in the plane $(o,\boldsymbol{e}_x,\boldsymbol{e}_y)$ from the y axis by an angle β.

A. Dynamic elastography in medicine

The first applications of SW imaging in bioacoustics were performed using M-mode ultrasonography[10] and Doppler for sonoelasticity,[11,12] MRI for magnetic resonance elastography,[13] or by means of real-time ultrafast ultrasonic imaging for both harmonic and transient elastographies.[14,15] Concerning transient elastography, most of modeling works concerned analytical or numerical simulation of SW generation obtained by impact methods[15] or by ultrasound radiation forces.[16,17] The interaction of SWs with confined heterogeneities in inhomogeneous tissues has not yet been precisely modeled despite the fact that to characterize mechanically many pathologies such as clotted vessels, tumors, etc., it is necessary to simulate and understand SW scattering by such inhomogeneities.

B. Objectives

An application of dynamic elastography to venous and arterial pathologies is presented here. The aim was to model analytically and to study, theoretically and experimentally, SW scattering by a venous clot, modeled by a cylindrical soft inclusion surrounded by an infinite soft medium. From a mechanical point of view, both inclusion and surrounding media were assumed to be made of homogeneous, isotropic, and linear viscoelastic materials.

A three-dimensional (3D) analytical model was developed to simulate the scattering of a harmonic plane SW by a cylindrical inclusion for an arbitrary incident angle. Following theoretical works of Faran,[18] White,[19] Fan *et al.*,[20] and Honarvar and Sinclair[21] on scattering of plane waves by cylindrical solids, the solution was formulated using a modal decomposition technique. The superposition principle of harmonic solutions (stationary displacement fields) served to obtain the 3D scattered field for an arbitrary incident transient plane SW. It is important to note that the model proposed in the present work can be coupled with magnetic resonance elastography or sonoelastography techniques.

Before validating experimentally the 3D model, the image acquisition process and the experimental setup that were used to produce ultrafast scanning are first described. The experimental material was composed of an agar-gelatin phantom (tissue mimicking material) containing a soft cylindrical inclusion, which simulated a venous clot. Both theoretical and experimental results are compared and discussed to investigate the validity of simulations. In the light of this discussion, a set of future possible applications and perspectives is presented, in particular, to solve the inverse problem allowing to characterize blood clot mechanical properties.

II. THEORY

To introduce the model, Fig. 1 shows a B-mode ultrasonic image of the 3D volume experimentally probed. The imaged soft agar-gelatin volume contained a circular cylindrical inclusion made of a mechanically different agar-gelatin material. One can observe that there is no echogenicity contrast between the inclusion and its surrounding medium despite their viscoelasticity differences. It will be shown (in Sec. IV) that a strong mechanical contrast appears between these two media when a SW propagates and interacts with the inclusion.

A. Problem formulation

The inclusion (blood clot phantom) was assumed to be a circular cylinder, of radius R, made of a soft material (medium 1) and surrounded by an infinite soft tissue (medium 2). Both media were assumed to be homogeneous, isotropic, linear, and viscoelastic. It is known that, for a certain frequency range (namely, for a few hundred hertz), the mechanical behavior of the agar-gelatin material is governed by the Hooke–Voigt viscoelastic behavior law.[22] This is in agree-

ment with a preliminary experimental work[23] that also suggested that coagulated blood viscoelasticity follows the same Hooke–Voigt model. This latter relies on the frequency domain, the stress, $\boldsymbol{\sigma}$, and the strain, $\boldsymbol{\varepsilon}$, tensors in the following way:

$$\boldsymbol{\sigma} = (\mathbf{C}' + i\omega\mathbf{C}'')\boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{C}'$ and $\mathbf{C}''$ are the stiffness and viscosity tensors, respectively. For an isotropic and viscoelastic material, the complex stiffness tensor, $\mathbf{C}=\mathbf{C}'+i\omega\mathbf{C}''$, depends on the complex Lamé coefficients λ and μ. According to this, and in order to easily take into account the viscosity, the elastodynamic problem was expressed in the frequency domain. Since studied materials are soft, the bulk moduli λ of both media are not significantly different and were chosen to ensure a compression wave speed equal to 1540 m/s (ultrasound speed in biological tissues) for tissue densities of 1100 $\mathrm{kg/m^3}$. Consequently, scattering of SWs in such materials depends strongly on the complex shear moduli.

The problem geometry, given in Fig. 1, was described in a Cartesian system of coordinates $(o,\boldsymbol{e}_x,\boldsymbol{e}_y,\boldsymbol{e}_z)$ but regarding the cylinder circular geometry and to simplify the expression of mechanical boundary conditions, the elastodynamic problem was expressed and solved in a cylindrical system of coordinates $(o,\boldsymbol{e}_r,\boldsymbol{e}_\theta,\boldsymbol{e}_z)$. The inclined incident plane SW, represented in Fig. 1, propagates with an incidence angle α, formed by the incident shear wavevector $\boldsymbol{k}_{T_{\mathrm{inc}}}$ and the polarization plane $(o,\boldsymbol{e}_x,\boldsymbol{e}_y)$. Finally, and for more generality, the incident wave polarization, $\mathbf{U}_{\mathrm{inc}}$ contained in the plane $(o,\boldsymbol{e}_x,\boldsymbol{e}_y)$, was doing an angle β with the y axis, see Fig. 1.

From these geometrical considerations, one can conclude that the scattering problem is 3D and that mode conversions of SWs can occur. Consequently, diffracted waves were expressed as a combination of compression and SWs in both media. The incident plane SW was first considered to be harmonic with an angular frequency ω. Then, the harmonic solution was derived to model the scattering of arbitrary transient plane SWs.

B. The stationary problem

In the following, the subscript 1 is assigned to the cylinder constitutive material, whereas the subscript 2 designates the surrounding medium. In addition, the longitudinal (compression) and transverse (shear) wavevectors in the medium j ($j=\{1,2\}$) are labeled by $\boldsymbol{K}_j$ and $\boldsymbol{k}_j$, respectively. Both of these wavevectors are decomposed into components parallel to the z axis, called $K_{j_z}=K_j \sin\alpha$ and $k_{j_z}=k_j \sin\alpha$, and other ones contained in the plane $(o,\boldsymbol{e}_r,\boldsymbol{e}_\theta)$ (perpendicular to the cylinder axis) given by $K_{j_\perp}=K_j\cos\alpha$ and $k_{j_\perp}=k_j\cos\alpha$.

The displacement field in both media satisfies the Navier elastodynamic wave equation.[24] Applying a Fourier transform to this equation, one obtains

$$\rho_j\omega^2\mathbf{U}_j + (\lambda_j + 2\mu_j)\,\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{U}_j) - \mu_j\,\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{U}_j) = \mathbf{0} \quad \text{with} \quad j=\{1,2\}, \tag{2}$$

where $\mathbf{U}_j=(U_{r_j}\ U_{\theta_j}\ U_{z_j})^T$ is the displacement field in a phase $j=\{1,2\}$, ρ_j is the density, and λ_j and μ_j are the complex viscoelastic Lamé coefficients. The Helmholtz decomposition technique permits to express the displacement field in both media into scalar and vector potentials:[25]

$$\mathbf{U}_j = \boldsymbol{\nabla}\varphi_j + \boldsymbol{\nabla}\times(\psi_j\boldsymbol{e}_z) + R\,\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times(\chi_j\boldsymbol{e}_z) \quad \text{with} \quad j=\{1,2\}. \tag{3}$$

Mechanically, $\varphi_j(r,\theta)$ is the displacement scalar potential associated with compression waves in phase j and $\psi_j\boldsymbol{e}_z$ and $\chi_j\boldsymbol{e}_z$ are the vector potentials of displacements polarized following the $(o,\boldsymbol{e}_r,\boldsymbol{e}_\theta)$ plane and another parallel to the z axis, respectively. Each of these potentials satisfies the well known Helmholtz wave equation.

Following the classical modal decomposition method, these potentials can be expressed as infinite series of Bessel and angular functions containing unknown coefficients, except for the known incident purely SW contained in the plane $(o,\boldsymbol{e}_r,\boldsymbol{e}_\theta)$. This latter depends only on the vector potential ψ_{inc} ($\mathbf{U}_{\mathrm{inc}}=\boldsymbol{\nabla}\times\psi_{\mathrm{inc}}\boldsymbol{e}_z$). Consequently, we can choose to express the shear incident field by its potential, which is written in a cylindrical system of coordinates as[18–21]

$$\psi_{\mathrm{inc}} = \mathrm{amp}(\omega)\sum_{n=0}^{+\infty}\frac{1}{\psi_0}\varepsilon_n(i)^n J_n(k_{2_\perp}r)\times\cos(n(\theta-\beta))e^{(ik_{2_z}z-i\omega t)}, \tag{4}$$

where ε_n is the Neumann factor, i.e., $\varepsilon_0=1$ and $\varepsilon_n=2$ for $n\geq 1$. One has to note that the incident wave's inclination angle β, in the plane $(o,\boldsymbol{e}_r,\boldsymbol{e}_\theta)$, acts like a rotation angle applied to the wave corresponding to $\beta=0$ (i.e., an incident wave propagating following the x axis). The incident wave amplitude, $\mathrm{amp}(\omega)$, is fixed to 1 for harmonic incident waves and is equal to the excitation amplitude (complex amplitude) in the case of a transient incident wave (see Sec. II C). The normalization coefficient, ψ_0, ensures that the maximum amplitude of the incident harmonic wave is equal to unity regardless of the frequency.

In a similar way, transmitted and reflected displacement fields can be expressed via their potentials. Considering the fact that the displacement at the cylinder center is finite, only the first kind Bessel functions $J_n(\cdot)$ serve to express the refracted potentials. The three displacement potentials within the inclusion are then given by

$$\varphi_1 = \sum_{n=0}^{+\infty} A_n J_n(K_1 r)\sin(n(\theta-\beta))e^{ik_{2_z}z},$$

$$\psi_1 = \sum_{n=0}^{+\infty} B_n J_n(k_1 r)\cos(n(\theta-\beta))e^{ik_{2_z}z},$$

$$\chi_1 = \sum_{n=0}^{+\infty} C_n J_n(k_1 r)\sin(n(\theta-\beta))e^{ik_{2_z}z}. \tag{5}$$

In these expressions and in the following, the harmonic time dependence term is omitted for simplification.

The displacement potentials in the surrounding medium are composed of those of the incident plane wave Eq. (4) and of the scattered ones. Knowing that scattered waves are out-

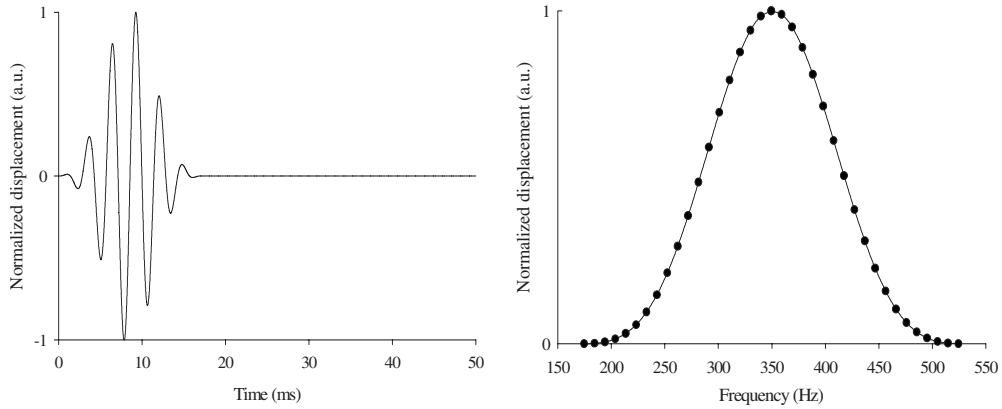


FIG. 2. Left: incident transient wave temporal form used in experiments and simulations. The wave is composed of six modulated pulses with 350 Hz central frequency. Right: the discretized incident wave spectrum representing the calculation points (●).

going, they are expressed by means of first kind Hankel functions $H_n^{(1)}(\cdot)$. One finally obtains the following scattered potential expressions:

$$\varphi_2 = \sum_{n=0}^{+\infty} D_n H_n^{(1)}(K_{2_\perp} r)\sin(n(\theta-\beta))e^{ik_{2_z}z},$$

$$\psi_2 = \sum_{n=0}^{+\infty} E_n H_n^{(1)}(k_{2_\perp} r)\cos(n(\theta-\beta))e^{ik_{2_z}z},$$

$$\chi_2 = \sum_{n=0}^{+\infty} F_n H_n^{(1)}(k_{2_\perp} r)\sin(n(\theta-\beta))e^{ik_{2_z}z}. \quad (6)$$

One can note that the axial wave number k_{2_z} is the same in both media. It has to be equal to the axial component of the incident SW since stress and displacements are continuous at the cylinder interface. Moreover, it is important to note that the angular dependencies in Eqs. (5) and (6) were chosen to be in agreement with the symmetry and antisymmetry properties of the different elastic waves (i.e., the incident and the scattered ones).[19]

At the cylinder boundary, the continuity of displacements and normal stress (given by $\boldsymbol{\sigma}\cdot\boldsymbol{n}$, where $\boldsymbol{\sigma}$ is the stress matrix and $\boldsymbol{n}$ is the unit vector normal to the cylinder) permits to determine the unknown coefficients A_n, B_n, C_n, D_n, E_n, and F_n in Eqs. (5) and (6). These conditions are summarized as follows:

$$\begin{aligned}&(U_{r1}\ \ U_{\theta 1}\ \ U_{z1})^T = (U_{r2}\ \ U_{\theta 2}\ \ U_{z2})^T\\ &(\sigma_{rr1}\ \ \sigma_{r\theta 1}\ \ \sigma_{rz1})^T = (\sigma_{rr2}\ \ \sigma_{r\theta 2}\ \ \sigma_{rz2})^T \quad \text{at} \quad r\\ &\quad = R \ \text{ and } \ \theta \in [0,2\pi].\end{aligned} \quad (7)$$

Using Eqs. (1) and (3), displacement and stress fields expressed at $r=R$ in Eq. (7) can be written with respect to potentials given in Eqs. (5) and (6). The orthogonality property of angular functions over the interval $[0,2\pi]$ allows to separate the boundary conditions in Eq. (7) into an infinite set of equations with respect to the infinite series orders n $(n=0,\ldots,+\infty)$. For a given mode of order n, it appears, after calculations, that the unknown coefficients A_n, B_n, C_n, D_n, E_n, and F_n are related to a set of known terms, which depend on the incident SW, by means of a linear system of equations:

$$T_n(A_n\ \ B_n\ \ C_n\ \ D_n\ \ E_n\ \ F_n)^T = \mathbf{b}_n \quad \text{with} \quad r=R \ \text{ and } \ n = 0,\ldots,+\infty. \quad (8)$$

In Eq. (8), the (6×6) T_n matrix depends on the geometry and the mechanical properties of media 1 and 2, while the second member $\mathbf{b}_n$ represents the incident wave contribution to the nth displacement and stress terms, expressed at the cylinder boundary. Both $\mathbf{T}_n$ and $\mathbf{b}_n$ elements are detailed in the Appendix. Solving the system of Eq. (8), by using a classical matrix inversion algorithm[26] to avoid fastidious calculations using the Cramer rule (which formulates solutions explicitly), provides for each integer order n the searched coefficients. These latter permit then to calculate the total stationary displacement field by means of expressions (4)–(6) and (3). Since one cannot calculate all terms of the infinite series appearing in the displacement field expression, the maximum calculation order is limited to a finite order. The truncation order, which depends on the inclusion size and on the incident wave frequency, was fixed for the simulations presented here to 50 terms.

C. Transient-wave scattering

Following the superposition principle in the spectral domain,[24] transient SW scattering by a homogeneous soft circular cylinder was modeled using the previous harmonic wave scattering model. Indeed, any incident wave, having an arbitrary temporal profile, can be expressed in the frequency domain by means of a temporal Fourier transform. In practice, the relevant frequencies are frequently limited to a certain interval with no significant information outside it (see the example in Fig. 2). Consequently, by choosing to express the transient wave scattering problem in the frequency domain, one can discretize the frequency range of interest into a finite number of calculation points (Fig. 2) and solve, for each frequency-wave amplitude couple $(\omega,\text{amp}(\omega))$ in Eq. (4), the associated harmonic problem using the model previously presented. This set of stationary solutions provides the diffracted transient wave spectrum. The temporal diffracted field was finally obtained from the spectral solution by

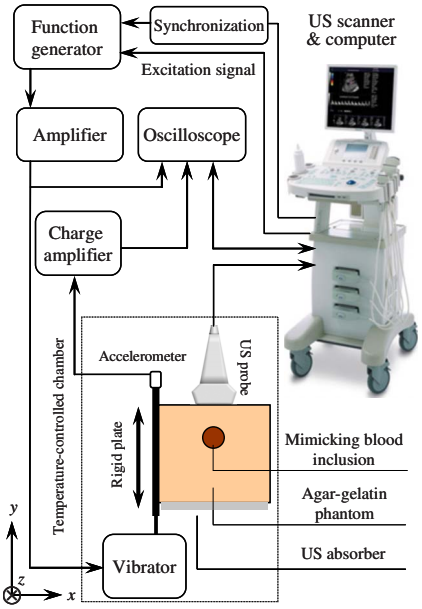


FIG. 3. (Color online) A schematic representation of the experimental setup: the SW generation and the ultrasonic acquisition devices.

means of an inverse Fourier transform. Compared to a temporal solving strategy, this spectral method is faster and permits to model more easily viscoelastic effects (since temporal convolution products appearing when viscosity is modeled are transformed into simple products in the frequency domain).

III. EXPERIMENTAL SETUP

A. Generation of shear waves and ultrafast acquisition system

In the experimental setup schematized in Fig. 3, SWs were generated by a large rigid plate (202×85 mm^2) attached to a vibrator (type 4810, Brüel&Kjær, Nærum, Denmark) and applied on the phantom surface. The low frequency harmonic or transient vibration was produced with a function generator (model 33250A, Agilent, Palo Alto, CA), amplified (type 2706, low frequency amplifier, Brüel&Kjær) before supplying the minishaker. In this configuration, the propagation direction of the plane SWs was orthogonal to the plate motion. A clinical array transducer (model L14-5/38, 38 mm width, 128 elements, Ultrasonix) connected to the Sonix RP scanner (Ultrasonix Medical Corporation, Burnaby, BC, Canada) was positioned parallel to the tissue motion to acquire and reconstruct radio-frequency (RF) sequences at a high frame rate (3850 Hz). The ultrafast imaging method used here was inspired from an electrocardiogram-gated image acquisition strategy to reach high frame rates.[27,28] In the present case, synchronization was achieved using SW gating by considering the starting of the mechanical excitation.

The probe excitation frequency, the sampling frequency, and the bit depth were 10 MHz, 40 MHz, and 16 bits, respectively. To avoid reverberation artifacts, an acoustical absorber was placed on the front of the probe. The true plate motion was acquired with an accelerometer (type 4375, Brüel&Kjær) connected to a charge amplifier (type 2692, Nexus Amplifier, Brüel&Kjær). The part of the experimental setup including the phantom was enclosed in a temperature-controlled chamber regulated at 20 °C.

The agar-gelatin phantom was made following a protocol described in a previous work.[29] The surrounding material was made (in proportion of the water weight) of 4.0% porcine skin gelatin and 3.0% agar powder (product No. G-1890 for gelatin and No. A-9799 for agar, Sigma Chemical, St. Louis, MO) and contained an agar-gelatin cylindrical inclusion composed of 2.5% porcine skin gelatin and 1.0% agar. The whole phantom had a parallelepipedal geometry ($20.5 \times 27.5 \times 10.5$ cm^3) and the inclusion was a 9.8 mm diameter cylinder. This latter formed an angle of $\alpha = 46°$ (i.e., the incidence angle) with the moving plate (positioned vertically) and had a length equal to 15.5 cm.

To image the 3D displacement field, RF ultrasonic signals at 14 consecutive planes (38.0×40.0 mm^2) were acquired every 2.5 mm along the z axis (for a 35.0 mm scanning distance) by adjusting automatically the probe position with a positioning step motor. A normalized cross-correlation algorithm applied to the acquired RF signals allowed to obtain the y displacement field component and its temporal evolution for the different slices. Finally, the 3D displacement field was reconstructed by superposing, along the z axis, the two-dimensional (2D) consecutive measured fields. In what follows, one has to note that both experimental and simulated displacement fields [in 3D, 2D and one-dimensional (1D) representations] have been normalized by their respective maximum displacement value.

B. Viscoelasticity of phantom components

Since mechanical properties have to be entered into the model to calculate displacement fields and in order to validate simulations, i.e., to solve the forward problem, it was necessary to assess viscoelastic properties of the gel materials that were used. The viscoelastic properties of phantom materials were determined by using 2D dynamic elastography[22] based on our ultrafast imaging system. Using this approach, the viscoelasticity was obtained by solving an inverse problem[22] based on the phase velocity and attenuation of plane SWs propagating in a homogeneous gel sample. A multifrequency study (using harmonic plane waves at different frequencies) allowed to assess the Hooke–Voigt viscoelastic parameters.

The two agar-gelatin materials (surrounding medium and inclusion) were characterized by using a separate homogeneous and parallelepipedal phantom for each of them. Viscoelastic properties were assessed between 50 and 440 Hz for the surrounding material and from 50 to 360 Hz for the inclusion material. Since this latter was very soft and viscous, high frequency plane SWs did not propagate sufficiently to characterize the inclusion material at frequencies above 360 Hz (as it can be shown by simulation, such a problem is avoided when the soft material is confined, for example, in a cylindrical form into a harder one). The experimental viscoelastic mean values and relative errors (i.e., the ratio between the standard deviation and the mean value) for the two samples were obtained by assessing plane SW veloc-

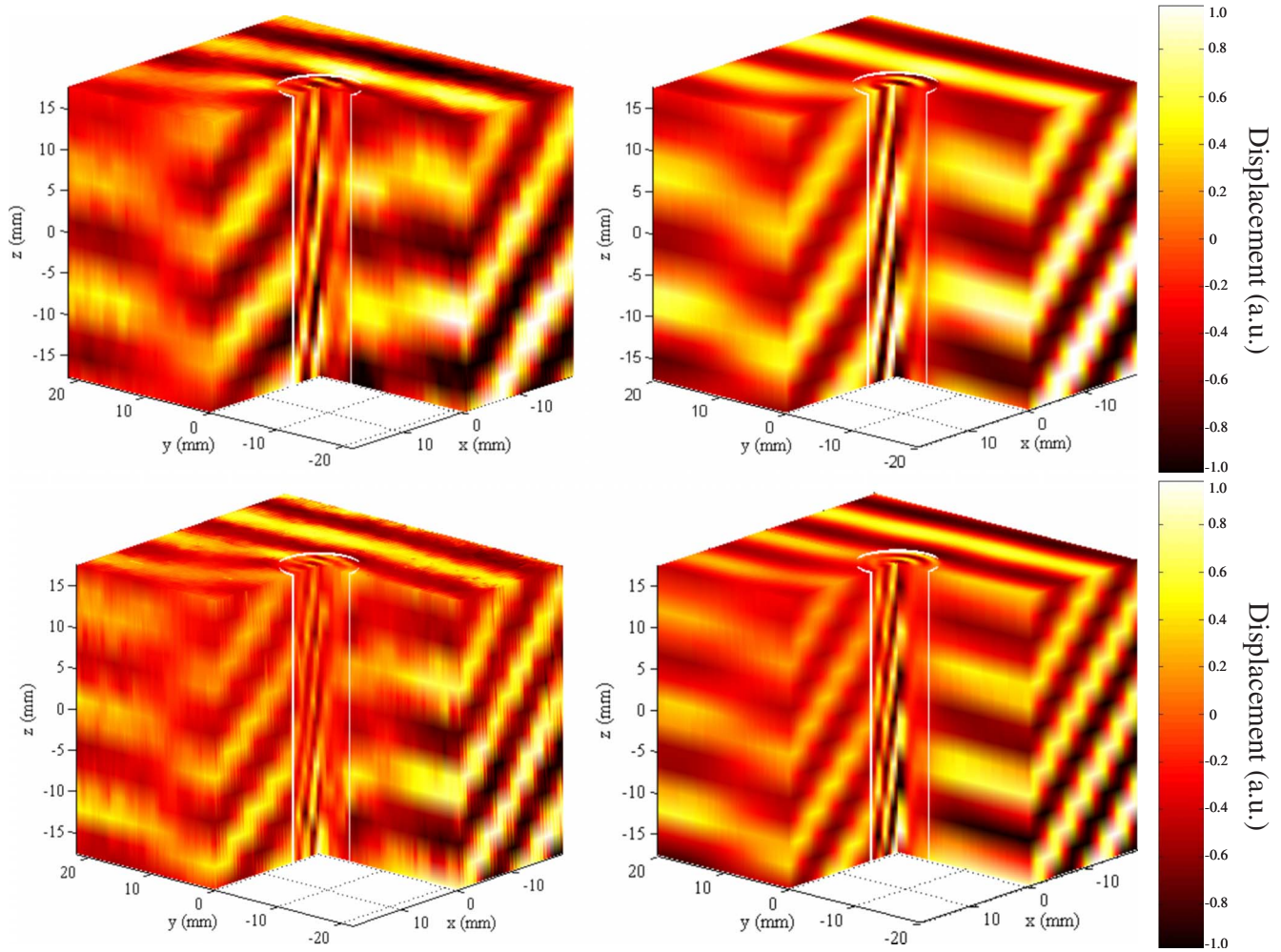


FIG. 4. (Color online) 3D representation of experimental (left) and simulated (right) normalized stationary displacement fields with $\alpha=46°$. The top and bottom images correspond to the scattering of 350 and 450 Hz harmonic incident waves, respectively.

ity and attenuation along 80 lines (of 38 mm width) at different depths. Viscoelasticity mean values were $\mu=(14600+0.7i\omega)$ Pa and $\mu=(1150+0.035i\omega)$ Pa for the surrounding medium and inclusion, respectively. Relative experimental errors were equal to $\pm 2.4\%$ and $\pm 14.0\%$ for elasticity and viscosity, respectively.

IV. EXPERIMENTAL VALIDATION OF THE 3D MODELS

A. Monochromatic case

The first experimental validation consisted in comparing experimental and theoretical stationary displacement fields due to the scattering of a monochromatic incident SW. This validation is necessary since the harmonic simulation also served to model a transient incident wave scattering. In addition, monochromatic excitation of soft tissues is commonly used in magnetic resonance elastography and in Doppler sonoelastography. Two harmonic excitations, 350 and 450 Hz, have been chosen to verify the validity of the harmonic model. This choice has been made because a large bandwidth was covered by the transient wave experimentally generated (see Fig. 2).

Figure 4 presents a 3D view of experimental and theoretical normalized displacements for the two harmonic excitations. A quarter volume was cut out from the entire scanned (or calculated) space in order to better visualize the displacement field into the inclusion. As one can see, a great similarity is noted between displacements within the inclusion and those within the surrounding medium for the two incident waves. One can also notice that the correspondence between simulations and measurements is not satisfied everywhere in the 3D space. In particular, it appears that the experimental frontwaves are not always strictly planar and that phase and amplitude shifts exist in certain regions. In order to compare more rigorously simulations with experiments and study the differences, the stationary fields were compared, first, following a plane Π (namely, the plane $z=-17.5$ mm), then along a line (1D) parallel to the x axis and contained in the plane Π, and finally along the z axis (i.e., the cylinder axis).

Figure 5 shows the 2D normalized stationary displacement fields in the plane Π for the two harmonic incident waves. The horizontal striations (jitter errors) appearing in the experimental displacements correspond to the piezoelectric elements of the ultrasonic linear array. It is noticeable that measurements contain experimental noise due to the relatively small amplitude of displacements at high frequencies (at lower frequencies, strong displacement modulations do not appear within the inclusion). Despite the presence of noise, one can see the good agreement between measured and theoretically predicted fields. For this example, the ex-

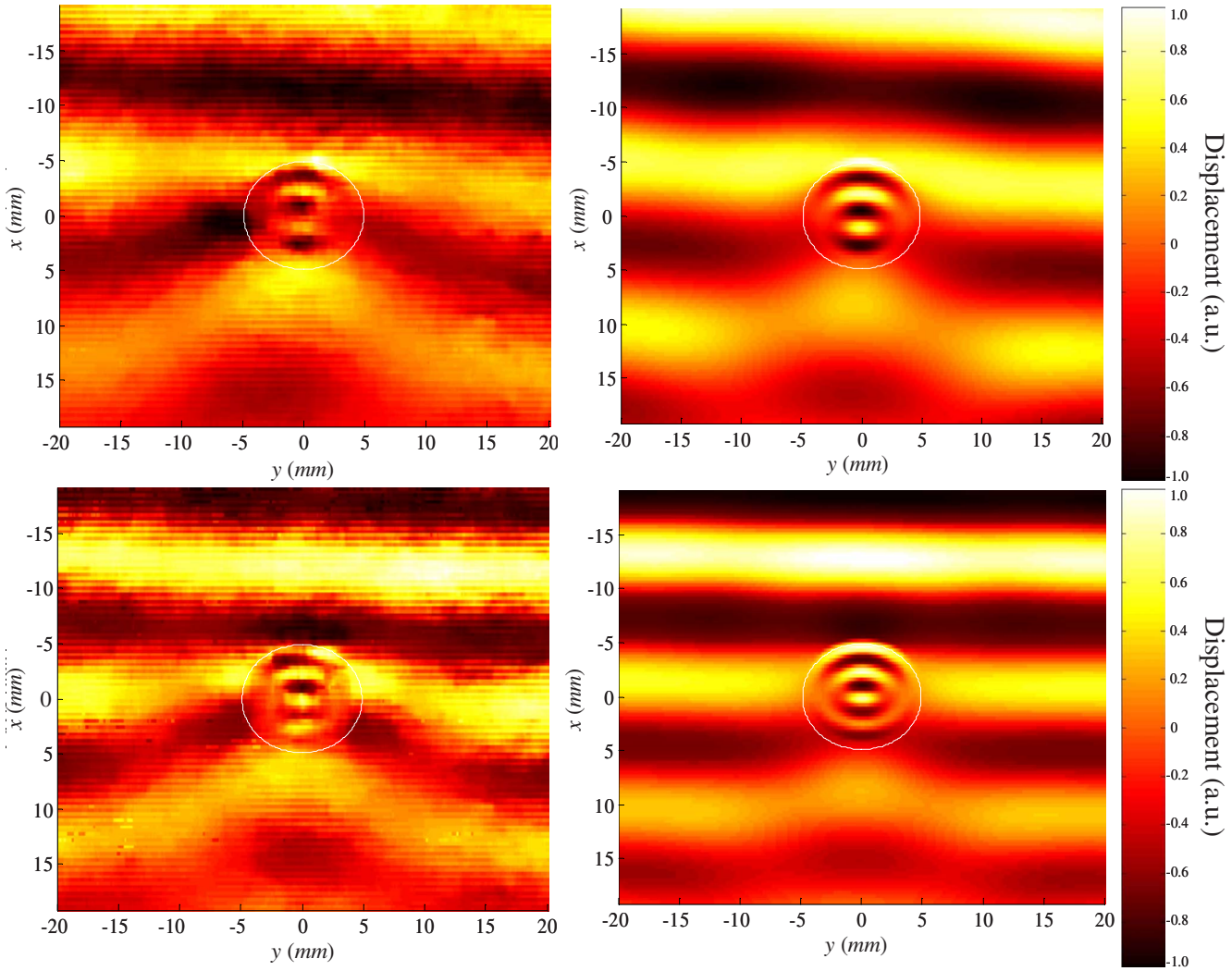


FIG. 5. (Color online) 2D representation of experimental (left) and simulated (right) normalized stationary displacement fields in the plane Π (corresponding to $z=-17.5$ mm). The top and bottom images correspond to the scattering of 350 and 450 Hz harmonic incident waves, respectively. The measured inclination β, in the plane $(o,\boldsymbol{e}_x,\boldsymbol{e}_y)$, equals −3°.

perimental incident wave was inclined, in the plane $(o,\boldsymbol{e}_x,\boldsymbol{e}_y)$, by an angle $\beta=-3°$ relative to the x axis. Figure 5 shows that this inclination, which has been measured and entered into the model, is correctly simulated.

One can also observe in Fig. 5 that the presence of oscillations in the inclusion is well simulated by the model. This is confirmed in Fig. 6 that represents, for both selected frequencies, the superposition of the experimental and simu-

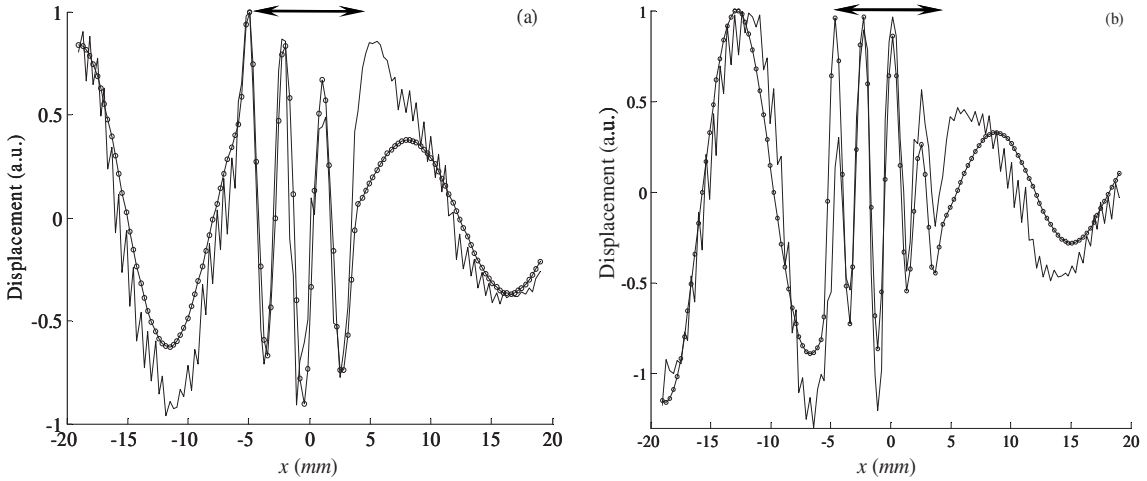


FIG. 6. Comparison of the experimental and simulated stationary displacements along the x axis in the plane Π for (a) a 350 Hz harmonic incident wave and (b) for a 450 Hz harmonic incident wave (— experimental; ⊖ simulation). The double arrow, in the top of figures, designates the inclusion spatial location.

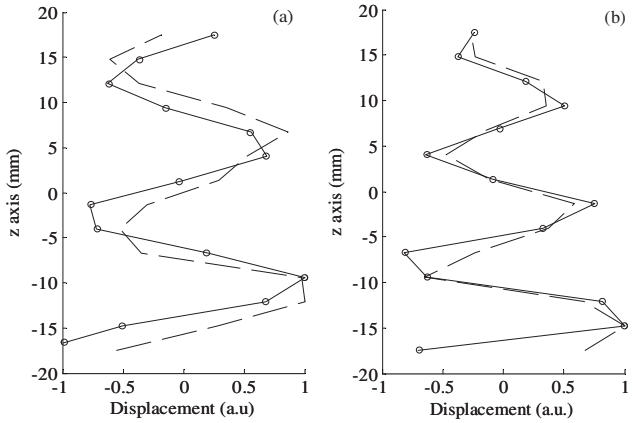


FIG. 7. Comparison of experimental and simulated stationary displacements along the z axis into the inclusion for (a) the 350 Hz harmonic incident wave and (b) for the 450 Hz harmonic incident wave (—— experimental; ⊖ simulation).

lated displacement profiles along a line ($y=0$) parallel to the x axis, crossing the inclusion and contained in the plane Π. A good correspondence between calculations and measurements is observable, except for the region immediately after the inclusion in the positive x direction. One can note the effect of attenuation increasing with frequency between the two plots, particularly in the region ranging between 5.0 and 20.0 mm. One may also notice differences in amplitude between measured and calculated displacements in Fig. 6 that are also observable in the stationary fields of Fig. 5. This is due to three principal reasons. The first one is the uncertainty in mechanical properties of materials composing inclusion and surrounding medium given in Sec. III B. Since theoretical displacements are calculated for mean values of the measured viscoelastic properties (which are global measurements), locally, slight differences between simulated and measured displacements (particularly for wave attenuation) can exist. The second source of differences is attributed to the relatively low spatial resolution in the lateral x direction because of the limited number of elements in the array transducer. The measured displacements were consequently subsampled (particularly into the inclusion where wavelengths are smallest) and an error on the inclusion localization occurred. Finally, the whole phantom dimensions were finite and the surrounding medium viscosity was relatively low. Experimentally, in the harmonic regime, this caused SW reflections at boundaries that perturbed the measured displacements into the inclusion and in its neighboring region.

The abovementioned sources of errors were simultaneously present but they did not dramatically affect the quality of measurements. This is also verified in Fig. 7, where experimental and theoretical stationary displacements along the z axis are compared for each tested frequency. As one can expect, the wavelength following the z direction, as in the other directions, is inversely proportional to the incident wave frequency (which is 14.5 and 11.3 mm at 350 and 450 Hz, respectively). The best agreement is thus noticed at the highest frequency of 450 Hz in Fig. 7. Existing differences could be explained by the spatial discretization error ($\Delta z = 2.5$ mm) and by elastic wave reflections at phantom boundaries that are strongest at low frequencies (since SWs are less attenuated at low frequencies).

Before concluding for the monochromatic case, it is important to point out that the presence of oscillations in the inclusion, observed in the horizontal planes (see, for example, Fig. 6) and along the vertical direction (as in Fig. 7), is directly related to the inclusion constitutive material viscoelasticity. Indeed, on one hand oscillation wavelengths are a function of elasticity, and on the other hand their amplitudes depend on both elasticity and viscosity. Another scattered wave characteristic is its spatial distribution in a given horizontal plane, i.e., the SW slowing down after crossing the inclusion, the orientation of diffraction lobes in the surrounding medium (see Fig. 5), the 2D shape of oscillations into the inclusion, the wave diffraction angle into the inclusion with respect to its axis, etc. All these behaviors are related to the internal and external viscoelastic properties and also to the contrast between them. Consequently, as it is developed in Sec. IV B, one could exploit this rich information to characterize mechanical properties of heterogeneous media.

B. Transient case

The harmonic 3D model is now used to simulate scattering of a transient SW with the previously described strategy. Both simulated and experimental incident waves were constituted by six oscillations, modulated by a Blackman temporal window to minimize Gibbs effects, with 350 Hz central frequency (see Fig. 2). Theoretical displacements are compared as a function of time with measurements obtained on the same phantom as the one used for the monochromatic excitation.

Figure 8 shows the transient wave propagation at different moments in half of the total 3D probed volume. This 3D view has been chosen to visualize more precisely the internal interaction of the incident SW with the cylindrical inclusion and the wave diffraction angle within the cylinder relative to the z axis. Time-varying experimental and simulated displacements are in good agreement in both inclusion and surrounding media. However, some comments can be made on the scattering behavior. First, it is noticeable that the experimental incident wave is not perfectly plane (this is observed at $t=21.8$ ms in the lower left part of the image). This explains slight differences between measured and predicted displacement fields. A second remark concerns oscillations observable in simulations in both inclusion and surrounding media before the incident wave arrival (see the right panel of Fig. 8 at $t=11.4$ ms). These artifacts are due to the Gibbs phenomenon, which appears when the scattering spectral solution is inversely Fourier transformed to the time domain. Indeed, since the scattering problem is solved for a finite frequency range, we applied the zero-padding technique outside this range to get the total spectral response. This numerical technique introduced a small discontinuity in the spectra and, in consequence, small nonphysical oscillations in the time response.

Despite these sources of errors, in addition to those referenced for the harmonic case (except reflections which are

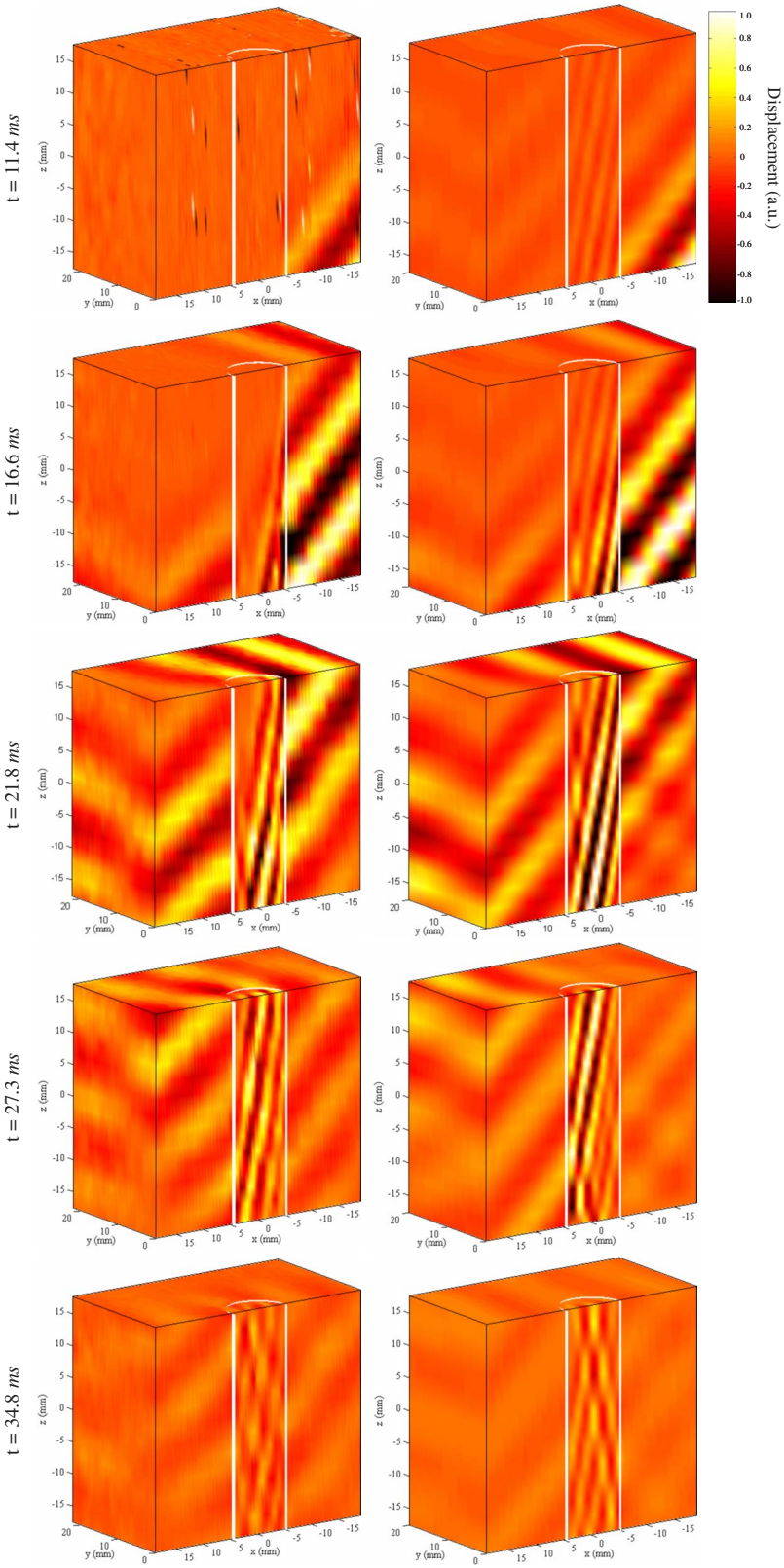


FIG. 8. (Color online) Time-varying 3D representation of the transient wave excitation scattered by the soft cylindrical inclusion. The view represents the half of the volume to visualize the behavior of the wave into the inclusion during the propagation. Inclination and incidence angles, β and α, were equal to $-3°$ and $46°$, respectively.

absent in the transient case), we can observe that the 3D transient SW scattering was well simulated by the model. It is important to notice that, as for the harmonic scattering problem, the wave distribution and amplitude are related to the medium viscoelasticity along the propagation path. Typically, the wavelength and the wave diffraction angle into the cylinder are related to the inclusion mechanical properties and to the contrast in viscoelasticity between both media, respectively.

V. CONCLUSION AND PERSPECTIVES

An analytical model was developed to simulate the scattering of an arbitrary (monochromatic or transient) plane SW by a circular cylindrical heterogeneity. Both inclusion and surrounding media constitutive materials were soft enough to allow the slow propagation of SWs which could, consequently, be imaged by an ultrafast ultrasonic system. The model validation was achieved experimentally on a heterogeneous phantom containing a very soft cylindrical inclusion. The comparison of measured and calculated 3D, 2D, and 1D stationary displacement fields has shown the model validity to simulate SW scattering by a cylindrical heterogeneity. The validated model then served to simulate the scattering of a transient plane SW (using the superposition principle). In this configuration, 3D theoretical temporal displacements were also in good agreement with measurements. When the geometrical and mechanical configurations of a scattering problem allow an analytical solving, analytical approaches are generally more precise, more stable, and faster than numerical methods (such as finite-element and finite-difference methods, well adapted to irregular scattering configurations). Moreover, analytical models allow the calculation of mechanical fields, such as displacement, strain, and stress, at any spatial position (a part of the plane, a line, or a set of points), without calculating in the whole volume like in numerical approaches. These advantages could be interesting to solve inverse problems using simulations.

In the context of dynamic elastography of soft structures, and more generally of living tissues, the approach developed in the present work could likely be applied to assess viscoelastic properties of venous blood clots, as introduced earlier. Indeed, these latter have a cylindrical shape (almost circular) and are constituted by coagulated blood with elasticity that does not exceed few kilopascals and a viscosity that is generally high.[5] Clot mechanical properties depend on several factors such as hematocrit level, the time elapsed since its formation in a blood vessel, chemical and metabolic conditions, etc. One can then consider to formulate and solve an inverse problem, into a strategy using both *in vivo* and simulated data, to perform mechanical property assessments of clots. The correlation of the assessed viscoelasticity with the clot age, its compactness, and biological constitution would provide to clinicians a complementary tool for venous thrombosis diagnosis and therapy planning. Moreover, the general 3D configuration considered here is compatible with real clinical conditions where mechanical excitation and ultrasound imaging angles can be arbitrary due, for example, to anatomical restrictions. The formulation and resolution of such inverse problem should deserve attention. Parallel to this direct application of the present work, the model could be developed to simulate scattering of SWs by a multilayer cylindrical inclusion. This extension would permit to treat the more general and realistic case of an inhomogeneous inclusion. To be more realistic, another avenue of development would be the simulation of SW interaction with geometrically and mechanically more complex heterogeneities (e.g., hollowed heterogeneities, elliptical cylinders, finite heterogeneities, anisotropic materials, etc.).

To conclude, this study showed that it is possible to analytically model SW interactions with soft bodies. The impact of this model is not limited to ultrasound elastography (in both harmonic and transient regimes) but also to MRI based elastography (monochromatic regime) that can be formulated in a similar framework. It is finally conceivable to extend this modeling strategy to other pathologies such as breast, brain, prostate and liver tumors, atherosclerosis, etc.

ACKNOWLEDGMENTS

This work was supported by a grant from the Canadian Institutes of Health Research (No. MOP-84358), by a National Scientist award of the Fonds de la Recherche en Santé du Québec (G.C.), and by postdoctoral (A.H.H.) and doctoral (C.S.) partial scholarships of the Groupe de Recherche en Sciences et Technologies Biomédicales of the Institute of Biomedical Engineering of the École Polytechnique and Université de Montréal.

Appendix

The (6×6) $\mathbf{T}_n$ matrix elements in Eq. (8) are given by

$$\mathbf{T}_n(1,1)=\frac{1}{2}K_1[J_{n-1}(K_1R)-J_{n+1}(K_1R)],\quad \mathbf{T}_n(1,2)=\frac{R}{2}k_1ik_{2_z}[J_{n-1}(k_1R)-J_{n+1}(k_1R)],$$

$$\mathbf{T}_n(1,3)=-\frac{n}{R}J_n(k_1R),\quad \mathbf{T}_n(1,4)=-\frac{1}{2}K_{2_\perp}[H_{n-1}^{(1)}(K_{2_\perp}R)-H_{n+1}^{(1)}(K_{2_\perp}R)],$$

$$\mathbf{T}_n(1,5)=-\frac{R}{2}k_{2_\perp}ik_{2_z}[H_{n-1}^{(1)}(k_{2_\perp}R)-H_{n+1}^{(1)}(k_{2_\perp}R)],\quad \mathbf{T}_n(1,6)=\frac{n}{R}H_n^{(1)}(k_{2_\perp}R),$$

$$\mathbf{T}_n(2,1)=\frac{n}{R}J_n(K_1R),\quad \mathbf{T}_n(2,2)=nik_{2_z}J_n(k_1R),\quad \mathbf{T}_n(2,3)=-\frac{1}{2}k_1[J_{n-1}(k_1R)-J_{n+1}(k_1R)],$$

$$\mathbf{T}_n(2,4)=-\frac{n}{R}H_n^{(1)}(K_{2_\perp}R),\quad \mathbf{T}_n(2,5)=-nik_{2_z}\times H_n^{(1)}(k_{2_\perp}R),$$

$$\mathbf{T}_n(2,6)=\frac{1}{2}k_{2_\perp}[H_{n-1}^{(1)}(k_{2_\perp}R)-H_{n+1}^{(1)}(k_{2_\perp}R)],$$

$$\mathbf{T}_n(3,1)=ik_{2_z}J_n(K_1R),$$

$$\mathbf{T}_n(3,2) = R\Bigg[-\frac{(k_1)^2}{4}(J_{n-2}(k_1R) - 2J_n(k_1R) + J_{n+2}(k_1R)) - \frac{k_1}{2R}(J_{n-1}(k_1R) - J_{n+1}(k_1R)) + \frac{n^2}{R^2}J_n(k_1R)\Bigg],$$

$$\mathbf{T}_n(3,3) = 0, \quad \mathbf{T}_n(3,4) = -ik_{2_z}H_n^{(1)}(K_{2_\perp}R),$$

$$\mathbf{T}_n(3,5) = R\Bigg[+\frac{(k_{2_\perp})^2}{4}(H_{n-2}^{(1)}(k_{2_\perp}R) - 2H_n^{(1)}(k_{2_\perp}R) + H_{n+2}^{(1)}(k_{2_\perp}R)) + \frac{k_{2_\perp}}{2R}(H_{n-1}^{(1)}(k_{2_\perp}R) - H_{n+1}^{(1)}(k_{2_\perp}R)) - \frac{n^2}{R^2}H_n^{(1)}(k_{2_\perp}R)\Bigg],$$

$$\mathbf{T}_n(3,6) = 0,$$

$$\mathbf{T}_n(4,1) = (\lambda_1 + 2\mu_1)\frac{(K_1)^2}{4}[J_{n-2}(K_1R) - 2J_n(K_1R) + J_{n+2}(K_1R)] - \lambda_1\left(\frac{n^2}{R^2} + (k_{2_z})^2\right)J_n(K_1R) + \lambda_1\frac{K_1}{2R}[J_{n-1}(K_1R) - J_{n+1}(K_1R)],$$

$$\mathbf{T}_n(4,2) = 2\mu_1\frac{R}{4}i(k_1)^2k_{2_z}[J_{n-2}(k_1R) - 2J_n(k_1R) + J_{n+2}(k_1R)] + 2\lambda_1\frac{1}{R}ik_{2_z}n^2J_n(k_1R),$$

$$\mathbf{T}_n(4,3) = +2\mu_1\frac{n}{R^2}J_n(k_1R) - \mu_1\frac{n}{R}k_1[J_{n-1}(k_1R) - J_{n+1}(k_1R)],$$

$$\mathbf{T}_n(4,4) = -(\lambda_2 + 2\mu_2)\frac{(K_{2_\perp})^2}{4}[H_{n-2}^{(1)}(K_{2_\perp}R) - 2H_n^{(1)}(K_{2_\perp}R) + H_{n+2}^{(1)}(K_{2_\perp}R)] + \lambda_2\left(\frac{n^2}{R^2} + (k_{2_z})^2\right)H_n^{(1)}(K_{2_\perp}R) - \lambda_2\frac{K_{2_\perp}}{2R}[H_{n-1}^{(1)}(K_{2_\perp}R) - H_{n+1}^{(1)}(K_{2_\perp}R)],$$

$$\mathbf{T}_n(4,5) = -2\mu_2\frac{R}{4}i(k_{2_\perp})^2k_{2_z}[H_{n-2}^{(1)}(k_{2_\perp}R) - 2H_n^{(1)}(k_{2_\perp}R) + H_{n+2}^{(1)}(k_{2_\perp}R)] - 2\lambda_2\frac{1}{R}ik_{2_z}n^2H_n^{(1)}(k_{2_\perp}R),$$

$$\mathbf{T}_n(4,6) = -2\mu_2\frac{n}{R^2}H_n^{(1)}(k_{2_\perp}R) + \mu_2\frac{n}{R}k_{2_\perp}[H_{n-1}^{(1)}(k_{2_\perp}R) - H_{n+1}^{(1)}(k_{2_\perp}R)],$$

$$\mathbf{T}_n(5,1) = -2\mu_1\frac{n}{R^2}J_n(K_1R) + \mu_1\frac{n}{R}K_1[J_{n-1}(K_1R) - J_{n+1}(K_1R)],$$

$$\mathbf{T}_n(5,2) = -2\mu_1\frac{n}{R}ik_{2_z}J_n(k_1R) + \mu_1ink_{2_z}k_1[J_{n-1}(k_1R) - J_{n+1}(k_1R)],$$

$$\mathbf{T}_n(5,3) = -\mu_1\frac{(k_1)^2}{4}[J_{n-2}(k_1R) - 2J_n(k_1R) + J_{n+2}(k_1R)] - \mu_1\frac{n^2}{R^2}J_n(k_1R) + \mu_1\frac{k_1}{2R}[J_{n-1}(k_1R) - J_{n+1}(k_1R)],$$

$$\mathbf{T}_n(5,4) = +2\mu_2\frac{n}{R^2}H_n^{(1)}(K_{2_\perp}R) - \mu_2\frac{n}{R}K_{2_\perp}[H_{n-1}^{(1)}(K_{2_\perp}R) - H_{n+1}^{(1)}(K_{2_\perp}R)],$$

$$\mathbf{T}_n(5,5) = +2\mu_2\frac{n}{R}ik_{2_z}H_n^{(1)}(k_{2_\perp}R) - \mu_2ink_{2_z}k_{2_\perp} \times[H_{n-1}^{(1)}(k_{2_\perp}R) - H_{n+1}^{(1)}(k_{2_\perp}R)],$$

$$\mathbf{T}_n(5,6) = +\mu_2\frac{(k_{2_\perp})^2}{4}[H_{n-2}^{(1)}(k_{2_\perp}R) - 2H_n^{(1)}(k_{2_\perp}R) + H_{n+2}^{(1)}(k_{2_\perp}R)] + \mu_2\frac{n^2}{R^2}H_n^{(1)}(k_{2_\perp}R) - \mu_2\frac{k_{2_\perp}}{2R} \times[H_{n-1}^{(1)}(k_{2_\perp}R) - H_{n+1}^{(1)}(k_{2_\perp}R)],$$

$$\mathbf{T}_n(6,1) = +2\mu_1ik_{2_z}K_1[J_{n-1}(K_1R) - J_{n+1}(K_1R)],$$

$$\mathbf{T}_n(6,2) = -\mu_1Rk_1(k_{2_z})^2[J_{n-1}(k_1R) - J_{n+1}(k_1R)] - \mu_1R\frac{(k_1)^2}{4}[J_{n-3}(k_1R) - 3J_{n-1}(k_1R) + 3J_{n+1}(k_1R) - J_{n+3}(k_1R)] + \mu_1\frac{k_1}{R}[J_{n-1}(k_1R) - J_{n+1}(k_1R)] - \mu_1\frac{(k_1)^2}{2}[J_{n-2}(k_1R) - 2J_n(k_1R) + J_{n+2}(k_1R)] - 4\mu_1\frac{n^2}{R^2}J_n(k_1R) + \mu_1\frac{n^2}{R}k_1[J_{n-1}(k_1R) - J_{n+1}(k_1R)],$$

$$\mathbf{T}_n(6,3) = -2\mu_1\frac{n}{R}ik_{2_z}J_n(k_1R),$$

$$\mathbf{T}_n(6,4) = -2\mu_2ik_{2_z}K_{2_\perp}[H_{n-1}^{(1)}(K_{2_\perp}R) - H_{n+1}^{(1)}(K_{2_\perp}R)],$$

$$\mathbf{T}_n(6,5) = +\mu_2 R k_{2_\perp}(k_{2_z})^2[H_{n-1}^{(1)}(k_{2_\perp}R) - H_{n+1}^{(1)}(k_{2_\perp}R)]$$
$$+\mu_2 R\frac{(k_{2_\perp})^2}{4}[H_{n-3}^{(1)}(k_{2_\perp}R) - 3H_{n-1}^{(1)}(k_{2_\perp}R)$$
$$+3H_{n+1}^{(1)}(k_{2_\perp}R) - H_{n+3}^{(1)}(k_{2_\perp}R)] - \mu_2\frac{k_{2_\perp}}{R}$$
$$\times[H_{n-1}^{(1)}(k_{2_\perp}R) - H_{n+1}^{(1)}(k_{2_\perp}R)] + \mu_2\frac{(k_{2_\perp})^2}{2}$$
$$\times[H_{n-2}^{(1)}(k_{2_\perp}R) - 2H_n^{(1)}(k_{2_\perp}R) + H_{n+2}^{(1)}(k_{2_\perp}R)]$$
$$+4\mu_2\frac{n^2}{R^2}H_n^{(1)}(k_{2_\perp}R) - \mu_2\frac{n^2}{R}k_{2_\perp}[H_{n-1}^{(1)}(k_2R)$$
$$-H_{n+1}^{(1)}(k_2R)],$$

$$\mathbf{T}_n(6,6) = +2\mu_2\frac{n}{R}ik_{2_z}H_n^{(1)}(k_{2_\perp}R).$$

The second member vector in Eq. (8) contains the nth components of displacement and stress due to the incident wave at the cylinder boundary. The $\mathbf{b}_n$ vector elements are as follows:

$$(b_1)_n = -\frac{\mathrm{amp}(\omega)}{\psi_0}\frac{n}{R}\varepsilon_n(i)^n J_n(k_{2_\perp}R),$$

$$(b_2)_n = -\frac{\mathrm{amp}(\omega)}{\psi_0}\frac{1}{2}k_{2_\perp}\varepsilon_n(i)^n[J_{n-1}(k_{2_\perp}R) - J_{n+1}(k_{2_\perp}R)],$$

$$(b_3)_n = 0,$$

$$(b_4)_n = \frac{\mathrm{amp}(\omega)}{\psi_0}2\mu_2\frac{n}{R^2}\varepsilon_n(i)^n J_n(k_{2_\perp}R)$$
$$-\frac{\mathrm{amp}(\omega)}{\psi_0}\mu_2\frac{nk_{2_\perp}}{R}\varepsilon_n(i)^n[J_{n-1}(k_{2_\perp}r) - J_{n+1}(k_{2_\perp}r)],$$

$$(b_5)_n = -\frac{\mathrm{amp}(\omega)}{\psi_0}\mu_2\frac{(k_{2_\perp})^2}{4}\varepsilon_n(i)^n[J_{n-2}(k_{2_\perp}R) - 2J_n(k_{2_\perp}R)$$
$$+J_{n+2}(k_{2_\perp}R)] + \frac{\mathrm{amp}(\omega)}{\psi_0}\mu_2\frac{k_{2_\perp}}{2R}\varepsilon_n(i)^n[J_{n-1}(k_{2_\perp}R)$$
$$-J_{n+1}(k_{2_\perp}R)] - \frac{\mathrm{Amp}(\omega)}{\psi_0}\mu_2\frac{n^2}{R^2}\varepsilon_n(i)^n J_n(k_{2_\perp}R),$$

$$(b_6)_n = -\frac{\mathrm{amp}(\omega)}{\psi_0}2\mu_2\frac{n}{R}\varepsilon_n(i)^{n+1}k_{2_z}J_n(k_{2_\perp}R).$$

[1]World Health Organization, http://www.who.int (Last viewed December 12, 2007).
[2]J. Hirsh and J. Hoak, "Management of deep vein thrombosis and pulmonary embolism: A statement for healthcare professionals from the Council on thrombosis (in consultation with the council on cardiovascular radiology)," Circulation **93**, 2212–2245 (1996).
[3]M. Kaibara, "Rheology of blood coagulation," Biorheology **33**, 101–117 (1996).
[4]J. V. Shah and P. A. Janmey, "Strain hardening of fibrin gels and plasma clots," Rheol. Acta **36**, 262–268 (1997).
[5]J.-L. Gennisson, S. Lerouge, and G. Cloutier, "Assessment by transient elastography of the viscoelastic properties of blood during clotting," Ultrasound Med. Biol. **32**, 1529–1537 (2006).
[6]S. Y. Emelianov, X. Chen, M. O'Donnell, B. Knipp, D. Myers, T. W. Wakefield, and J. M. Rubin, "Triplex ultrasound: Elasticity imaging to age deep venous thrombosis," Ultrasound Med. Biol. **28**, 757–767 (2002).
[7]J. M. Rubin, S. R. Aglyamov, T. W. Wakefield, M. O'Donnell, and S. Y. Emelianov, "Clinical application of sonographic elasticity imaging for aging of deep venous thrombosis: Preliminary findings," J. Ultrasound Med. **22**, 443–448 (2003).
[8]R. M. Lerner, K. J. Parker, J. Holen, R. Gramiak, and R. C. Waag, "Sonoelasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets," Acoust. Imaging **16**, 317–327 (1988).
[9]Y. Yamakoshi, J. Sato, and T. Sato, "Ultrasonic imaging of internal vibration of soft tissue under forced vibration," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 45–53 (1990).
[10]A. Eisencher, E. Schweg-Toffler, G. Pelletier, and P. Jacquemard, "La palpation échographique rythmée: Echosismographie," J. Radiol. **64**, 255–261 (1983).
[11]T. A. Krouskop, D. R. Dougherty, and F. S. Vinson, "A pulsed Doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue," J. Rehabil. Res. Dev. **24**, 1–8 (1987).
[12]K. J. Parker and R. M. Lerner, "Sonoelasticity of organs: Shear waves ring a bell," J. Ultrasound Med. **11**, 387–392 (1992).
[13]R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," Science **269**, 1854–1857 (1995).
[14]S. Catheline, F. Wu, and M. Fink, "A solution to diffraction biases in sonoelasticity: The acoustic impulse technique," J. Acoust. Soc. Am. **105**, 2941–2950 (1999).
[15]L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear modulus imaging with 2D transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 426–435 (2002).
[16]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).
[17]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).
[18]J. J. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418 (1951).
[19]R. M. White, "Elastic wave scattering at a cylindrical discontinuity in a solid," J. Acoust. Soc. Am. **30**, 771–785 (1958).
[20]Y. Fan, A. N. Sinclair, and F. Honarvar, "Scattering of a plane acoustic wave from a transversely isotropic cylinder encased in a solid elastic medium," J. Acoust. Soc. Am. **106**, 1229–1236 (1999).
[21]F. Honarvar and A. N. Sinclair, "Acoustic wave scattering from transversely isotropic cylinders," J. Acoust. Soc. Am. **100**, 57–63 (1996).
[22]S. Catheline, J.-L. Gennisson, G. Delon, M. Fink, R. Sinkus, S. Abouelkaram, and J. Culioli, "Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach," J. Acoust. Soc. Am. **116**, 3734–3741 (2004).
[23]C. Schmitt, A. Hadj Henni, and G. Cloutier, "Characterization of time-varying mechanical viscoelastic parameters of mimicking deep vein thrombi with 2D dynamic elastography," in IEEE International Ultrasonics Symposium, New York (2007), pp. 28–31.
[24]J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).
[25]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[26]Matlabs® User's Guide, Mathworks, Inc., Natick, MA, http://www.mathworks.com
[27]E. Chérin, R. Williams, A. Needles, G. Liu, C. White, A. S. Brown, Y.-Q. Zhou, and F. S. Foster, "Ultrafast frame rate retrospective ultrasound microimaging and blood flow visualization in mice *in vivo*," Ultrasound Med. Biol. **32**, 683–691 (2006).
[28]M. Pernot, K. Fujikura, S. D. Fung-Kee-Fung, and E. Konofagou, "ECG-gated, mechanical and electromechanical wave imaging of cardiovascular tissues *in vivo*," Ultrasound Med. Biol. **33**, 1075–1085 (2007).
[29]J.-L. Gennisson and G. Cloutier, "Sol-Gel transition in agar-gelatin mixtures studied with transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 716–723 (2006).

Acoustic characterization of high intensity focused ultrasound fields: A combined measurement and modeling approach

Michael S. Canney, Michael R. Bailey, and Lawrence A. Crum
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Vera A. Khokhlova and Oleg A. Sapozhnikov
Department of Acoustics, Faculty of Physics, Moscow State University, Moscow 119992, Russia and Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

(Received 30 December 2007; revised 24 June 2008; accepted 27 June 2008)

Acoustic characterization of high intensity focused ultrasound (HIFU) fields is important both for the accurate prediction of ultrasound induced bioeffects in tissues and for the development of regulatory standards for clinical HIFU devices. In this paper, a method to determine HIFU field parameters at and around the focus is proposed. Nonlinear pressure waveforms were measured and modeled in water and in a tissue-mimicking gel phantom for a 2 MHz transducer with an aperture and focal length of 4.4 cm. Measurements were performed with a fiber optic probe hydrophone at intensity levels up to 24 000 W/cm^2. The inputs to a Khokhlov–Zabolotskaya–Kuznetsov-type numerical model were determined based on experimental low amplitude beam plots. Strongly asymmetric waveforms with peak positive pressures up to 80 MPa and peak negative pressures up to 15 MPa were obtained both numerically and experimentally. Numerical simulations and experimental measurements agreed well; however, when steep shocks were present in the waveform at focal intensity levels higher than 6000 W/cm^2, lower values of the peak positive pressure were observed in the measured waveforms. This underrepresentation was attributed mainly to the limited hydrophone bandwidth of 100 MHz. It is shown that a combination of measurements and modeling is necessary to enable accurate characterization of HIFU fields.
 [DOI: 10.1121/1.2967836]

PACS number(s): 43.80.Vj, 43.25.Zx, 43.80.Ev, 43.25.Cb [ROC] Pages: 2406–2420

I. INTRODUCTION

High intensity focused ultrasound (HIFU) is an evolving medical technology for noninvasive surgery and cancer therapy. HIFU devices are currently under investigation for use as surgical tools, for example, to thermally ablate solid tumors of the prostate,[1] liver,[2] breast,[3] kidney,[4] and brain,[5] as well as for cauterizing internal bleeding.[6]

In HIFU medical treatments, ultrasound (US) energy is focused into a small volume to heat and destroy the targeted tissue while ideally not damaging tissue outside the focal region. The acoustic characterization of HIFU fields is important both for the accurate prediction of US induced bioeffects in tissues and for the development of standards to ensure the safety and efficacy of treatments. Modern HIFU devices operate at very high focal intensity levels from 1000 W/cm^2 to greater than 25 000 W/cm^2 *in situ* and are highly focused to a millimeter or even a submillimeter sized focal spot.[2,7] The high pressures and tight focusing of HIFU devices make accurate acoustic field measurements challenging.

For medical work, the characterization of acoustic output is often performed in two parts. First, measurements are performed in water. Next, a method is used to relate the measurements in water to what is expected in tissue. Although this approach is widely used in medical US, the challenge of directly measuring and modeling HIFU fields is worth revisiting since experimental tools and numerical models may improve over time, and no standard method, such as that for diagnostic US,[8] has yet been agreed upon for HIFU devices.[9]

The most common parameter to characterize the acoustic output of a HIFU device is the spatially averaged intensity (I_{SAL}).[10] The calculation of I_{SAL} for a given source is achieved through measurement of the acoustic power output of the device and the focal cross-sectional area of the HIFU beam. The acoustic power output is measured using a radiation force balance. The focal cross-sectional area is determined by scanning a hydrophone across the focal plane of the source. The spatially averaged intensity is then calculated by dividing the acoustic power that passes within the −6 dB pressure contour[11] by the area of the contour. Acoustic propagation is assumed to be linear in that the frequency content of the pressure waveform and the corresponding beam width are assumed to be independent of the source power. The peak positive and peak negative pressures for sinusoidal waveforms are therefore equal and only one needs to be measured. The low-power measurements of I_{SAL} are then linearly extrapolated to higher output levels of the HIFU device. That is, the focal pressure amplitude is assumed to scale linearly with applied source voltage so that the focal intensity (I_{SAL}) increases as a square of the voltage.

Once the output intensity is determined in water, it can be used to estimate the focal intensity in tissue. This step is achieved by multiplying by a compensation factor depending only on the propagation path and linear attenuation coefficient of tissue at the operating frequency of the source. This approach is useful for comparing exposures at relatively low intensities as it provides an estimate of the *in situ* focal intensity. Tissue heating, the common bioeffect of interest, can also be estimated from the product of the tissue absorption coefficient and I_{SAL}. However, the acoustic fields of most HIFU clinical devices are nonlinear; therefore, errors are introduced at every characterization step where linear acoustic propagation is assumed.[12,13]

At currently reported HIFU intensity levels, the combined effects of nonlinear propagation and diffraction in water or in tissue lead to generation of higher harmonics, narrowing of harmonic beam widths, asymmetric distortion of pressure waveforms, and formation of steep shock fronts.[14,15] Nonlinear broadening of the spectrum to higher frequencies and the formation of shocks can significantly increase the amount of energy converted to heat, as absorption in tissue increases with frequency.[16,17] Since nonlinear effects accumulate with propagation distance and depend on the initial pressure amplitude, waveforms have different shape and spectral content at different locations of a HIFU beam and at different source power levels. Acoustic characterization of a nonlinear HIFU field thus requires measurements not only of the pressure amplitude or intensity but of the entire acoustic waveform at each source power and location. At a minimum, the focal waveform, which contains the broadest spectral content, is necessary to obtain the spatial peak heating rate under nonlinear propagation conditions. Because nonlinear effects accumulate differently in water and in tissue, the focal waveform measured in water cannot necessarily be directly translated to tissue.

The characterization of medical acoustic devices that operate at high output levels has been a research topic and an issue of practical concern for several decades.[18] The importance of nonlinear effects has been considered and addressed even at diagnostic levels of US.[13,19] In lithotripsy and HIFU, these effects are critical as acoustic pressures of up to 100 MPa or higher can be reached; such pressures are two or even three orders higher in magnitude than diagnostic US. Direct measurements of pressure waveforms in HIFU or lithotripsy fields require a hydrophone that is robust to mechanical damage from cavitation and that has a large bandwidth to capture sharp shock fronts than can develop at the focus. Waveform measurements in HIFU are even more difficult than in lithotripsy because of the explicit need to accurately measure peak amplitudes of shock waves within a relatively small focal region. Accurate measurement of peak pressures is especially relevant in HIFU because the prediction of heating rates is sensitive to shock amplitudes. When shocks are present, the amount of heat deposition is determined by the cube of the shock amplitude.[17] This imposes additional requirements on hydrophone bandwidth. In addition, focal beam dimensions of clinical HIFU devices are usually much smaller than in lithotripsy. The typical size of the HIFU focal zone is about an order of magnitude smaller in both the transverse and axial directions. As such, a hydrophone with a very small active diameter is required to accurately measure HIFU fields.

As mentioned, direct measurements of high amplitude fields can be performed in water. Methods have been described to measure nonlinear fields of diagnostic US transducers[13,19] and lithotripters.[18] Recently, shocked waveforms from a HIFU source were measured using a fiber optic probe hydrophone (FOPH).[20] Such hydrophones claim advantages of self-calibration, small size, and broad bandwidth relative to polyvinylidene fluoride (PVDF) hydrophones.[20,21]

Alternatively, numerical modeling has been used to predict high amplitude acoustic fields from medical devices. One advantage of modeling is that it can be used to determine the acoustic field in both water and tissue. Numerical algorithms, most commonly based on the nonlinear parabolic Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation, have been developed and applied to the nonlinear fields of lithotripters,[22] unfocused ultrasonic piston sources,[23,24] diagnostic US transducers operating in tissue harmonic imaging mode,[25,26] focused US sources,[27] and HIFU sources.[17,28,29] For highly focused sources, the Rayleigh integral can be used to capture diffraction effects more accurately in the model.[30,31] More comprehensive models based on full-wave nonlinear equations have also been developed, but they are much more computationally intensive.[32]

In this study, a new method to characterize HIFU field parameters at and around the focus is proposed and examined in water and in a tissue-mimicking gel phantom. First, at low source excitation, the pressure amplitude along and across the HIFU axis beyond the −6 dB points is measured using a hydrophone with a spot size smaller than the HIFU wavelength. Second, the voltage amplitude applied to the source and focal pressure is measured. Third, the source is modeled as a curved uniform piston using nominal measures of curvature and aperture and the axial and transverse pressure plots are calculated in the linear regime and compared with the measurements. The exact source geometry to be used in the model is determined by adjusting the curvature and aperture to best fit the dimensions of the measured focal region. Fourth, the source pressure amplitude, p_0, is determined by matching the measured and calculated focal pressures. In this last step, the source voltage measured at any excitation can be converted to a source pressure, p_0, used as an input to the model. With these steps completed, the model can then be used to calculate the nonlinear fields in water or in tissue. Here, the simulations are validated against experimental measurements in water and in a tissue-mimicking phantom. Methods for further refining the techniques and for broadening them to other source geometries are also discussed.

The clinically relevant goal of this work is to demonstrate methods for accurately determining HIFU pressure waveforms. This paper is timely as new clinical devices enter the market and many of them operate at amplitudes much higher than used in previous research. The problems of measuring and modeling nonlinear shocked HIFU waveforms as well as potential reasons for discrepancies observed between measured and modeled results are examined and discussed.

While direct low amplitude measurements of a HIFU field in water are necessary to characterize the spatial pattern of the field and to establish precise boundary conditions for the model, simulations provide more accurate results than measurements at high source amplitude levels.

II. MATERIALS AND METHODS

A 2 MHz single element HIFU transducer with a diameter and focal length of 44 mm was used in measurements and calculations and is representative of many HIFU sources. Intensity levels *in situ* reached 24 000 $\mathrm{W/cm^2}$, a level used by some clinical devices.[7] Calibration measurements and modeling were performed in water and in a transparent tissue-mimicking gel phantom. The gel phantom consisted of polyacrylamide and a 7% concentration of bovine serum albumen (BSA) and has been used in several studies as a test phantom for HIFU dosimetry studies.[28,33–35] While the acoustic properties of the phantom material such as sound speed and density are similar to tissue, the absorption coefficient is about three times lower than in tissue. The lower absorption leads to more pronounced nonlinear effects than might be observed in tissue at similar source power levels.

A. Numerical model

Numerical modeling of experimental conditions was performed using a KZK-type nonlinear parabolic equation, generalized for the frequency-dependent absorption properties of the propagation medium:

$$\frac{\partial}{\partial\tau}\left[\frac{\partial p}{\partial z}-\frac{\beta}{\rho_0 c_0^3}p\frac{\partial p}{\partial\tau}-L_{\mathrm{abs}}(p)\right]=\frac{c_0}{2}\Delta_\perp p, \tag{1}$$

where p is the acoustic pressure, z is the propagation coordinate along the axis of the beam, $\tau=t-z/c_0$ is the retarded time, c_0 is the ambient sound speed, ρ_0 is the ambient density of the medium, β is the coefficient of nonlinearity, $\Delta_\perp$ is the Laplacian with respect to the transverse coordinate r, and L_{abs} is the linear operator that accounts for the absorption and dispersion of the medium.[17,28]

For simulations in water, thermoviscous absorption was included as

$$L_{\mathrm{abs}}=\frac{b}{2\rho_0 c_0^3}\frac{\partial^2 p}{\partial\tau^2}, \tag{2}$$

where b is the dissipative parameter of water. For simulations in gel, the propagation path for US comprised a two-layer medium consisting of water followed by tissue-mimicking gel phantom. The frequency-dependent absorption of US in the gel was included in the model according to a nearly linear power law combined with a weak thermoviscous absorption, as in water:

$$\alpha(f)=2\pi^2 f^2 b/\rho_0 c_0^3+\alpha_0(f/f_0)^\eta. \tag{3}$$

Here α_0 is the absorption parameter of the gel at the fundamental frequency, f_0, and variation of the sound speed with frequency was calculated for the power law term (η) in Eq. (3) using the local dispersion relations.[36,37]

The boundary condition for Eq. (1) was set by translating the pressure amplitude, p_0, uniformly distributed over the curved surface of the source to the plane $z=0$. The translation of the amplitude was performed using a geometrical acoustics approximation following the spherical convergence of the field. The focusing phase shift along the radial coordinate was introduced in the parabolic approximation as

$$p(z=0,r,\tau)=\frac{p_0}{\sqrt{1+r_0^2/F^2}}\sin(2\pi f_0(\tau+r^2/2c_0F)), \tag{4}$$

if $r<r_0/\sqrt{(1-r_0^2/F^2)}$ and $p(z=0,r,\tau)=0$ elsewhere. Here, $2r_0$ is the aperture of the source and F is its radius of curvature. To account for nonuniform vibration of the surface of the transducer,[38] additional simulations were performed with more accurate boundary conditions that were reconstructed using an experimental acoustic holography method.[39]

Equation (1) was solved numerically in the frequency domain using a previously developed finite difference algorithm.[17,24] The acoustic pressure waveform was represented as a Fourier series expansion as

$$p(z,r,\tau)=\sum_{n=1}^{\infty}c_n(z,r)e^{-in2\pi f_0\tau}, \tag{5}$$

where c_n is the complex amplitude of the n th harmonic. A set of nonlinear coupled differential equations for the amplitudes of the harmonics was derived and integrated numerically using the method of fractional steps and an operator-splitting procedure. The simulations were performed assuming radial symmetry of the HIFU source.

To characterize the HIFU output level in the focal zone in water or in gel, two values of the spatial peak intensity were introduced. These values will be referred to in the paper as focal intensities. The first value, I_N, was calculated from the numerically modeled nonlinear waveform as a combination of the focal intensities of all harmonic components:

$$I_N=f_0\int_0^{1/f_0}\frac{p^2}{\rho_0 c_0}d\tau=\frac{2}{\rho_0 c_0}\sum_{n=1}^{\infty}|c_n|^2. \tag{6}$$

The second value, I_L, was calculated based on the results of the linear acoustic propagation modeling as

$$I_L=p_F^2/2\rho_0 c_0, \tag{7}$$

where p_F is the focal (i.e., spatial peak) pressure amplitude *in situ*. The linear focusing gain of the source was defined as a ratio of the focal and source pressure amplitude obtained from linear modeling in water:

$$G=p_F/p_0. \tag{8}$$

The values of the physical constants used for the modeling in water were $\rho_0=1000\ \mathrm{kg/m^3}$, $c_0=1486\ \mathrm{m/s}$, $\beta=3.5$, and $b=4.33\times10^{-3}\ \mathrm{kg\ s^{-1}\ m^{-1}}$. In the gel phantom, the constants were $\rho_0=1044\ \mathrm{kg/m^3}$, $c_0=1544\ \mathrm{m/s}$, $\beta=4$, $\alpha_0=1.6\ \mathrm{m^{-1}}$ at 1 MHz, $\eta=1$, and $L=0.033$ mm.[33]

The simulations were performed for the following parameters of the numerical scheme: $\Delta z/z=2.5\times10^{-5}$ and $\Delta r/r_0=0.5\times10^{-3}$ were the steps in the axial and transverse directions, respectively, $z_{\max}=2F$ and $r_{\max}=1.25r_0$ were the

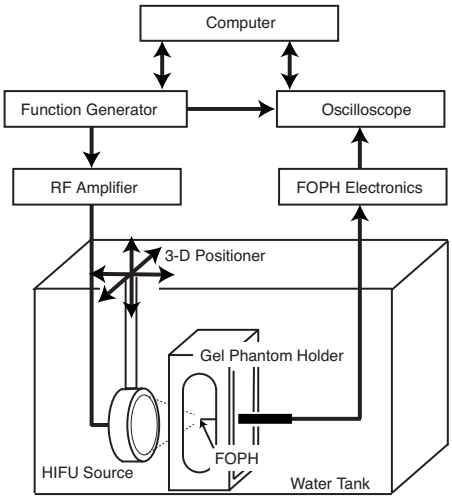


FIG. 1. A diagram of the experimental arrangement used for measurements of pressure waveforms.

spatial windows, and $N_{max}=1000$ was the maximum number of harmonics retained in the code. To suppress reflections from the radial boundary of the grid, an artificial absorption was introduced for the last 100 radial grid points. Within this layer, the absorption coefficient increased quadratically with the radial coordinate from zero to 100 Np/cm and was equal for all harmonics included in calculations.

B. Experimental arrangement

A diagram of the experimental arrangement used to perform the high-power measurements is shown in Fig. 1. The US source was a single element spherically focused piezoceramic crystal (PZ 26, Ferroperm Piezoceramics, Denmark) with a resonant frequency of 2.158 MHZ that was air backed and mounted in a custom-designed brass housing. The transducer was driven by a function generator (Agilent 33250A, Agilent, Palo Alto, CA) and a linear rf amplifier (ENI A-300, ENI, Rochester, NY). A high-voltage probe was attached in parallel with the transducer for monitoring the driving voltage amplitude to the source. The transducer was operated in a pulse mode with a 30-cycle sine-wave burst. Waveforms from the hydrophones and the driving voltage were recorded using a digital oscilloscope (500 megasamples/s, model LT344, Lecroy, Chestnut Ridge, NY). Waveform averaging was performed of multiple tone bursts to increase the signal-to-noise ratio. The US source was attached to a three-axis computer-controlled positioning system (Velmex Inc., Bloomfield, NY) for alignment with the hydrophone. A custom LABVIEW (National Instruments, Austin, TX) computer program was used to control the function generator, oscilloscope, and the positioning system during measurements. All of the experiments were performed in a large water tank at room temperature (20 °C). The water was purified using a reverse osmosis system and was degassed prior to measurements using a pinhole degassing system to limit cavitation.[40] The dissolved oxygen content of the water was measured using an oxygen meter (WTW Oxi 330i, Weilheim, Germany) and was less than 25% of saturation during all experiments.

C. Calibration measurements in water

Experiments performed to establish the boundary conditions for the numerical model involved measurements in water with two different hydrophones. First, a SEA needle hydrophone [uncalibrated, model GL-150-1A with 150 μm active diameter, Specialty Engineering Associates (SEA), Soquel, CA] was used for measuring axial and transverse profiles of the source under linear propagation conditions. Next, a calibrated NTR PVDF membrane hydrophone (0.168 V/MPa sensitivity, model MHA-200 with 200 μm active diameter, NTR Systems, Seattle, WA) was used as a reference hydrophone to measure focal waveforms to establish the source pressure amplitude, p_0, for a given source driving voltage. The FOPHs were not used for these measurements because their sensitivity is too low. Focal waveforms were then measured at moderate source powers using both the NTR and the FOPH to ensure agreement between the calibrated hydrophones. Two FOPHs (FOPH 500 and FOPH 2000 with 100 μm active diameters, RP Acoustics, Germany) were used for pressure measurements at the higher levels of operation of the HIFU source. The FOPH 500 and 2000 are similar in design but differ in the stated bandwidth of the photodiode in each unit—30 MHz for the FOPH 500 and 100 MHz for the FOPH 2000. Pressure waveforms were measured with the fiber parallel to and the sensitive cleaved face of the fiber perpendicular to the HIFU axis. Measured waveforms were deconvolved using the impulse responses of the hydrophones provided by the manufacturers. Calibration of FOPH signals from voltage to acoustic pressure in water was performed using the manufacturer's calibration equations and protocols.[41]

D. Waveform measurements in gel phantom

Although fiber optic hydrophones have been employed in various calibration measurements of high-power lithotripsy and HIFU fields in water,[20,41] no results have been reported on the application of such techniques to measurements in tissues or phantoms. In this study, the FOPH was used to measure pressure waveforms at the focus of the HIFU source in a transparent tissue-mimicking gel phantom. To perform measurements in the phantom, the fiber tip was cast in a large rectangular mold that was designed for HIFU dosimetry experiments with tissue phantoms. The mold's dimensions were 6 cm in the axial direction, 5 cm in the transverse direction, and 13 cm in the other transverse direction. Preparation of the gel phantom comprised three steps: degassing of the unpolymerized liquid mixture, positioning of the FOPH tip in the liquid, and lastly addition of a polymerization agent. For these experiments, the propagation path from the transducer to the focus included water, then 33 mm of the phantom.

The FOPH signal measured in the gel phantom was converted from voltage to acoustic pressure based on the method described by the manufacturer for measurements in water. However, the method could not be directly applied to measurements in gel. Therefore, we discuss here the main equations and explain how they can be modified for using the FOPH in gel phantom.

The output of the FOPH photodiode is a voltage that is proportional to the reflection coefficient of laser intensity, R, at the FOPH tip. If the tip is embedded in some medium, the calibration of the measured signal is performed based on three equations.[41] The first equation is the relation between the reflection coefficient, R, and the optical refractive index of the medium:

$$R = (n - n_g)^2/(n + n_g)^2. \tag{9}$$

Here n_g is the refractive index of the glass fiber and n is the refractive index of the surrounding medium. The acoustic measurements rely on the fact that n is a function of the medium density, ρ, which changes with acoustic pressure: $\rho=\rho(p)$.

The second equation is the relation for refractive index as a function of density, which can be obtained based on the Gladstone–Dale model:

$$\frac{n(\rho) - 1}{\rho} = \text{const} = \frac{n_0 - 1}{\rho_0}. \tag{10}$$

The static refractive index $n_0=n(\rho_0)$ of the gel phantom in Eq. (10) was calculated from Eq. (9) by comparing dc voltage output levels of the photodiode with the FOPH tip embedded in the gel phantom or placed in water, which has a known refractive index of $n_0=1.329$ at 20 °C. Using this method, the refractive index of the gel was measured as $n_0=1.357$.

The third equation used for calibration of the hydrophone is the relation between the change of the density and acoustic pressure $\rho=\rho(p)$. It can be used in the form of the Tait equation $(P+Q)/\rho^{\gamma}=\text{const}$, where P is the sum of the acoustic and ambient pressure and Q and γ are constants. The manufacturer used the Tait equation for water measurements with parameters $Q=295.5$ MPa and $\gamma=7.44$.[41] However, the values of Q and γ are not available for the gel. Moreover, the Tait parameters (Q,γ) for water were originally obtained by fitting experimental data for which pressure varied isothermally up to 600 MPa using pressure steps of 50 MPa.[42,43] For the pressure range less than 100 MPa that typifies HIFU and lithotripsy, the inaccuracy of this approach is obvious from the fact that the Tait equation gives the wrong value of the nonlinear parameter of water $\beta=(\gamma+1)/2\approx 4.2$ at 20 °C whereas acoustic measurements give $\beta\approx 3.5$.[14]

The relation $\rho=\rho(p)$ in the gel was obtained from the expansion used in nonlinear acoustics to second order:[14]

$$\rho = \rho_0 + \left(\frac{1}{c_0^2}\right)p - \left(\frac{\beta - 1}{\rho_0 c_0^4}\right)p^2 + \cdots. \tag{11}$$

The neglected terms in Eq. (11) depend on the acoustic pressure to a power greater than 2 and on entropy changes, which are insignificant for pressure changes of less than 100 MPa, as considered in this paper. The necessary parameters of the medium (ρ_0, c_0, β) in Eq. (11) are known for both water and gel.[14,33]

The combination of Eqs. (9)–(11) is used here to determine the change of the refractive index with the acoustic pressure in the gel phantom. For measurements in water, calibrations using both the Tait equation and Eq. (11) were compared. At the highest output level ($p_0=0.57$ MPa), peak pressures of 61.8 and 60.4 MPa were obtained with the Tait equation and Eq. (11), respectively. The small discrepancy of about 2% validates the calibration approach for the gel phantom using Eq. (11). For measurements in water, calibrations were performed with the Tait equation.

The calibration described above is applicable only to low frequency measurements when the fiber diameter is much smaller than the acoustic wavelength. At higher frequencies, the signal is distorted because of diffraction of the acoustic wave at the fiber tip.[44] Additional damping of high frequencies is introduced by the FOPH electronics. To compensate for these effects, a deconvolution procedure must be applied using the frequency response of the hydrophone. For measurements in water, this response was provided by the manufacturer. For measurements in gel, the response was assumed to be the same because the acoustic properties of the gel and water (provided in Sec. II) are very close and therefore diffraction effects at the fiber tip should be similar.

III. RESULTS

A. Low amplitude calibration of the HIFU source in water

The radius of curvature and diameter of the transducer were nominally stated by the manufacturer as 44.5 mm and 44.45 mm, respectively. However, both the transducer housing and surface waves might limit and distort the vibration of the source. Therefore, to model the source as a uniform piston, it was necessary to obtain the effective operational aperture and radius of curvature of the source. These parameters were determined by measuring pressure distributions along the axis of the HIFU beam and transverse to the axis in the focal plane. Measurements were acquired with the SEA hydrophone under low amplitude linear propagation conditions. The acoustic axis of the transducer relative to the mechanical axis of the positioning system was determined using methods described by Cathignol *et al.*[45] to ensure that the axial scan corresponded to the true axial field distribution. Measured distributions were compared to results of simulations based on a linearization of Eq. (1). Simulations were performed for various values of the source aperture and radius of curvature to obtain the "best fit" with hydrophone measurements in the focal zone. The effective aperture of the transducer, $2r_0$, and radius of curvature, F, were found to be 40.0 mm and 44.4 mm, respectively. A comparison of measurements and simulation results for the one dimensional axial and transverse focal field is shown in Fig. 2. In addition, pressure distributions calculated using analytic O'Neil solutions for the linear focused field[11] (with $2r_0=42$ mm and $F=44.4$ mm) are also shown in Fig. 2. The results of both models, calculated with the best fit aperture and radius of curvature, are in good agreement with the experimental data within the focal region of the pressure field of the HIFU transducer. These plots demonstrate that simplified models of uniformly vibrating focused transducers can be applied to model the focal zone of a single element HIFU source. However, discrepancies between simulations and measurements

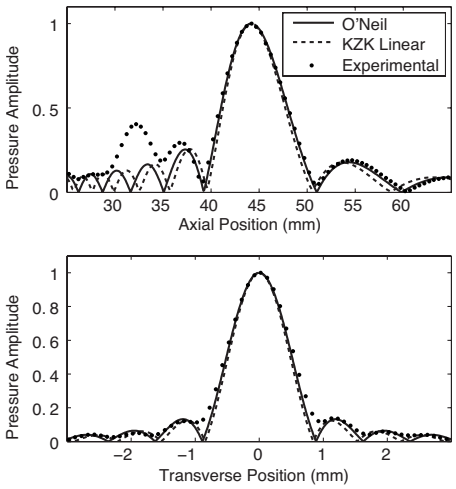


FIG. 2. Comparison of axial and focal scans of the low amplitude ("linear") pressure field measured in water by the SEA hydrophone and calculated with the linearized KZK equation, along with the O'Neil analytic solution (Ref. 11). The beam plots were used to determine the radius of curvature and aperture of the source input to the KZK model.

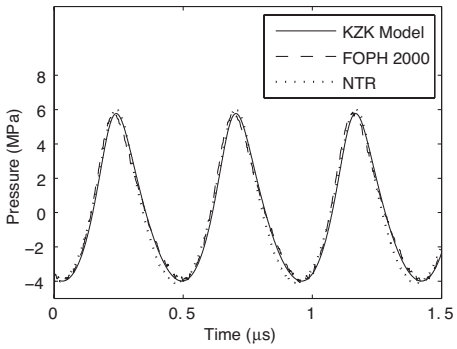


FIG. 3. Comparison of focal waveforms simulated with the KZK model and measured with the FOPH 2000 and NTR hydrophones in water for a source pressure amplitude of $p_0=0.1$ MPa ($I_L=700$ W/cm^2, $I_N=720$ W/cm^2). Under slightly nonlinear propagation conditions, the focal pressure level was sufficiently high to be measured with two calibrated hydrophones.

in the axial prefocal distributions were observed; therefore, more precise boundary conditions should be employed in the modeling to capture the nearfield of a real transducer with better accuracy. The location of the spatial pressure peak pressures determined by both modeling and measurements essentially coincided with the geometric focus of the transducer. Consequently, further measurements and calculations of the focal waveforms were performed at the distance $z=F$ from the source.

After establishing the effective radius of curvature and aperture of the source, the relationship between the voltage applied to the source and the source pressure amplitude was determined. First, a focal maximum pressure (spatial peak) of 0.47 MPa was measured using the NTR hydrophone at a low driving voltage for linear propagation conditions. The hydrophone was positioned at the geometric focus of the source (44.4 mm) using time-of-flight measurements. The linear focusing gain of the source as defined by Eq. (8) was determined from the linear modeling results and was found to be $G=48$. Hence, the source pressure amplitude in the model that corresponds to a focal peak pressure of 0.47 MPa can be calculated as $p_0=p_F/G=0.0097$ MPa. As calculated from Eqs. (6) and (7), a peak focal pressure of 0.47 MPa corresponds to focal intensities of $I_L=I_N=7$ W/cm^2. For other output levels, the source pressure amplitude, p_0, was assumed to be linearly proportional to the driving voltage to the source.[23]

To compare calibration of the NTR and FOPH 2000 hydrophones, further measurements of focal waveforms were performed with both hydrophones at a higher source level, $p_0=0.1$ MPa ($I_L=700$ W/cm^2, $I_N=720$ W/cm^2). Figure 3 shows the focal waveforms obtained with the NTR and FOPH 2000 hydrophones as well as with the KZK nonlinear model. At this output level, the waveforms are slightly distorted due to nonlinear propagation. All three waveforms show very good agreement with a peak positive pressure of 6 MPa and a peak negative pressure of −4 MPa. This comparison demonstrates that under conditions of nonlinear propagation, the numerical modeling agrees with experimental measurements, while the FOPH calibration is accurate in comparison with a reference hydrophone.

B. High amplitude calibration of the HIFU source in water and gel phantom

At pressures above those described above, waveforms were measured with the FOPH 2000 and modeled in both water and the gel phantom. Measurements and simulations included source pressure levels up to $p_0=0.57$ MPa.

1. Water measurements and modeling

Figure 4 shows the focal waveforms and corresponding envelopes of the spectra measured and modeled for $p_0=0.29$ MPa ($I_L=6500$ W/cm^2, $I_N=8200$ W/cm^2), $p_0=0.39$ MPa ($I_L=11\,000$ W/cm^2, $I_N=16\,000$ W/cm^2), and $p_0=0.57$ MPa ($I_L=24\,000$ W/cm^2, $I_N=29\,000$ W/cm^2). At an initial pressure amplitude of $p_0=0.29$ MPa, the experimental and simulated waveforms are noticeably distorted with a measured peak positive pressure of 35 MPa and a peak negative pressure of −10 MPa and show excellent agreement. At higher-power levels, the waveforms also agree well, particularly in the smooth rarefaction part of the waveform and corresponding peak negative pressure values. However, some discrepancy between the measurement and modeling is observed within the sharp positive part of the waveform. The peak positive pressure obtained with modeling is consistently higher than in the measured waveform. For $p_0=0.39$ MPa, the peak positive pressure modeled is 63 MPa, which is 10 MPa or 19% higher than the measured pressure of 53 MPa. The peak negative pressure for both waveforms is −11 MPa. For $p_0=0.57$ MPa, the peak negative pressure is −14.5 MPa, but the peak positive pressure is 25% higher in the model (81 MPa versus 65 MPa from the measurement). The discrepancy between measurements and modeling is also apparent in the frequency domain, wherein measurements consistently exhibit less energy content at high frequencies.

Distributions of the peak positive $(p+)$ and peak negative $(p-)$ pressures were also measured and modeled along the acoustic axis and in the transverse focal plane. In Fig. 5,

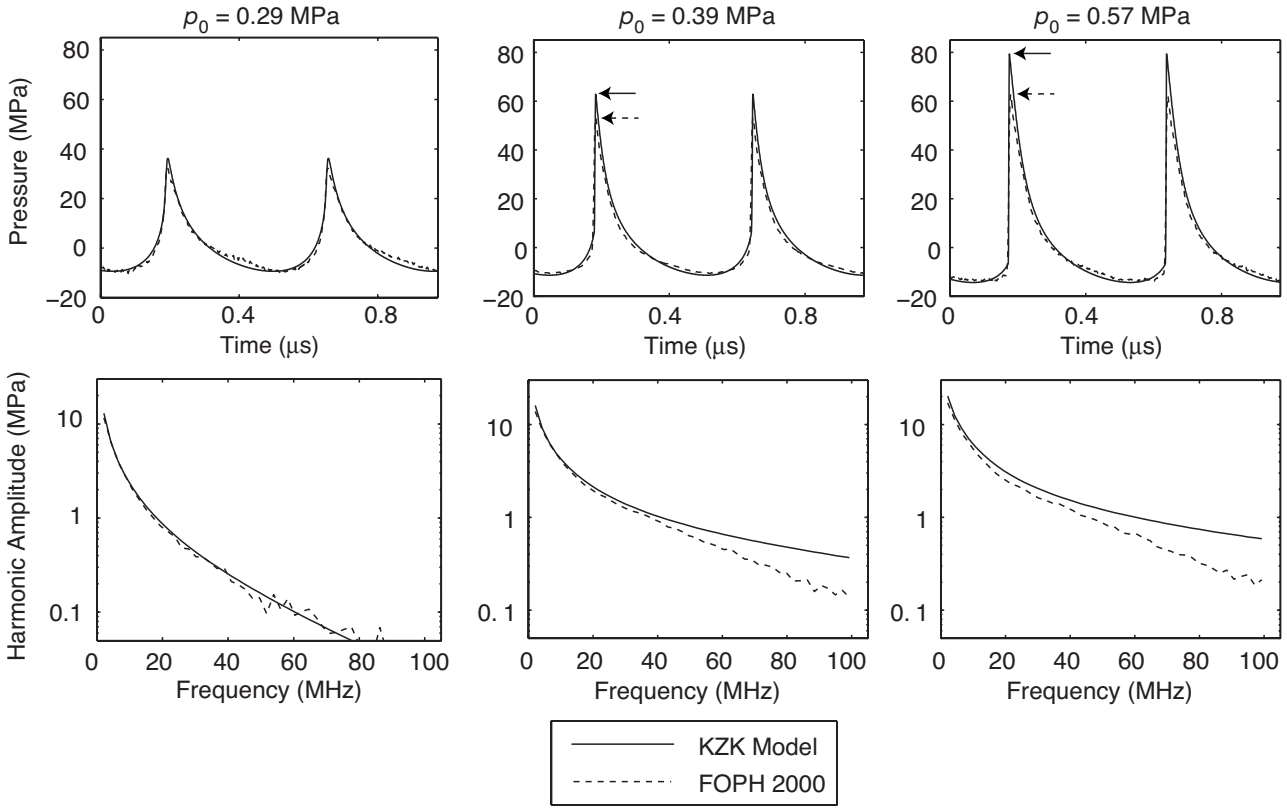


FIG. 4. Comparison of focal waveforms and corresponding spectra simulated with the KZK model and measured with the FOPH 2000 in water for p_0 =0.29 MPa (I_L=6500 W/cm^2, I_N=8200 W/cm^2), p_0=0.39 MPa (I_L=11 000 W/cm^2, I_N=16 000 W/cm^2), and p_0=0.57 MPa (I_L=24 000 W/cm^2, I_N =29 000 W/cm^2). The horizontal arrows in the waveform plots depict the peak positive pressure measured and modeled. At these output levels, the waveforms are strongly distorted, contain shocks, and the harmonic content of the waveforms extends beyond 100 MHz. The agreement between modeling and measurement results is good, although, when sharp shocks are present (p_0=0.39 and 0.57 MPa), the measurements show lower peak positive pressures and lower values of harmonic amplitudes at high frequencies.

axial and transverse peak pressure distributions are shown along with distributions of the first four harmonics for p_0 =0.39 MPa. As indicated by Fig. 4 for the same drive level, asymmetric shocks develop in the focal region and the measured peak positive pressure is slightly lower in value than the modeled one. Both experiment and simulation show that the peak positive pressure has a dramatic increase in gain from the linear value of G=48 to G_{p+}=130 in measurements and G_{p+}=160 in modeling. The −6 dB spatial distribution of the peak positive pressure becomes less than half the width in the transverse direction (0.4 mm in modeling versus 1 mm in the linear case) and shorter in the axial direction (35 mm in modeling versus 70 mm in the linear case) compared to linear propagation conditions, which are shown in Fig. 2.

The measured and modeled distributions of the first four harmonics from Fig. 5 show good agreement and demonstrate that higher frequencies are increasingly localized. Only four harmonics are shown in Fig. 5, but as seen in Fig. 4, the focal waveform has a frequency content beyond 100 MHz. Stronger focusing of higher harmonics will result in enhanced localized heating of the propagation medium because the absorption coefficient grows with frequency, as described by Eqs. (2) and (3). This effect will significantly increase if shock fronts develop. Figure 4 shows qualitatively that the shock amplitude is nearly equal to the value of the peak positive pressure. Because heat deposition at a shock front is proportional to the cube of its amplitude,[14] the spatial distribution of heating in the presence of shocks will be even sharper than the highly localized distribution of peak positive pressure. Moreover, an underrepresentation of the peak positive pressure and thus of the shock amplitude in measurements can have a large impact on predictions of heat deposition rates.

The agreement in the distribution of peak negative pressure is excellent between measurements and simulations. In contrast with the impact of nonlinear propagation on peak positive pressures, the focusing gain for the peak negative pressures is reduced to G_{p-}=30 and the −6 dB size of the focal zone (1.25 mm in the transverse direction and 80 mm in the axial direction) is larger than in the linear case. Because peak negative pressure in the acoustic waveform is primarily responsible for cavitation effects, the volume of cavitation in nonlinear focused fields is much less localized than that of absorptive heating. To characterize the acoustic field with regard to possible cavitation effects, direct measurements or nonlinear modeling of focal waveforms is necessary to avoid overestimation of peak negative pressures.

A summary of the measured and modeled spatial peak pressures over the range of power levels used in experiments and modeling is shown in Fig. 6. Peak positive and negative pressure values are normalized by the source pressure amplitude and therefore represent the focusing gain of the transducer. The focusing gains for p+ and p− are plotted versus the source pressure amplitude. The error bars show the stan-

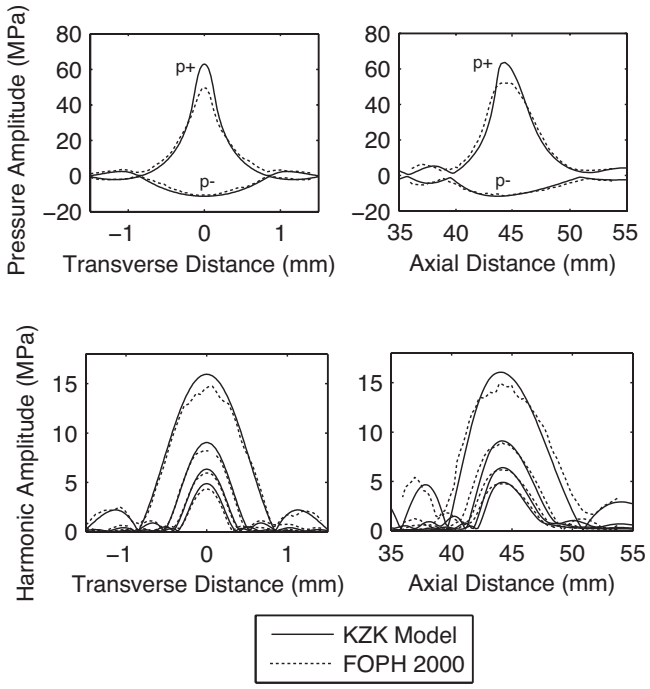


FIG. 5. Comparison of measured and modeled pressures axially and in the focal plane in water for $p_0=0.39$ MPa ($I_L=11\,000$ W/cm^2, $I_N=16\,000$ W/cm^2). Shown on the left are the transverse distributions for the peak positive, p_+, and peak negative, p_-, pressures as well as for the first four harmonics. The axial data are shown on the right. The combination of nonlinear propagation and diffraction results in narrowing of the harmonic beamwidths and asymmetry of the waveform at the focus, which was captured by both measurement and modeling.

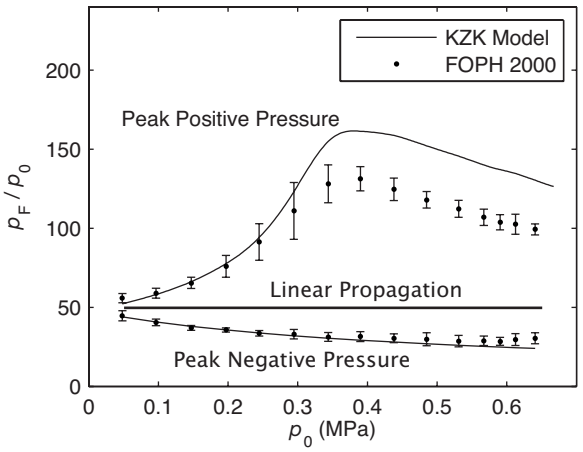


FIG. 6. Summary of focal pressures measured and modeled with increase in the source operation level. The pressures are normalized by the source pressure amplitude, p_0, illustrating the change in the focusing gain due to combined nonlinear and diffraction effects. The error bars indicate the standard deviation of three different measurements. Nonlinear propagation effects, which are stronger at higher p_0, lead first to a higher gain in peak positive pressure. However, the gain falls off as shocks form prefocally, leading to increased attenuation before reaching the focus. The focusing gain for the peak negative pressure monotonically decreased as the source pressure amplitude increased.

dard deviations of three separate measurements. At low source pressure, the linear focusing gain of the transducer is 48 for both peak positive and peak negative pressures. Under nonlinear conditions, the peak positive pressure focusing gain grows higher (a maximum of 160 is obtained in modeling) and then decreases as shocks form proximal to the focus and result in higher prefocal losses. Meanwhile, the gain for peak negative pressures monotonically decreases with increasing source pressure amplitude. Figure 6 again illustrates that the measured and modeled peak negative pressures show an excellent agreement over the range of input source amplitudes, while the peak positive pressures are lower in the measurement than in the modeling.

2. Gel measurements and modeling

Before waveform measurements were made in the gel phantom, an experiment was performed to test the calibration method that is described in Sec. II. In this experiment, the focal waveform was measured with the FOPH tip in both water and gel phantom and then the calibrated waveforms were compared. First, the FOPH tip was cast in a 7% BSA-acrylamide gel phantom. The gel phantom was cast in the mold described in Sec. II, but was only 1 cm thick in the axial direction. The FOPH tip was cast so that it was embedded less than 1 mm into the gel phantom. The HIFU transducer was aligned so that the FOPH tip was positioned at the spatial maximum of the peak positive pressure and waveform measurements were recorded up to a source level of $p_0=0.39$ MPa. Next, measurement of the focal waveform was taken with the FOPH tip positioned in water, less than 1 mm behind the 10 mm thick slab of gel phantom. The voltage signals measured in water and in gel were calibrated following the method of Sec. II and compared for the same source pressure levels. Note that the difference in absorption between the two cases is negligible, but the pressure slightly changes in transmission from gel to water. The corresponding transmission coefficient is $T=2Z_{\text{water}}/(Z_{\text{water}}+Z_{\text{gel}})=0.96$, where $Z=\rho c$ denotes the acoustic impedance of the water or gel. The relevant material parameters used to calculate Z are provided in Sec. II. When waveforms were compared over a range of source levels, they showed very good agreement (within 5%), thus validating the proposed calibration approach for using the FOPH to measure pressure waveforms in the gel phantom.

The method of testing the theoretical gel calibration described above can also be viewed as a "substitution" calibration of the FOPH in the new medium (gel phantom). The calibration equations of the FOPH and the results of waveform measurements "in" and "out" of the gel phantom lead to the result that after the waveforms are deconvolved with the impulse response of the hydrophone, the transfer function from voltage to pressure is nearly linear. If the calibration for water was used for measurements in gel, then pressure waveforms in gel were linearly shifted to a value 20% higher. The "linear" sensitivity of the FOPH thus was found to be 20% higher in gel than in water. Accordingly, the FOPH calibration for water can be corrected for measurements in gel without requiring the explicit calibration technique described in Sec. II D.

The measurements and modeling in the gel phantom with the FOPH 2000 show very good agreement. At low drive levels, the agreement was excellent. However, at higher-power levels when shocks were present, there was again a discrepancy between the measured and modeled peak

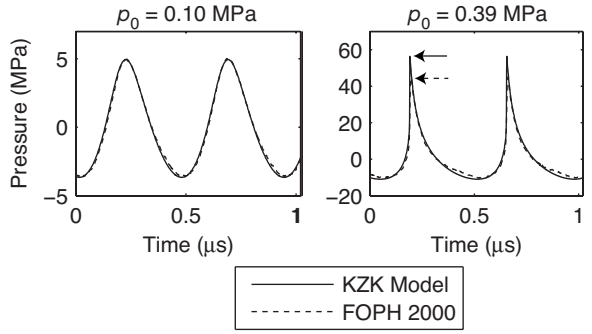


FIG. 7. Comparison of focal waveforms measured and modeled in a 7% BSA-acrylamide tissue-mimicking gel phantom for p_0=0.1 MPa (I_L=I_N =560 W/cm^2) and p_0=0.39 MPa (I_L=9100 W/cm^2, I_N=12 000 W/cm^2). The arrows in the p_0=0.39 MPa plot indicate the peak pressures measured and modeled. The agreement is good, but the peak positive pressure is lower for the measured waveform when the shock is present. The waveforms were able to be measured in the phantom, and although of lower amplitude because of increased attenuation, they are of identical shape to comparable amplitude waves measured in water.

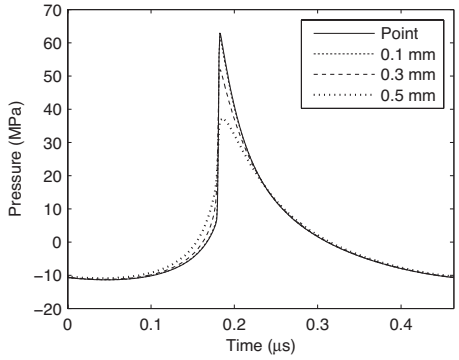


FIG. 8. The waveforms for p_0=0.39 MPa calculated using the KZK model at focus and averaged over 100, 300, and 500 μm diameter circular cross sections, representing increasing hydrophone sizes. Spatial averaging over the finite hydrophone size caused less than <2% reduction in peak amplitude at the narrowest beamwidth for the 100 μm diameter FOPH hydrophone but would cause a significant reduction with larger hydrophones.

positive pressures in the waveforms. Figure 7 depicts one such comparison of a focal waveform in the gel phantom at p_0=0.1 MPa (I_L=I_N=560 W/cm^2 *in situ*) and p_0 =0.39 MPa (I_L=9100 W/cm^2, I_N=12 000 W/cm^2 *in situ*). The peak positive pressure in the shocked waveform is 46 MPa by measurement and 56 MPa (i.e., 22% higher) by modeling; the peak negative pressure for both waveforms is −11 MPa. The peak pressures in gel are lower than in water for the same values of p_0 because of the higher attenuation in gel (see Figs. 3 and 4). These results demonstrate that measurements can be made in a tissue-mimicking phantom and that the FOPH can be calibrated in media other than water using the proposed method based on medium properties (β,n_0).

C. Reasons for discrepancy in measured and modeled peak positive pressure

The results summarized in Fig. 6 for measurements in water indicate that at lower drive levels (p_0<0.25 MPa) there is very good agreement between the simulated and measured focal waveforms. At higher drive levels, the values of the peak positive pressure and thus the shock amplitude are lower in the experiment than in the modeling. Accurate measurement of the peak positive pressure is important for predicting thermal effects during HIFU treatments. For example, a 25% difference in measured and modeled values of p_+ in Fig. 4 results in a twofold difference in the heat deposition rate at the focus, as calculated using weak shock theory.[17] Several hypotheses for the discrepancy in focal waveforms between measurement and modeling were tested and are discussed below.

1. Spatial averaging

One of the factors that can affect acoustic measurements is the finite size of the receiver used to measure pressure waveforms.[9] If the acoustic field at the point of measurement does not have a uniform planar pressure distribution over the active area of the receiver, then the true acoustic pressure at the focus can be underestimated. This effect is referred to as spatial averaging.[46–48]

Although the 100 μm diameter FOPH hydrophone used in this study is much smaller than the transverse dimensions of the acoustic field under linear acoustic propagation conditions (1 mm at −6 dB level), it may not be sufficiently small in the nonlinear field. For example, the field distribution for the peak positive pressure becomes narrower with increasing source pressure amplitude. Thus, the measurement would become increasingly sensitive to spatial averaging at higher output levels. This hypothesis is consistent with the increasing discrepancy in measured and modeled focal values of peak positive pressure for higher values of the source pressure amplitude. The results of modeling show that the narrowest distribution of the peak positive pressure occurs at p_0=0.39 MPa, in which case the shock fronts form very close to the focus. This output level corresponds to the maximum focusing gain of p_+ (shown in Fig. 6). For higher source pressures, shocks form prefocally and the beam width at the focus widens.[15]

To determine whether spatial averaging had an effect on the measurement of focal waveforms, the KZK model was used to predict the waveform that would be measured by a hydrophone of finite diameter. Waveforms were simulated in the focal plane around the focus and then were averaged over a circular area corresponding to virtual hydrophones of a range of diameters. The results of these simulations are shown in Fig. 8. The predicted waveforms that would be measured by 100 (FOPH diameter), 300, and 500 μm (commonly used in lithotripsy[49]) diameter hydrophones are shown. The predicted focal waveform for a 100 μm diameter hydrophone was only 2% lower in peak positive pressure than the waveform predicted for an ideal "point" receiver at p_0=0.39 MPa, the drive level at which the −6 dB positive pressure region was the smallest. However, when a 500 μm diameter hydrophone was simulated, there was a significant distortion of the measured waveform. Spatial averaging results mainly in a decrease in the measured value of the peak positive pressure and a widening of the shock front. Never-

theless, for the conditions of this study, it was determined that spatial averaging was not a significant source of error in the measured waveforms.

2. Nonuniform vibration of the HIFU source

In the modeling described in Sec. II, it was assumed that the vibration pattern of the transducer surface was uniform. However, few real transducers operate as perfect uniformly vibrating pistons. The piezoceramic transducer used in this study, like others, had a nonuniform vibration pattern due to the excitation of surface waves.[38,45] Nonuniform vibration can influence the acoustic field generated by the device and lead to prefocal differences between the actual acoustic field and the calculated field assuming uniform vibration. The effect of nonuniform vibration can be observed in the axial pressure distributions shown in Fig. 2, in which the measured axial scan has two prefocal peaks that are different than those obtained using the O'Neil formula. Since nonlinear effects accumulate as the acoustic wave propagates to the focus, the discrepancy observed in modeling the transducer nearfield may affect the focal waveform.

To determine whether nonuniform source vibration had an effect on modeled waveforms, the transducer vibration pattern was measured using acoustic holography[39] and included in the numerical model. The holographic technique was implemented by measuring the amplitude and phase of a small amplitude signal in a plane located at a distance of 30 mm in front of the transducer perpendicular to its acoustic axis. The measurements were performed in water using the SEA hydrophone. A square grid, 60 mm in width, was raster scanned with a step size of 0.3 mm. The Rayleigh integral was then used to reconstruct the velocity distribution at the plane $z=0$ (where the boundary conditions to the KZK model are specified) by backpropagating the measurements obtained at $z=30$ mm. Figure 9 depicts the normal velocity distribution in the plane $z=0$ after the reconstruction. The annular pattern shown in Fig. 9 demonstrates that the transducer did not operate with a uniform velocity distribution.

The reconstructed nonuniform velocity distribution was incorporated into simulations by altering the boundary conditions of the KZK model. First, the two dimensional velocity distribution was transformed to an axisymmetric form by radially averaging over 300 equally spaced rings. The number of rings used in the model was chosen to be similar to the step size of the grid used for holographic reconstruction. Next, the source pressure amplitude in the numerical model was adjusted so that the focal pressure predicted by linear modeling matched experimental measurements. Figure 9 shows the uniform and nonuniform boundary conditions used in the model as well as the axial pressure distribution of the transducer. For the axial distribution, the experimentally measured result is shown as well as the results of linear modeling using both uniform and nonuniform boundary conditions. The agreement between both uniform and nonuniform models with measurements is excellent in the focal region, although prefocally, the uniform and nonuniform models yield different pressure patterns. The results of the nonuniform model in the nearfield are closer to measurements, but a discrepancy is still observed. Imperfect agreement in the nearfield is caused by the parabolic approximation used to govern diffraction effects in the KZK equation.

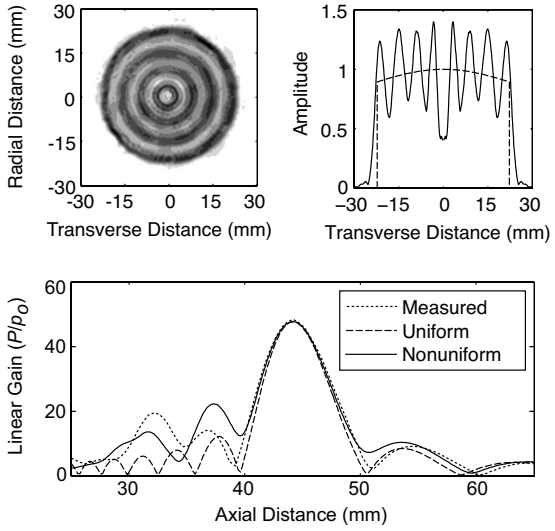


FIG. 9. The reconstructed velocity distribution across the transducer surface from low amplitude measurements using acoustic holography (top left). The surface is seen to have very periodic radial surface waves where black color corresponds to higher and white to lower values of velocity. A cross section of the radially averaged source velocity that was used as the boundary condition in the numerical model assuming uniform (dashed) and nonuniform (solid) vibrations (top right). Comparison of low amplitude axial pressures measured and modeled with uniform and nonuniform boundary conditions (bottom).

Nevertheless, further simulations performed up to the highest source pressure level considered in this work indicate that nonuniform boundary conditions did not have a significant effect on calculated focal waveforms. The excellent agreement is illustrated by Fig. 10, which shows the axial distributions for the peak positive (the curves above the "zero" axis) and peak negative pressures (the curves below the zero axis) as well as focal waveforms modeled assuming either uniform ($p_0=0.39$ MPa) or nonuniform boundary conditions. Although there is disagreement between the results of the two models prefocally, the pressures within the focal region agree very well. Most nonlinear effects occur within the high amplitude focal region; therefore, differences that occur prefocally do not contribute significantly to nonlinear distortion of the pressure field at the focus. The calculated focal waveforms, shown in Fig. 10, are almost identical. The difference in peak positive pressure values of the waveforms is less than 5%. Thus, nonuniform source vibration was not a significant source of error in this study.

3. Frequency response of the FOPH

In the experiments performed in this paper, the FOPH fiber was oriented parallel to the HIFU acoustic axis. In this orientation, as compared to the case when the FOPH tip is at an angle with respect to the acoustic axis, the effect of averaging of the received signal along the FOPH face is minimized. However, besides spatial averaging, diffraction effects at the hydrophone tip also alter the measured waveform. An effective doubling of the measured pressure occurs at frequencies above 10 MHz, resulting in a nonuniform frequency response of the hydrophone.[44] To determine

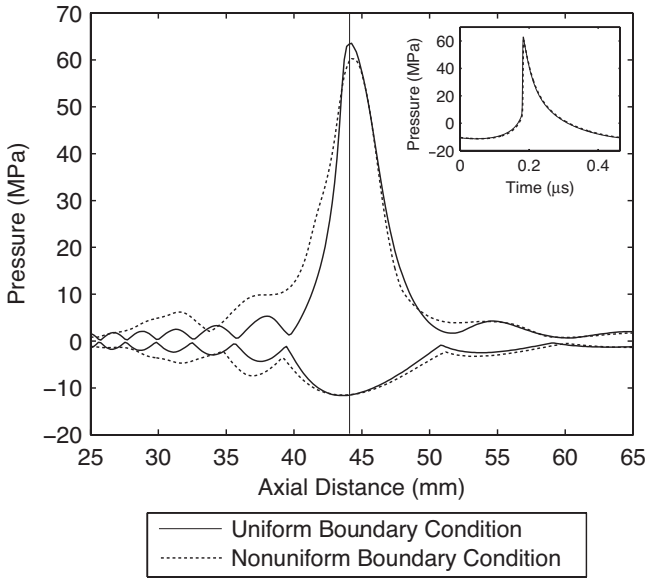


FIG. 10. Axial distributions of the peak positive and negative pressures calculated for p_0=0.39 MPa with uniform and nonuniform boundary conditions. The vertical line in the picture indicates the location of the focus, where pressure waveforms are calculated under the same conditions and shown on the top right. The nonuniform boundary condition is more accurate in modeling nearfield pressure patterns and may be useful for predicting side effects such as skin burns. However, the focal waveforms are almost indistinguishable when calculated with both uniform and nonuniform boundary conditions.

the actual acoustic pressure waveform, the measured signal must be deconvolved with the impulse response of the hydrophone. For the fiber optic hydrophones used in this study, the impulse responses provided by the manufacturer were used to deconvolve the measured signal. Typical shocked waveforms obtained with the FOPH 2000 system are shown in Fig. 11. The figure depicts a waveform that is calibrated to acoustic pressure with and without deconvolution with the FOPH impulse response. If deconvolution is not performed, the measured peak positive pressure is overpredicted and the waveform shape is distorted. Thus, deconvolution is necessary to obtain accurate pressure waveforms when using the FOPH.

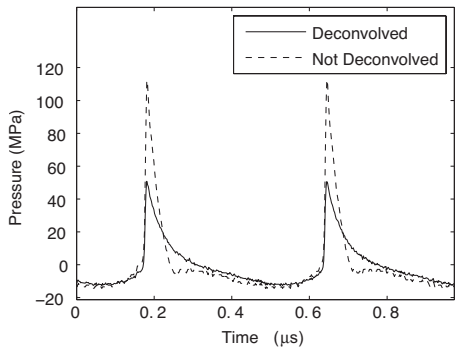


FIG. 11. A typical focal waveform measured using the FOPH hydrophone at p_0=0.39 MPa calibrated with and without deconvolution with the manufacturer's impulse response. At low values of p_0, where the wave is sinusoidal, the need for deconvolution is not obvious and a linear calibration from volts to pressure can be used. However, at higher p_0, the directly measured waveform significantly overpredicts the peak positive pressure and distorts the waveform shape; therefore deconvolution is necessary.

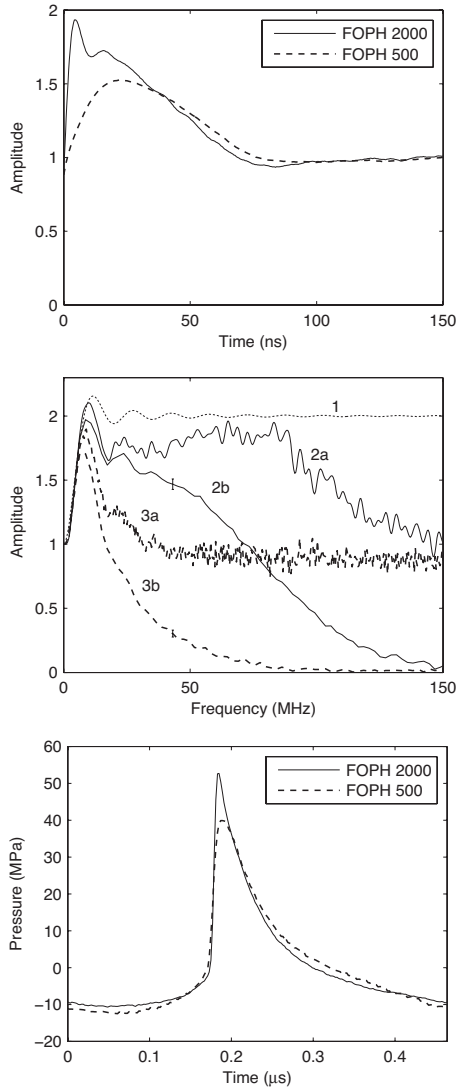


FIG. 12. The step response of the FOPH 500 and FOPH 2000 hydrophones calculated by integrating the manufacturer's provided impulse response (top). A comparison of the transfer function of the FOPH hydrophone (middle). The figure is composed of five different curves. Labeled (1) is the theoretical transfer function resulting from diffraction at a 100 μm tip (Ref. 44). The frequency transfer functions for the FOPH 2000 (2) and FOPH 500 (3) calculated from the step response (a) and obtained by comparison with the model predictions (b). Comparison of focal waveform measured with the FOPH 2000 and FOPH 500 hydrophones for p_0=0.39 MPa (bottom). The smaller bandwidth of the FOPH 500 hydrophone results in a diminished peak positive pressure and shock amplitude as compared to the FOPH 2000.

Thus far, only measurements with the FOPH 2000 have been presented. To examine the effect of different bandwidth hydrophones on pressure measurements, the focal waveforms were recorded using two models of fiber optic hydrophones. At low-power levels, where the focal waveforms were nearly sinusoidal, the agreement between the two hydrophones was excellent. As the power increased and shocks developed at the focus, the FOPH 500 measured shocks with longer rise times and smaller values of peak positive pressure than those obtained with the FOPH 2000. The bandwidth difference between the two hydrophones is illustrated in Fig. 12, which is composed of three subfigures—the step response, the transfer function, and a waveform comparison. The step response of the FOPH 500 and FOPH 2000 hydro-

phones is shown, as calculated by integrating the impulse response provided by the manufacturer. The transfer function of the FOPH hydrophone is shown and is composed of five plots. The first plot, labeled (1), is the published result that accounts for diffraction at a 100 μm diameter hydrophone oriented perpendicular to the propagation direction.[44] The next two plots depict the transfer function of the FOPH 2000. The first one (2a) was calculated by taking the discrete Fourier transform of the manufacturer's provided impulse response. The second (2b) was calculated by dividing the spectra of the measured peak focal waveforms without deconvolution by the modeled peak focal waveforms when shocks formed. The curve (2b) is an average of five different measurement and model comparisons (using waveforms from p_0=0.39 MPa to p_0=0.57 MPa) and the error bar at 40 MHz depicts the standard deviation. Lastly, the FOPH 500 transfer function calculated using the impulse response (3a) and the measured and modeled results (3b) are shown in the same way as for the FOPH 2000. Comparison of the focal pressure waveforms measured with the FOPH 2000 and FOPH 500 hydrophones and deconvolved by the respective impulse responses provided by the manufacturer at p_0 =0.39 MPa is shown at the bottom of Fig. 12. The peak positive pressure measured with the FOPH 2000 is more than 10 MPa higher than that obtained with the FOPH 500 and thus closer to the value obtained in modeling (Fig. 4). As this result demonstrates, the bandwidth of the receiver is important for capturing the shock fronts and the peak positive pressure. Our conclusion is that despite the increased bandwidth of the FOPH 2000, the shock front at the focus of the HIFU source was still not resolved with sufficient accuracy.

IV. DISCUSSION

The Food and Drug Administration in the U.S. and other regulatory agencies have required waveform measurements in water as part of the approval process for US medical devices. Specific protocols have been described for diagnostic US (Ref. 8) and shock wave lithotripsy (SWL),[50] but not yet for HIFU. The US and SWL protocols alone are insufficient for HIFU because HIFU operates with higher amplitudes and is more focused than US. Moreover, even the US and SWL regulatory protocols are still debated despite the fact that such devices are in widespread clinical use.

The specific challenges of designing protocols for US device approval depend on what part of the waveform is critical to measure for determining bioeffects. The combined amplitude and duration of the negative pressure phase is responsible for cavitation, which contributes to bioeffects such as tissue injury in US and stone comminution in SWL. The spatially averaged intensity, I_{SAL}, has been used in HIFU as a metric for reporting exposure levels. In the linear case, the product of twice the pressure attenuation coefficient and the intensity describes the heating rate, and heating is the primary mechanism responsible for inducing bioeffects in HIFU treatments. However, at intensities typically used in HIFU (order of 1000 W/cm^2) and certainly at higher intensities used in newer HIFU devices (order of 10 000 W/cm^2), the heating depends on the frequencies generated due to nonlinear acoustic propagation. At these intensity levels, nonlinear effects lead to shock formation where energy loss at the shock is the dominant heating mechanism.[17] The energy loss at the shock is proportional to the shock amplitude cubed,[14] as opposed to the square of the pressure amplitude in linear acoustics. Hence, the heating rate becomes more sensitive to accurate measurement of shocks when they are present. Therefore, the peak positive pressure, which determines the peak of the shock, is perhaps the most critical feature in the HIFU waveform.

Comparison of the measured waveforms using two models of fiber optic hydrophones demonstrated that waveforms acquired with the wider bandwidth hydrophone resulted in better agreement between measurements and simulations. Even though agreement was significantly better, it was nevertheless concluded that modeling provided better accuracy in predicting shock waveforms than measurements. Additional evidence for trusting that the calculated peak pressures were more accurate than the bandwidth limited measurements has been demonstrated by comparing time to boiling in a gel phantom due to shock wave heating.[51] Initiation of boiling in milliseconds was detected using high speed video imaging and calculated using weak shock theory from the measured and modeled focal waveforms. The measured time to observe boiling agreed better with calculated time to achieve 100 °C when modeled, not measured waveforms were used.

However, it remains possible to improve the measurements made with the FOPH. In this work, each measured FOPH signal was deconvolved with a hydrophone impulse response provided by the manufacturer. It was concluded that the limited bandwidth of the hydrophones caused a reduction in the peak positive pressure. If the hydrophones were completely insensitive above a critical frequency, this would be the major measurement limitation. However, it is more likely that the hydrophone is simply less sensitive at high frequencies than was indicated by the impulse response. Figure 12 indicates that this is more likely the case, at least for the FOPH 2000. Therefore, the impulse response of the hydrophone can be refined to improve the accuracy of the measurements. Refined impulse responses can be determined directly from curves (2b) or (3b) in Fig. 12 or can be obtained from a recent publication that reports newly measured impulse response curves for fiber optic hydrophones.[52]

This paper underscores the importance of very broad bandwidth (and accurate correction for variable sensitivity in that bandwidth) and of small active spot size hydrophones when measuring HIFU waveforms at clinical excitation levels. Nonlinear acoustic propagation causes spectral broadening of the wave as well as narrowing of the transverse distribution of the peak positive pressure. The −6 dB peak positive pressure beam width was as small as 300 μm for the high gain 2 MHz source used in this study. For this case, the 100 μm active diameter of the FOPH was sufficient but HIFU devices that have higher gains or that operate at higher fundamental frequencies may require an even smaller hydrophone.

The position of the hydrophone for all of the focal waveform measurements presented in this paper was at the geo-

metric focus of the source and was not adjusted to compensate for possible small changes in the axial focal position of the spatial maximum that can occur with changes in source power. For a source with a high gain and small amplitude (linear) propagation, the geometric focus nearly corresponds to the location of the spatial pressure maximum on the acoustic axis. However, the location of the maximum can shift in the axial direction with an increase in the source amplitude. Moreover, the axial shift is different for the peak positive and peak negative pressures. From the modeling results, the accuracy of measurements performed in the geometric focus can be assessed by determining how much the pressure maxima shift axially over the range of source power levels used. Here, the maximum shift was less than 300 μm in either direction from the geometric focus. Furthermore, the difference in peak pressure values at the position of the true maximum, as compared to the geometric focus, was less than $<2\%$ for the peak positive pressure and negligible for the peak negative pressure.

In light of these measurement challenges, modeling provided a potentially more accurate determination of the pressure field in water at high amplitudes as well as the ability to readily determine the acoustic field in either absorptive phantoms or tissue. However, low amplitude measurements were necessary for obtaining the boundary conditions for the model. The source curvature and aperture were determined by matching the dimensions of the modeled and measured focal regions. Next, the source pressure amplitude was determined by matching the modeled and measured focal pressure; i.e., the source pressure p_0 was the measured focal pressure p_F divided by the linear gain G. Alternatively, the source pressure amplitude could be calculated using acoustic radiation force balance (ARFB) measurements by dividing the acoustic power of the device by the transducer area and then using the plane-wave relation between intensity and pressure.[53] The effect of focusing could be included in the force balance calculation using Beissner's correction factor.[54] However, when this calibration method was applied, the focal pressures predicted by the model were 22% higher than the measured pressures. This discrepancy occurs because part of the source power goes to sidelobes radiated by the nonuniform velocity distribution. These sidelobes are directed away from the focus and thus do not contribute to the focal pressure.[45] The ARFB method therefore does not provide an accurate calibration of the uniform piston model for simulating nonuniformly vibrating transducers. On the contrary, the method proposed in Sec. II is not influenced by the source nonuniformity.

A uniform piston model was used to simulate the focal field of a single element transducer. It was shown that even though the transducer vibrated nonuniformly, the simplified piston model provided a good agreement for pressure waveforms within the focal lobe both in water and in gel up to 24 000 W/cm^2 *in situ*. A holographic method was reported to refine the boundary conditions used in calculations by measuring and modeling the nonuniform vibration of the source as caused in this study by surface waves or as might be caused by individual array elements in a more complex source. The majority of clinical devices are much more complex and can involve hundreds to thousands of individual elements to allow for electronic steering of the transducer focus. The method of matching measured and modeled focal beamwidths and the techniques to model a nonlinear source described here provide a framework applicable to characterizing focal fields of these more complex sources.

A further development of the proposed method of HIFU characterization would be to better govern the nearfield of the HIFU source in modeling. The modifications would include a more accurate diffraction approach based on using the Rayleigh integral, extension of the code to nonaxisymmetric sources, and the use of acoustic holography to obtain appropriate boundary conditions for the model. For the single element transducer considered in this paper, we have shown that the holographic technique and incorporation of nonuniform boundary conditions in the numerical model did not have a significant effect on focal pressure values. However, for more complicated sources, the holographic method may be more significant for correctly determining focal pressures. Nonuniform source vibration or complicated array designs can also lead to larger pressures in the nearfield than expected with the uniform piston model. These nearfield peaks are important during HIFU surgery because they can contribute to skin burns or other unintentional tissue damage.[7]

V. CONCLUSIONS

Acoustic characterization of HIFU fields is difficult and a standard method has not yet been established. The root of the difficulty is nonlinear acoustic propagation effects, especially since the clinical trend has been toward increasing pressure amplitudes in HIFU.

In this paper, a new characterization method based on measurement and modeling was proposed and validated at clinical output intensities in water and in a tissue-mimicking phantom. In summary, the focal pressure and beam dimensions are first measured with a hydrophone at low output levels where acoustic propagation is linear. Using these measurements, the boundary conditions to the model can be defined. The model then can be used to determine pressure waveforms in and around the focus at higher source operation levels. With the described technique, the source can be modeled as a uniformly vibrating focused piston. It was further demonstrated that acoustic holography can be used to improve the boundary conditions to the model or to model more complex source geometries.

Simulations and measurements were compared in water and in tissue phantom at high source pressure levels and agreed well. To obtain these results, the FOPH, was extended to *in situ* pressure measurements in a material other than water. Strongly asymmetric waveforms with a peak positive pressure of up to 80 MPa and a peak negative pressure of up to 15 MPa were obtained both numerically and experimentally found in very good agreement. However, when steep shocks were present in the waveform at focal intensity levels higher than 6000 W/cm^2, lower values of the peak positive pressure were observed in the measured waveforms. The lower measured values were attributed to the broad, but still

limited, hydrophone bandwidth. It was therefore concluded that both measurement and modeling were needed. Low amplitude calibration measurements in water were necessary to establish boundary conditions for modeling, but at higher amplitudes, simulations of shocked waveforms were more accurate than measurements. In addition, once validated, the model can be used without further modification to translate the pressures determined in water to *in situ* pressures in tissue.

ACKNOWLEDGMENTS

The work at the Center for Industrial and Medical Ultrasound (CIMU) at the Applied Physics Laboratory has been supported by the National Space Biomedical Research Institute (SMS00402) in consortium agreement with the National Aeronautics and Space Administration and the National Institutes of Health (DK43881 and EB007643). The work at Moscow State University has been supported by the Russian Foundation for Basic Research (06-02-16860 and 08-02-00368) and the International Science and Technology Center (3691). The authors gratefully acknowledge the help of Andrew Proctor and Fran Olson in the design and construction of the experimental setup. They also wish to thank the members of CIMU and the members of the Consortium for Shock Waves in Medicine (CSWM).

[1]L. Poissonnier, J.-Y. Chapelon, O. Rouviere, L. Curiel, R. Bouvier, X. Martin, J. M. Dubernard, and A. Gelet, "Control of prostate cancer by transrectal HIFU in 227 patients," Eur. Urol. **51**, 381–387 (2007).
[2]J. Kennedy, F. Wu, G. ter Haar, F. Gleeson, R. Phillips, M. Middleton, and D. Cranston, "High-intensity focused ultrasound for the treatment of liver tumours," Ultrasonics **42**, 931–935 (2004).
[3]P. Huber, J. Jenne, R. Rastert, I. Simiantonakis, H. Sinn, H. Strittmatter, D. von Fournier, M. Wannenmacher, and J. Debus, "A new noninvasive approach in breast cancer therapy using magnetic resonance imaging-guided focused ultrasound surgery," Cancer Res. **61**, 8441–8447 (2001).
[4]R. Illing, J. Kennedy, F. Wu, G. ter Haar, A. Protheroe, P. Friend, F. Gleeson, D. Cranston, R. Phillips, and M. Middleton, "The safety and feasibility of extracorporeal high-intensity focused ultrasound (HIFU) for the treatment of liver and kidney tumours in a Western population," Br. J. Cancer **93**, 890–895 (2005).
[5]K. Hynynen and G. Clement, "Clinical applications of focused ultrasound—The brain," Int. J. Hyperthermia **23**, 193–202 (2007).
[6]R. Siegal, S. Vaezy, R. Martin, and L. Crum, "Therapeutic ultrasound, Part II. High intensity focused ultrasound: A method of hemostasis," Echocardiogr. **18**, 309–315 (2001).
[7]F. Wu, Z. Wang, W. Chen, J. Zou, J. Bai, H. Zhu, K. Li, F. Xie, C. Jin, H. Su, and G. Gao, "Extracorporeal focused ultrasound surgery for treatment of human solid carcinomas: Early Chinese clinical experience," Ultrasound Med. Biol. **30**, 245–260 (2004).
[8]National Electrical Manufacturers Association, "NEMA standards publication UD 2-2004, acoustic output measurement standard for diagnostic ultrasound equipment, Revision 3," (2004).
[9]G. Harris, "Progress in medical ultrasound exposimetry," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 717–736 (2005).
[10]C. R. Hill, J. C. Bamber, and G. ter Haar, *Physical Principles of Medical Ultrasonics*, 2nd edition (Wiley, London, 2004).
[11]H. O'Neil, "Theory of focusing radiators," J. Acoust. Soc. Am. **21**, 516–526 (1949).
[12]P. Christopher and E. Carstensen, "Finite amplitude distortion and its relationship to linear derating formulae for diagnostic ultrasound systems," Ultrasound Med. Biol. **22**, 1103–1116 (1996).
[13]F. Duck, "Estimating in situ exposure in the presence of acoustic nonlinearity," J. Ultrasound Med. **18**, 43–53 (1999).
[14]*Nonlinear Acoustics*, edited by M. Hamilton and D. Blackstock (Academic, London, 1998).
[15]D. Dalecki, E. Carstensen, and K. Parker, "Absorption of finite amplitude focused ultrasound," J. Acoust. Soc. Am. **89**, 2435–2447 (1991).
[16]K. Hynynen, "Demonstration of enhanced temperature elevation due to nonlinear propagation of focused ultrasound in dog's thigh in vivo," Ultrasound Med. Biol. **13**, 85–91 (1987).
[17]E. Filonenko and V. Khokhlova, "Effect of acoustic nonlinearity on heating of biological tissue by high-intensity focused ultrasound," Acoust. Phys. **47**, 541–549 (2001).
[18]G. Harris, "Medical ultrasound exposure measurements: Update on devices, methods, and problems," Proc.-IEEE Ultrason. Symp. **2**, 1341–1352 (1999).
[19]D. Bacon, "Finite amplitude distortion of the pulsed fields used in diagnostic ultrasound," Ultrasound Med. Biol. **10**, 189–195 (1984).
[20]Y. Zhou, L. Zhai, R. Simmons, and P. Zhong, "Measurement of high intensity focused ultrasound fields by a fiber optic probe hydrophone," J. Acoust. Soc. Am. **120**, 676–685 (2006).
[21]J. Parsons, C. Cain, and J. Fowlkes, "Cost-effective assembly of a basic fiber-optic hydrophone for measurement of high-amplitude therapeutic ultrasound fields," J. Acoust. Soc. Am. **119**, 1432–1440 (2006).
[22]M. Averkiou and R. Cleveland, "Modeling of an electrohydraulic lithotripter with the KZK equation," J. Acoust. Soc. Am. **106**, 102–112 (1999).
[23]S. Nachef, D. Cathignol, J. Tjotta, A. Berg, and S. Tjotta, "Investigation of a high intensity sound beam from a plane transducer: Experimental and theoretical results," J. Acoust. Soc. Am. **98**, 2303–2323 (1995).
[24]V. Khokhlova, R. Souchon, O. Sapozhnikov, and D. Cathignol, "Numerical modeling of finite-amplitude sounds beams: Shock formation in the near field of a cw plane piston source," J. Acoust. Soc. Am. **110**, 95–108 (2001).
[25]X. Yang and R. Cleveland, "Time domain simulation of nonlinear acoustic beams generated by rectangular pistons with application to harmonic imaging," J. Acoust. Soc. Am. **117**, 113–123 (2005).
[26]V. Khokhlova, A. Ponomarev, M. Averkiou, and L. Crum, "Nonlinear pulsed ultrasound beams radiated by rectangular focused diagnostic transducers," Acoust. Phys. **52**, 481–489 (2006).
[27]M. Averkiou and M. Hamilton, "Nonlinear distortion of short pulses radiated by plane and focused circular pistons," J. Acoust. Soc. Am. **102**, 2539–2548 (1997).
[28]V. Khokhlova, M. Bailey, J. Reed, B. Cunitz, P. Kaczkowski, and L. Crum, "Effects of nonlinear propagation, cavitation, and boiling in lesion formation by high intensity focused ultrasound in a gel phantom," J. Acoust. Soc. Am. **119**, 1834–1848 (2006).
[29]P. Meaney, M. Cahill, and G. ter Haar, "The intensity dependence of lesion position shift during focused ultrasound surgery," Ultrasound Med. Biol. **26**, 441–450 (2000).
[30]J. Tavakkoli, D. Cathignol, R. Souchon, and O. Sapozhnikov, "Modeling of pulsed finite-amplitude focused sound beams in time domain," J. Acoust. Soc. Am. **104**, 2061–2072 (1998).
[31]P. Christopher and K. Parker, "New approaches to nonlinear diffractive field propagation," J. Acoust. Soc. Am. **90**, 488–499 (1991).
[32]S. Ginter, M. Liebler, E. Steiger, T. Dreyer, and R. Riedlinger, "Full-wave modeling of therapeutic ultrasound: Nonlinear ultrasound propagation in ideal fluids," J. Acoust. Soc. Am. **111**, 2049–2059 (2002).
[33]C. Lafon, V. Zderic, M. Noble, J. Yuen, P. Kaczkowski, O. Sapozhnikov, F. Chavrier, L. Crum, and S. Vaezy, "Gel phantom for use in high-intensity focused ultrasound dosimetry," Ultrasound Med. Biol. **31**, 1383–1389 (2005).
[34]W. Chen, C. Lafon, T. Matula, S. Vaezy, and L. Crum, "Mechanisms of lesion formation in high intensity focused ultrasound therapy," Proc.-IEEE Ultrason. Symp. **2**, 8–11 (2002).
[35]C. Thomas, C. Farny, C. Coussios, R. Roy, and R. Holt, "Dynamics and control of cavitation during high-intensity focused ultrasound application," ARLO **6**, 182–187 (2005).
[36]M. O'Donnell, E. Jaynes, and J. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[37]K. Waters, M. Hughes, J. Mobley, G. Brandenburger, and J. Miller, "On the applicability of Kramers–Krönig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[38]D. Cathignol, O. Sapozhnikov, and Y. Theillere, "Comparison of acoustic fields radiated from piezoceramic and piezocomposite focused radiators," J. Acoust. Soc. Am. **105**, 2612–2617 (1999).
[39]O. Sapozhnikov, Y. Pishchal'nikov, and A. Morozov, "Reconstruction of the normal velocity distribution on the surface of an ultrasonic transducer from the acoustic pressure measured on a reference surface," Acoust.

Phys. **49**, 416–424 (2003).
[40]A. Kaiser, C. Cain, E. Hwang, J. Fowlkes, and R. Jeffers, "A cost effective degassing system for use in ultrasonic measurements: The multiple pinhole degassing system," J. Acoust. Soc. Am. **99**, 3857–3860 (1995).
[41]J. Staudenraus and W. Eisenmenger, "Fibre-optic probe hydrophone for ultrasonic and shock-wave measurements in water," Ultrasonics **31**, 267–273 (1993).
[42]P. Bridgman, "The pressure-volume-temperature relations of the liquid, and the phase diagram of heavy water," J. Chem. Phys. **3**, 597–605 (1935).
[43]J. Macdonald, "Some simple isothermal equations of state," Rev. Mod. Phys. **38**, 669–679 (1966).
[44]Z. Wang, P. Lauxmann, C. Wurster, M. Kohler, B. Gompf, and W. Eisenmenger, "Impulse response of a fiber optic probe hydrophone determined with shock waves in water," J. Appl. Phys. **85**, 2514–2516 (1999).
[45]D. Cathignol, O. Sapozhnikov, and J. Zhang, "Lamb waves in piezoelectric focused radiator as a reason for discrepancy between O'Neil's formula and experiment," J. Acoust. Soc. Am. **101**, 1286–1297 (1997).
[46]B. Zeqiri and A. Bond, "The influence of waveform distortion on hydrophone spatial-averaging corrections-theory and measurement," J. Acoust. Soc. Am. **92**, 1809–1821 (1992).
[47]R. Smith, "Are hydrophones of diameter 0.5 mm small enough to characterise diagnostic ultrasound equipment?," Phys. Med. Biol. **34**, 1593–1607 (1989).
[48]E. Radulescu, P. Lewin, A. Goldstein, and A. Nowicki, "Hydrophone spatial averaging corrections from 1 to 40 MHz," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1575–1580 (2001).
[49]A. Maxwell, O. Sapozhnikov, and M. Bailey, "A new PVDF membrane hydrophone for measurement of medical shock waves," Proc.-IEEE Ultrason. Symp. 1608–1611 (2006).
[50]"IEC 61846, Ultrasonics–pressure pulse lithotripters–characteristics of fields, International Standard, International Electrotechnical Commission, 1998."
[51]V. Khokhlova, M. Canney, M. Bailey, and L. Crum, "Efficient heating and localized millisecond boiling in tissue phantoms by high intensity focused ultrasound due to formation of shocks," International Congress on Acoustics (2007).
[52]O. Sapozhnikov, Y. Pishchal'nikov, A. Maxwell, and M. Bailey, "Calibration of PVDF hydrophones using a broad-focus electromagnetic lithotripter," Proc.-IEEE Ultrason. Symp. 112–115 (2007).
[53]S. Maruvada, G. Harris, B. Herman, and R. King, "Acoustic power calibration of high-intensity focused ultrasound transducers using a radiation force technique," J. Acoust. Soc. Am. **121**, 1434–1439 (2007).
[54]K. Beissner, "Radiation force calculations," Acustica **62**, 255–263 (1987).

Erratum: "Continuous thermoacoustic mixture separation" [J. Acoust. Soc. Am., Vol. 120, pp. 2648–2657 (2006)]

G. W. Swift and D. A. Geller
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, NM 87545

(Received 19 June 2008; accepted 20 June 2008)
[DOI: 10.1121/1.2967473]

PACS number(s): 43.35.Ud, 43.20.Mv, 43.35.Ty

A typographical error appears in Eq. (A12). In the second line, $(\sigma-1)$ should be $(1-\sigma)$. (The correct equation was used for all the plots in the paper.)

 0001-4966/2008/124(4)/2421/1/$23.00

Session 1aAA

Architectural Acoustics: Acoustics of Small, Multipurpose Performance Spaces

Michael R. Yantis, Cochair
Sparling, 720 Olive Way, Ste. 1400, Seattle, WA 98101-1833

Bill Dohn, Cochair
630 Quintana Rd., Morro Bay, CA 93442

Chair's Introduction—8:30

Invited Papers

8:35

1aAA1. Translating acoustical concepts from large to small multipurpose performance spaces. Richard Silva and William Chu (McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Ste. 25, Westlake Village, CA 91362, rsilva@mchinc.com)

The design of small or large multipurpose performance spaces involves similar subjective listening parameters when considering the same types of acoustical programs. Such design includes sound absorptive material selection for reverberation control, and wall and reflector shaping for sound reflection or diffusion. However, simply scaling the same physical features does not necessarily achieve the same acoustical design goals. Drawing from a few case studies, this paper explores several design principles for small performance spaces and compares their acoustical effectiveness to their larger performance space counterparts.

8:55

1aAA2. Balancing reflection strength and loudness in small concert halls. Tim Gulsrud (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302, tgulsrud@kirkegaard.com)

Small concert halls, such as those encountered at many music education institutions, are frequently called upon to support a wide range of musical ensembles, including full symphony orchestra, within a volume significantly smaller than in traditional concert halls. On one hand, a degree of early reflection strength and reverberation is needed to support unamplified instruments and musician hearing on stage. On the other hand, constrained cubic volumes, limited audience seating, and sound-reflective finishes can lead to overloudness in these small halls. This paper will discuss recent listening and measurement experiences in small concert halls as they relate to achieving a suitable balance between reflection strength and loudness control.

9:15

1aAA3. Granoff Music Center: A multipurpose facility. K. Anthony Hoover (McKay Conant Hoover Inc., 5655 Lindero Canyon Rd., Westlake Village, CA 91362) and Matthew Moore (Cavanaugh Tocci Assoc. Inc., 327 F Boston Post Rd., Sudbury, MA 01776)

The Granoff Music Center at Tufts University is distinguished by its emphasis on classical, traditional, and world music. Completed in February 2007, it finally collects the entire Music Department under one roof after a century of inadequate office, rehearsal, and performance spaces that were scattered throughout the campus. Its intent was to house rehearsal and performance spaces whose acoustical environments are each distinctive so that the entire complex becomes multipurpose. Furthermore, the design essentially dismissed electroacoustic reinforcement or enhancement in favor of ''natural'' acoustics. This paper reviews the history, design, value engineering, and results of the facility, and the emergence and rapid escalation during construction of a multimedia program including multichannel recording, sound system, and A/V system features.

9:35

1aAA4. Acoustic design in small performance venues: "Unplugged" amplification, spatial presentation of dance, unamplified loudness control. Scott Pfeiffer (53 West Jackson Blvd. Ste. 1734, Chicago, IL 60604, spfeiffer@thresholdacoustics.com)

Facilities for the 150-seat S.N. Shure Theatre and 400-seat MUNTU Dance Theatre and for a small family of 500-seat high school auditoria all employ techniques particular to the venue size. At Shure, an emphasis on an "unplugged" sound led to a room design that is intimate and controlled. Architectural features that would be cost prohibitive on a larger scale are implemented gracefully here. The MUNTU Dance Theatre is small enough to allow for the audience to be convincingly provided with directional realism. Area microphones on stage directly feed left, center, and right clusters to allow the acoustic sounds made by the dancers to be naturally "panned" to their actual location on stage, allowing for a seamless placement of the performer. Lastly, "one-room" design approaches for schools at Nazareth Academy, Marengo High School, and Benet Academy allowed for a basic theatrical function while at the same time linking the stage volume to the house. Absorption on the tops of the ceiling system and in the materials selected for basic finishes controls loudness for circumstances where ensembles are large when compared to the audience size. Variations in approach to these similar-sized venues are discussed.

9:55—10:10 Break

10:10

1aAA5. Case studies of variable acoustic design: A repertory theater and performing arts chapel. David Conant (McKay Conant Hoover Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362)

Two multipurpose halls required variable acoustic design with minimal alteration of their appearance. The 400-seat Performing Arts Chapel at Fuller Theological Seminary (awaiting construction) will accommodate everything from pipe organ concerts to lectures-in-the-round by employing multiple variable acoustics methods with adjustable treatments deployed above the visual ceiling and along sidewalls and the performance platform. The 550-seat Virginia Piper Theater at the Mesa Arts Center (recently opened) successfully hosts everything from drama to choral works via novel, inexpensive, and wonderfully simple means, entirely beyond audience view.

10:30

1aAA6. Who cares if small high school auditoria sound any "good" for music? Bill Dohn (630 Quintana Rd., 312, Morro Bay, CA 93442, bill.dohn@gte.net)

High school music instructors, high school music students, and acoustical consultants certainly care if their auditoria sound good for music, but not all high school districts or architects appear to be all that interested. Examples of some recently completed small high school (multipurpose) auditoria are presented. Some of these have not proven particularly satisfactory for music, and (thankfully) some have. Basic programming and design criteria to achieve satisfactory music performance and listening environments in small mulitpurpose auditoria are suggested.

10:50

1aAA7. Two new terms for the comparison between acoustical measurements, calculations (ray or image models), and subjective impressions. Jason R. Duty, Hari V. Savitala, and David R. Schwind (Charles M. Salter Assoc., Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104)

The sound energy growth (SEG) curve is a new parameter depicting the growth in sound energy during the first 200–300 ms. An instantaneous sound envelope (ISE) is a new parameter yielding the sound amplitude and arrival time of individual reflections. ISE is especially useful for comparison with calculated reflectograms generated in computer models. ISE shows the contribution of individual reflections to the SEG. The two new terms provide a better comparison between measurements (ISE), calculations (from ray or image computer models), and subjective impressions (SEG) based on energy parameters.

Contributed Papers

11:10

1aAA8. Case study of measurement results for medium-sized multipurpose halls. Hari V. Savitala and Jason R. Duty (Charles M. Salter Assoc., Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104)

This case study focuses on presenting measurement results for medium-sized multipurpose halls (400–600 seats). Objective acoustical parameters, such as Reverberation Time (T20) and Clarity (C80), were correlated to subjective impressions of the spaces by the end users. The results are presented to better understand how to design medium-sized multipurpose halls.

11:25

1aAA9. Audibility of absorption, diffusion, or a specular reflector on the first lateral reflection. Benjamin Bridgewater (7383 Stagecoach Trail Rd., Manhattan, MT 59741, benbridgewater@gmail.com)

To better understand the importance of treatment on the first reflection in a small room, a study of the audibility of the treatment of the first lateral reflections in a small room was conducted. Subjects completed the ABX tests comparing the different treatments of the first lateral reflection to determine if there was a noticeable difference between the treatments of the first lateral reflection.

Session 1aMU

Musical Acoustics: Dynamical Approaches in the Study of Music Perception and Performance I

Edward W. Large, Chair
Florida Atlantic Univ., Ctr. for Complex Systems, 777 Glades Rd., Boca Raton, FL 33431

Chair's Introduction—8:30

Invited Papers

8:35

1aMU1. A dynamic approach to attending: New issues. Mari Riess Jones (Dept. of Psych., Univ. of California, Santa Barbara, 5388 Traci Dr., Santa Barbara, CA 93111, jones.80@osu.edu)

Musical events entail crafted serial changes in frequency, intensity, and time that are explicitly patterned to engage attending activities over relatively long periods of time (i.e., versus brief time periods required for single tones). Serial changes vary in salience, with some functioning as accents outlining higher-order time spans between nonadjacent tones. Accordingly, attending to music must be dynamic and sustained, capable of anticipatory behaviors while continuously reacting and adapting to various serial changes, accented and unaccented. Issues of meter perception and melodic phrasing also mean that attending must be responsive to different time levels marked by accents. Approaches to attending that ground attentional tracking in biological entrainment have potential for describing these aspects of musical listening. They also raise new issues, as illustrated in experiments inspired by entrainment theories; as reviewed here, these address effects of rhythm on time discrimination, time judgments, and responses to melodic changes. Also considered are general issues relating to development of new paradigms, age-specific preferred rate (i.e., preferred intrinsic periods), and synchronization limits.

9:05

1aMU2. Perspectives on perceptual timing. J. Devin McAuley (Dept. of Psych., Bowling Green State Univ., Bowling Green, OH 43403, mcauley@bgsu.edu)

There has been a notable increase in research on the neural bases of timing in recent years. A focus of much of this work has been on identifying the locus and functional characteristics of a putative central clock mechanism under the assumptions that (1) both human and nonhuman animal timings involve similar pacemaker-accumulator mechanisms and (2) the functional characteristics of the internal clock are the same whether timing an isolated duration, such as a stoplight, or dancing to music, even though the latter requires perceiving a periodic beat, while the former does not. This talk presents research from an alternative dynamic perspective that emphasizes the importance of assessing (1) the timing of extended temporal sequences characteristic of music rather than isolated durations and (2) individual differences. [Work supported by the GRAMMY Foundation.]

9:35

1aMU3. Neural dynamics of beat perception and production. John Iversen (The Neurosci. Inst., 10640 John J Hopkins Dr., San Diego, CA 92121)

The perceptual experience of a rhythm depends on its metrical interpretation. Prior work using magnetoencephalography has shown that brain activity reflects a listener's metrical interpretation of simple ambiguous rhythms: neural response to a note coinciding with the listener's internal sense of the beat is enhanced in the beta frequency range (20–30 Hz). An unanswered question raised by this result is where in the brain the transient periodic enhancement of evoked response to the beat might arise. Ongoing research on brain responses to syncopated rhythmic patterns is presented. In these syncopated patterns, some beats are never marked by physical tones. It is asked if the increase in transient beta activity observed in prior experiments can also indicate internal beat placement even in the predictable absence of a physical input. If so, it may reveal the nature of the internal modulation of cortical responses to sound and enable localization of its source. The presentation will conclude with a discussion of what brain architectures may be able to support beat perception and synchronization, with reference to nonhuman animals. [Work supported by Neurosciences Research Foundation.]

10:05—10:30 Break

10:30

1aMU4. Time courses of cortical beta and gamma-band activity during listening to metronome sounds in different tempi. Takako Fujioka (Rotman Res. Inst., Baycrest, Univ. of Toronto, 3560 Bathurst St., Toronto, ON M6A 2E1 Canada, tfujioka@rotman-baycrest.on.ca), Edward Large (Florida Atlantic Univ., Boca Raton, FL 33486), Laurel Trainor (McMaster Univ., Hamilton, ON L8S 4K1, Canada), and Bernhard Ross (Univ. of Toronto, Toronto, ON M6A 2E1, Canada)

Oscillatory cortical activities in beta-band (13–20 Hz) are related to a somatomotor system, and gamma-bands (20 Hz) are involved with feature binding in perception. Previously gamma-band activity in electroencephalography was found to modulate with musical pulse with a two-beat metric accent. The present study examined beta- and gamma-band activities in auditory cortices recorded via magnetoencephalography (MEG) when subjects listened to musical pulse in various tempi. Tones were presented with intervals of (1) 390 ms, (2) 585 ms, and (3) 780 ms, and (4) with irregular intervals between 390 and 780 ms. In addition, the same tones were presented with a 390-ms interval using a two-beat accent, while occasionally either (5) an accented tone or (6) an unaccented tone was omitted. Beta-band activity decreased immediately after the stimulus and returned to the previous level just before the next stimulus, regardless of the tempo, except in the irregular condition. The tone omission resulted in an extra beta rebound. Gamma-band activity increased right after the pulse or the omission. We propose that beta oscillations may encode the timing of the next sound in a regular pulse sequence, whereas gamma oscillations likely reflect processing of the current auditory events including omissions.

11:00

1aMU5. Tempo fluctuation and predictive synchrony in music. Summer Rankin, Edward Large, and Philip Fink (Ctr. for Complex Systems and Brain Sci., Florida Atlantic Univ., Boca Raton, FL 33431, rankin@ccs.fau.edu)

The goal of this study was to understand the ability of musicians and nonmusicians to entrain to musical performances with naturally fluctuating tempo. In experiment 1, we investigated the nature of tempo fluctuations produced by a musician. We collected four music performances from a skilled pianist who was instructed to play with natural expression. We extrapolated the beat times by comparing the performances with the musical scores and performed a spectral analysis and a rescaled range analysis on the interbeat intervals. The results indicate fractal scaling of the performance tempo in each case. Thus, musical tempo fluctuations exhibit long-range correlations characterized by a $1/f$ type long memory process. The stimuli in experiment 2 included two performances from experiment 1 and mechanical versions of the same pieces. Participants were asked to synchronize at slower (quarter note) and faster (eighth note) metrical levels. Musicians and nonmusicians synchronized successfully and were able to accommodate large tempo fluctuations. Entrainment was superior at the slower metrical level. Participants were less variable for the mechanical versions. Most importantly, participants predicted the tempo fluctuations observed in experiment 1, suggesting a possible relationship between fractal tempo scaling, pulse synchronization, and neural oscillation in the perception of rhythm. [Work supported by NSF grant BCS-0094229.]

11:30

1aMU6. Dynamics of emotional communication in performed music. Heather Chapin, Edward Large (Ctr. for Complex Systems, Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431, chapin@ccs.fau.edu), Kelly Jantzen (Western Washington Univ., Bellingham, WA 98225), J. A. S. Kelso (Florida Atlantic Univ., Boca Raton, FL 33431), and Fred Steinberg (Univ. MRI of Boca Raton, Boca Raton, FL 33431)

Musicians vary performance parameters, such as tempo and sound intensity, to express emotion dynamically. Our goals were to link specific parameters of performance expression with listeners' emotional responses and neural activity. Ten musically trained and ten untrained participants listened to two versions of a romantic piano composition. The expressive performance included variations in timing and sound intensity. The mechanical performance maintained constant tempo and sound intensity throughout. Participants reported emotional responses in a two-dimensional response space (emotional intensity and valence) before and after fMRI scanning. In the fMRI scanner, participants listened to both versions without reporting emotional responses. Ratings of emotional intensity were positively correlated with tempo. The fMRI analysis revealed a main effect of performance, showing increased BOLD responses for the expressive performance in an emotion-related network. A main effect of training showed increased BOLD responses for trained participants in areas associated with emotion processing and reward. We also observed BOLD responses that were temporally correlated with tempo fluctuations. In summary, performance expression predicted listeners' dynamic ratings of emotional intensity. Moreover, limbic areas responded to expressive versus mechanical performances, and dynamic BOLD changes tracked performance expression and reported emotional intensity over time. [Work supported by NSF BCS-0094229.]

Session 1aNS

Noise and Animal Bioacoustics: Advances in Measurement and Noise and Noise Effects on Humans and Non-Human Animals in the Environment I

Anne E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Fluid Mechanics and Engineering, Einsteinufer 25, 10587 Berlin, Germany

Chair's Introduction—8:30

Invited Papers

8:35

1aNS1. New approaches in soundscapes: About the triangulation of measurements. Brigitte Schulte-Fortkamp (Inst. of Fluid Mech. and Eng. Acoust., TU-Berlin, Einsteinufer 25, D-10587 Berlin, Germany)

A procedure is developed following the soundscape approach, which will provide acoustical indicators and parameters describing urban and other outdoor living areas with respect to physical conditions and their relevance for life. In particular, the contribution of important sources like traffic noise to the overall sound exposure and its influence on the evaluations by humans has to be determined. Moreover, the question is to which degree does a single source determine the soundscape of the environment with respect to the perception and evaluation? Qualitative evaluations have been conducted introducing the concept of new experts with respect to data triangulation.

8:55

1aNS2. Spatial variation in marine mammal habitat soundscapes within one coastal region. Jennifer L. Miksis-Olds (Appl. Res. Lab., The Penn. State Univ., P.O. Box 30, State College, PA 16803, jlm91@psu.edu) and Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105)

Manatees and dolphins inhabit coastal waterways in which they are constantly exposed to sound produced by physical processes, biological activity, and human activities. Ambient sound levels can change rapidly on the scale of minutes to hours, but variability in sound level is the only one component of the acoustic environment that has the potential to impact marine mammals. Soundscapes, or acoustic footprints, of a region provide a visual snapshot of the acoustic components contributing to the ambient sound level. Acoustic footprints provide information on the identity of sound sources as well as sound levels. A passive aquatic listening instrument was deployed in 24 sites around Sarasota Bay, FL. Recordings were made in 24 sites frequented by manatees and dolphins. Generation of acoustic footprints from each site illustrates the temporal and spatial variabilities of source components within and between the habitat types. Vessel noise is a more dominant sound source in sites used less frequently by manatees compared to more frequently used sites. Sites used most heavily by manatees tended to be dominated by biological activity or wind.

9:15

1aNS3. Underwater acoustic scene analysis: Exploration of appropriate metrics. William T. Ellison, Adam S. Frankel, David Zeddies, Kathleen J. Vigness Raposa, and Cheryl Schroeder (Marine Acoustics, Inc., 809 Aquidneck Ave., Middletown, RI 02842)

A recent review of published accounts of acoustic noise exposure and behavioral response found that the reported results were often limited by the lack of supporting information on basic acoustical metrics as well as potentially important contextual variables. [Southall *et al.*, Aquat. Mamm. **33**, 411–521(2007)]. Human studies have also shown that simple measures such as sound level are insufficient to represent complex acoustic scenes [Dubois, J. Acoust. Soc. Am. **122**, 3011–(2007)]. While source and received levels are frequently reported for underwater studies, other characteristics of the resultant acoustic field are rarely described. Understanding the full spatial, temporal, and spectral nature of the sound field can provide additional quantitative measures useful in elucidating the animals' response to the incident sound field exposure. The animals' perception of received acoustic field may also be influenced by the biological meaningfulness of the signal and the experience of the receiver with signals of a given type. Here we provide guidelines for future studies corresponding to three phases: preplanning, field measurements, and postmeasurement analysis, addressing acoustic field metrics as well as the seasonal distribution and baseline behavior of the animals of interest. Examples from selected field measurements will be used to illustrate the salient features.

9:35

1aNS4. Is it time to seriously consider the effects of acoustic particle motion on hearing and behavior in fish? Mardi Hastings (Penn State Univ., Appl. Res. Lab, P.O. Box 30, State College, PA 16804, mch26@psu.edu)

Although sound exposure level (SEL) and sound pressure level (SPL) are the metrics being applied to assess effects of underwater sound on fish, no correlation exists between SEL or SPL and temporary threshold shift (TTS) or behavior. Certainly particle motion plays a major role in the auditory scene of fishes because their inner ears are literally accelerometers that detect signals via relative motion between sensory epithelia and solid otoliths in the auditory endorgans. Most all hearing data, however, continue to be measured with respect to SPL, and TTS with respect to SEL or SPL. Neither metric accounts for effects of particle motion. An analysis of findings from several recent acoustic impact studies on various species using different types of sources indicates that differences in observed effects are most likely due to differences in received levels of acoustic particle motion. Field studies with real sources in open water predict an onset of TTS at higher SEL than do studies conducted in confined spaces or with specimens located in the source near a field where acoustic particle motion is higher than estimated by simple pressure–particle velocity relationships. Thus monitoring only sound pressure may not be sufficient to protect fish from adverse impacts.

9:55—10:15 Break

10:15

1aNS5. Results of giving a bottlenose dolphin control over exposure to a net alarm: Implications for the measurement of aversion and detection thresholds. Ann Bowles (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109, abowles@hswri.com) and Rindy Anderson (Duke Univ., Durham, NC 27708-0338)

When exposed to pings from a Dukane Netcom 1000 net alarm, bottlenose dolphins responded with avoidance and aggression. These species-typical behaviors were tested as a tool for measuring the threshold for aversion. Two bottlenose dolphins were exposed to pings with varying signal-to-noise ratios (SNRs of 5–70 dB) in randomized order. Trials were terminated when the dolphins exhibited any agonistic behavior. Sixty-nine animal trials were terminated by the avoidance or aggressive approach. Seventy-seven trials were terminated by fluke-slapping, for which latency could be measured precisely. At SNR of 10 dB or below, the dolphin did not fluke-slap. Above 10 dB SNR, the probability of response increased and latency to slap decreased rapidly. The shape of the latency-SNR curve was similar to equal latency curves measured in audiometric studies using operant conditioning, suggesting that control over pings was intrinsically rewarding. Cessation of pings may have been the reward, but because the dolphin chose to terminate stimuli at levels just above the limit of detection, another possibility should be considered: control over the stimulus could reinforce itself. If so, conditioning experiments could be designed for audiometric research on wild cetaceans [Work supported by NMML and SeaWorld San Diego; NMFS Permit 1016.]

10:35

1aNS6. The impacts of energy development noise playback on lek attendance and behavior in greater sagegrouse. Jessica Blickley (Dept. of Evolution and Ecology, Univ. of California, Davis, One Shields Ave., Davis, CA 95616, jlblickley@ucdavis.edu), Diane Blackwood (Baywing Assoc., Williamston, MI 48895), and Gail Patricelli (Univ. of California, Davis, Davis, CA 95616)

The spread of human development has caused a dramatic increase in noise across the landscape. Recent studies suggest that noise may negatively impact wildlife, yet little is known about the causes and consequences of this impact. Furthermore, most previous studies have not been designed to isolate noise impacts from other confounding factors. This study is investigating the impacts of energy development noise on greater sage grouse (*Centrocercus urophasianus*), a species of management concern across western North America. Sage grouse are declining in areas of energy development and circumstantial evidence suggests that noise is a cause of this decline. To test this hypothesis, control leks and leks with experimentally-introduced energy development noise were monitored. Results from three seasons of experimental noise playback will be presented, comparing lek attendance, territory selection, and display behavior between noise and control leks. This is the first long-term playback experiment investigating the chronic impacts of noise on any wild population, offering a unique opportunity to experimentally address noise impacts on avian behavior and breeding site selection while informing conservation efforts for this species.

10:55

1aNS7. The effects of gas well compressor noise on nesting birds in pinyon-juniper woodlands. Clinton D. Francis (Dept. of Ecology and Evolutionary Biology, Univ. of Colorado, UCB 334, Boulder, CO 80026, clinton.francis@colorado.edu) and Catherine P. Ortega (Fort Lewis College, Durango, CO 81301)

Anthropogenic noise has been implicated as a cause of declines in avian species diversity and densities; however, these patterns are not conclusive because noise has not been uncoupled from potential confounding variables that accompany noisy human activities. During the summers of 2005–2007, noise was isolated through the use of natural gas wells with and without noisy compressors (treatment and control sites, respectively). No difference in overall nest density on plots with and without compressors was found; however, noise significantly reduced nesting species richness. Individual species also demonstrated responses to noise: Mourning Dove nests were more abundant on control sites, whereas Black-chinned Hummingbird and House Finch nests were more abundant on treatment sites. Several species were also found to nest significantly farther from well pads on treatment sites than on control sites. Nest success was higher on treatment sites than on control sites, due to lower levels of predation. In areas influenced by anthropogenic noise, predators may be present, but unable to locate nests, or they are in lower densities or absent. Anthropogenic noise appears not only to influence avian nesting patterns, thereby altering the avian community, but also to influence nest success through changes to predation patterns.

Contributed Paper

11:15

1aNS8. Effects of varying levels of noise and vibration on behavioral and physiological stress responses in two species of freshwater aquarium fish: Fathead minnows (*Pimephales promelas*) and bluegill sunfish (*Lepomis macrochirus*). Alexandra Rose, Jeanette Thomas (Biological Sci. Dept., Western Illinois Univ.-Quad Cities, Moline, IL 61265), and Allen LaPointe (John G. Shedd Aquarium, Chicago, IL 60614)

Aquariums provide beautiful underwater environments for an array of fish species, but to maintain optimum water quality and aesthetically pleasing displays, a variety of mechanical systems that unintentionally produce noise, including pumps, filters, skimmers, bubblers, etc., are necessary. The objectives of this study are to measure the acoustic properties of noise in an aquarium and examine the effects of different noise levels on the stress experienced by two species of freshwater fishes: fathead minnows (*Pimephales promelas*), hearing specialists, and bluegill sunfish (*Lepomis macrochirus*), hearing generalists. We conducted a noise mapping of several fish galleries at the Shedd Aquarium using a BK 2560 SPL meter; Leq, SEL, and minimum & maximum decibel levels were measured up to 2 kHz using a DolphinEar hydrophone. We then tested the effects of four noise levels on the stress experienced by these two fish species by measuring oxygen consumption/respiration rates with Loligo respirometry chambers and cortisol release rate with a noninvasive radioimmunoassay that measures free cortisol in water samples [Ellis *et al.* 2004]. These measurements will help to evaluate the effects of noise in standard aquarium settings on fish welfare and may aid in the determination and regulation of acceptable anthropogenic noise levels in aquariums.

11:30—1:00
Session Luncheon and Moderated Discussion

MONDAY MORNING, 10 NOVEMBER 2008 LEGENDS 7, 9:00 A.M. TO 12:00 NOON

Session 1aSC

Speech Communication: Perception of Consonants and Vowels (Poster Session)

Catherine L. Rogers, Chair
Univ. of South Florida, Comm. Sci. and Disorders, 4202 E. Fowler Ave., Tampa, FL 33620-8150

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

1aSC1. Duration and spectral shape as cues to vowel identity. Webster Tilton, IV and James R. Sawusch (Psych. Dept., State Univ. of New York at Buffalo, 206 Park Hall, Buffalo, NY 14260)

Theories of vowel recognition based on the vowel formant frequencies (or spectral peaks) are less accurate in identifying vowels than human listeners. This has been true even if the model uses dynamic spectral information such as the changes in the spectral peaks over time [H. Houde (2002)]. Studies with synthetic speech and edited natural speech have implicated duration and fundamental frequency as additional perceptual cues. The present work focused on duration, fundamental frequency, and spectral shape using synthetic vowels modeled on the speech of a male and a female talker. Series from /i/ to /I/, /E/ to /ae/, /uh/ to /a/, and /U/ to /u/ were created in which only duration, fundamental frequency, or the shape of the short-term spectrum were varied. Subsequent studies manipulated the formant frequencies and the duration or spectral shape orthogonally. Results indicated that the duration and spectral shape of a steady state vowel exert a potent influence on vowel identification, while varying the fundamental frequency appeared to have little effect.

1aSC2. Identification of transition-only and steady-state vowels by young and older normal-hearing listeners. Elizabeth Talmage, Gail Donaldson, and Catherine Rogers (Dept. of Comm. Sci. and Disord., Univ. of South Florida, PCD 1017, 4202 E. Fowler Ave., Tampa, FL, gdonalds @cas.usf.edu)

Normal-hearing listeners can identify vowels on the basis of either dynamic or steady-state (SS) cues. To determine which cues are more effective, vowel identification was measured for full, transition-only (TN) and SS versions of naturally produced exemplars of the syllables "beeb, bib, babe, beb, bab, and bob." TN stimuli retained 10, 20, 30, 40, 60, or 80 ms of the consonant-vowel and vowel-consonant transitions and were neutralized in overall duration. SS stimuli retained 10, 20, 30, 40, 60, or 80 ms of the vowel center but eliminated transitions. Young normal-hearing (YNH) and older normal-hearing (ONH) listeners were assessed. Performance declined as the duration of acoustic information decreased. On average, decrements were similar for TN and SS stimuli except at the shortest duration (TN10 and SS20), where average performance was poorer for the TN stimulus. However, relative performance for short-duration TN and SS stimuli varied substantially across vowels. TN and SS stimuli produced similar performance for some vowels, but TN stimuli produced much poorer performance than SS stimuli for other vowels. Preliminary findings indicate a trend for slightly lower performance among ONH listeners relative to YNH listeners but suggest similar patterns of performance across groups. [Work supported by NIH-NIDCD 5R03DC005561.]

1aSC3. Perceptual accommodation to sinewave speech. James M. Hillenbrand (Dept. of Speech Pathol. Audiol., Western Michigan Univ., Kalamazoo, MI 49008, james.hillenbrand@wmich.edu), Michael J. Clark (Western Michigan Univ., Kalamazoo, MI 49008), Robert A. Houde (Ctr. for Commu Res., Rochester, NY 14623), Michael W. Hillenbrand (Western Michigan Univ., Kalamazoo, MI 49008), and Kathryn S. Hillenbrand (Vicksburg High School, Vicksburg MI 49097)

Many studies have reported good intelligibility for sine wave replicas of sentences (e.g., R. Remez *et al.*, Science **212**, 947–950 (1981)]. Recent work, however, has shown poor intelligibility (~55%) for vowels in isolated syllables [J. Hillenbrand and M. Clark, J. Acoust. Soc. Am. **123**, 3326 (2008)]. While enhanced intelligibility for sentences undoubtedly reveals the importance of top-down mechanisms, it is also possible that sentence-length utterances allow listeners to make (as yet unknown) perceptual accommodations to the unfamiliar acoustic properties of sine wave speech (SWS). In this study, the intelligibility of SWS replicas of 16 vowels/diphthongs in isolated syllables ("heed," "hid," and "hide") was compared to that of the same syllables when preceded by a seven-word SWS carrier phrase (CP) spoken

by the same talker. Intelligibility was ~24 percentage points higher when the SWS syllables were preceded by the SWS CP than when the same utterances were presented in isolation. Furthermore, the effect was observed even when the CP and test-syllable talkers did not match, showing that the effect involves more than just talker normalization. Finally, a same-talker natural speech CP preceding the SWS syllable produced a decrement rather than an improvement in intelligibility.

1aSC4. The role of speech-specific signal characteristics in vowel normalization. Matthias J. Sjerps, James M. McQueen, and Holger Mitterer (Max Planck Inst. for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, matthias.sjerps@mpi.nl)

Listeners adjust their vowel perception to the characteristics of a particular speaker. Six experiments investigated whether speech-specific signal characteristics influence the occurrence and amount of such normalization. Previous findings were replicated with first formant ($F1$) manipulations of naturally recorded speech; target sounds on a /pIt/ (low $F1$) to /pt/ (high $F1$) continuum were more often labeled as /pIt/ after a precursor sentence with a high $F1$, and more often labeled as /pt/ after one with a low $F1$ (Experiment 1). Normalization was also observed, though to a lesser extent, when these materials were spectrally rotated, and hence sounded unlike speech (Experiment 2). No normalization occurred when, in addition to spectral rotation, the silent intervals and pitch movement were removed and the syllables were temporally reversed (Experiment 3), despite spectral similarity of these precursors to those in Experiment 2. Reintroducing only pitch movement (Experiment 4), or silent intervals (Experiment 5), or spectrally rotating the stimuli back (Experiment 6) did not result in normalization, so none of these factors alone accounts for the effect's disappearance in Experiment 3. These results show that normalization is not specific to speech but still depends on more than the overall spectral properties of the preceding acoustic context.

1aSC5. Measuring the impact of native inventory and native contrast on vowel perception. Marc Ettlinger and Keith Johnson (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720)

Theories of speech perception differ as to whether experience with sounds or with sound contrasts is more important when discriminating novel sounds. To explore this question, 21 American English, 16 Turkish, and 16 French speakers were tested on their ability to discriminate high front vowels in German (/i/,/y/,/Y/,/I/). The crucial difference between the languages is that English only uses the tense-lax contrast (/i/-/I/), while Turkish and French only use the rounding contrast (/i/-/y/). If contrast is crucial, then English speakers should be better at discriminating /y/-/Y/, whereas if the inventory is crucial, then they should be better at discriminating /I/-/Y/. The results of a fixed discrimination task show that there is no significant effect of language and that the tense-lax contrast (/i/-/I/; /y/-/Y/) is universally more easily discriminated than the rounding contrast (/i/-/y/; /I/-/Y/). The results of a rating task do reflect an influence of language, however, with English-listeners rating /I/-/Y/ as more distinct and French and Turkish speakers rating /y/-/Y/ as more distinct. So, because English has /I/, but no rounding contrast, what matters in the perception of new sounds is not experience with contrasts, but rather experience with particular sounds.

1aSC6. Perceptual compensation for /u/-fronting in American English. Reiko Kataoka (Dept. of Linguist., Univ. of California at Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, kataoka@berkeley.edu)

Listener's identification of speech sounds are influenced by both perceived and expected characteristics of surrounding sounds. For example, Ohala and Feder (1994) demonstrated that American listeners judge a vowel stimulus which is ambiguous between /i/ and /u/ more frequently as /u/ in alveolar context than in bilabial context, and do so both in cognitively "restored" contexts as well as acoustic contexts. This paper reports the results of three-part perception experiments with 31 native speakers of American English, aiming to confirm and extend these results. Experiment 1 replicated the findings of Ohala and Feder. Furthermore, it shows that reaction time (RT) for /u/ judgment is shorter in alveolar than in bilabial context. Experiment 2 showed that perceptual compensation became greater as the speech rate of the precursor sentence increased. The results from experiments 1 and 2 might indicate that listeners use both cognitively based categorical compensation and mechanically based gradient compensation. Experiment 3 investigated the role of speech production in perceptual compensation. Moderate correlation between the degree of /u/-fronting in production and the perceptual boundary of /i/-/u/ categories was obtained, suggesting a link between speech production and speech perception.

1aSC7. Dialectal differences in dynamic formant patterns in vowels. Robert Allen Fox and Ewa Jacewicz (Speech Percept. and Acoust. Labs, Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, fox.2@osu.edu)

This study examines basic acoustic variation in the dynamic patterns of vowel formants among three regional variants of American English spoken in southeastern Wisconsin (affected by the Northern Cities Shift), western North Carolina (Appalachian English affected by the Southern Vowel Shift), and central Ohio (not considered to be affected currently by any vowel shift). Three groups of speakers (including men and women) produced vowels in citation form (in a /hVd/ context) and in sentences. The sentence material elicited two degrees of vowel emphasis (high and low) in the /bVd/ and /bVt/ contexts. The frequencies and amplitudes of the first three formants were extracted at points corresponding to 20%-35%-50%-65%-80% of the vowel's duration. A set of dynamic measures was then calculated using these values that included overall signed/unsigned change in formant frequencies, vector and trajectory lengths, direction/angle of vowel movement in the $F1$ by $F2$ plane, and rate of frequency change. The results show significant cross-dialectal differences in formant patterns in monophthongs, phonemic diphthongs, and nonphonemic diphthongs. The nature and extent of these formant changes varied as a function of speaker dialect in a manner apart from expected variation resulting from phonetic and prosodic context and speaking style. [Work supported by NIH.]

1aSC8. Production and perception of two vowels in Northeast Ohio. Anna Schmidt and Lois Powell (School of Speech Path Aud., Kent State Univ., Kent, OH 44242, aschmidt@kent.edu)

In Northeast Ohio, three dialect regions converge: the Inland North, North Midland, and Pittsburg regions (Labov *et al.* 2005). It is believed that the vowel in "bat" is raising and fronting while the vowel in "bet" is lowering and backing as part of the Northern cities vowel shift in the Inland North region. Earlier data from 50 talkers (aged 18–25) from these regions suggest vowel movement in each region with a full merger or reversal of the two vowels only for Inland North talkers. Talkers from all regions produced longer bat vowels than bet vowels. The current study examined perception as well as production of the vowels by talkers from these regions. Participants heard two synthetic vowel continua and identified the vowels: a duration continuum, in which the formant pattern was constant but the vowel duration varied, and a formant continuum, in which the duration was constant but the formant pattern varied from a low backed vowel to a high fronted vowel. Perception patterns were compared to production patterns in order to determine if (a) longer vowels were identified as bat vowels and (b) vowel identification followed vowel production.

1aSC9. Comparison of vowels in maternal speech to adults and to children with hearing loss. Laura Dilley (Dept. of Commun. Disord. and Dept. of Psych., Bowling Green State Univ., 247 Health Ctr., Bowling Green, OH 43403, dilley@bgsu.edu) and Tonya Bergeson (Indiana Univ. School of Medicine, Indianapolis, IN 46202)

Previous research has indicated that when speaking to normal-hearing children, maternal caregivers' speech is characterized by a vowel space which is expanded relative to that produced when speaking to other adults. The present study investigated the characteristics of the vowel spaces produced by mothers interacting with hearing-impaired children who had recently received a cochlear implant (CI), relative to those produced when interacting with another adult. Mothers were recorded interacting with their hearing-impaired children with CIs in quiet play sessions at 3- or 6-month interval postimplantation as well as in semistructured interviews with another adult. Measurements of $F1$ and $F2$ for the vowels /i/, /ɑ/, and /u/ in stressed syllables were taken, and the areas of the resulting vowel triangles for the two production conditions were compared. Preliminary results show that when speaking to their hearing-impaired children with CIs, vowel spaces of mothers were not expanded relative to when speaking to another adult. The results will be compared with vowel spaces produced to normal-hearing children and hearing-impaired children with hearing aids. These

findings have implications for developing intervention strategies aimed at caregivers for improving speech and language skills in children with CIs. [Work supported by NIH-NIDCD R01DC008581.]

1aSC10. The impact of caregivers' speech on infants' discrimination of a speech sound contrast. Alejandrina Cristià (Linguist., Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907, acristia@purdue.edu)

The acquisition of speech perception has aimed primarily to detect variation across linguistic populations, assuming that the linguistic input to infants within a given linguistic community is essentially the same. In this presentation, individual differences due to variability within a linguistic community are investigated using a contrast that is phonemic in the language, but which is produced with a great deal of variation across speakers. English-learning 5-month-old and 13-month-old infants were tested for their ability to discriminate English sibilants, and the variance in their performance was correlated with the acoustic characteristics of the speech of their caregivers in order to investigate some of the possible sources of individual differences in speech perception.

1aSC11. Infants' perception of non-native sibilants following different distributions of frication and vocalic cues. Alejandrina Cristià (Linguist., Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907, acristia@purdue.edu), Grant McGuire (Univ. of California at Santa Cruz, Santa Cruz, CA 95064), Amanda Seidl, and Alexander Francis (Purdue Univ., West Lafayette, IN 47907)

Previous work suggests that learning to perceive speech categories in infancy may be influenced by the distributions of acoustic cues that underlie them. However, studies on this topic have focused on distributions of single cues, especially voice onset time (VOT), whereas natural phonetic categories are typically defined according to multiple cues and, moreover, VOT may constitute a special case as nonhuman animals are known to exhibit categorical perception of this cue. In order to explore the role of cue distributions in a more natural context, and to extend this research beyond the domain of VOT, in this study we exposed infants to different distributions of cues to a non-native contrast of place of articulation in fricatives. Results suggest that young infants are indeed able to track multiple acoustic cue distributions at once, and that these distributions shape infants' perception after a brief exposure, possibly by enhancing the perceptual similarities between acoustically similar tokens.

1aSC12. Transmission and reception of visual speech signals produced by cued speech transliterators. Katherine Pelley and Jean C. Krause (Dept. of Comm. Sci. and Dis., Univ. of South Florida, 4202 E. Fowler Ave., PCD 1017, Tampa, FL 33620, jkrause@cas.usf.edu)

Although it is common practice for deaf individuals to use interpreters as a means of accessing spoken information, few investigations of interpreter skills have been reported and virtually none have focused on interpreter intelligibility [Kluwin and Stewart, Odyssey **2**, 15–17 (2001)]. In order to begin quantifying the contribution of factors affecting interpreter (transliterator) intelligibility for one English-based communication mode (cued speech), two experiments were conducted. In the first experiment, 12 cued speech transliterators were asked to transliterate materials at three different speeds (slow, normal, fast). Two characteristics of the visual speech signal produced by these transliterators were then examined: (1) accuracy, measured as a percent-correct score based on the target cue sequence, and (2) lag time, reported as the average delay between the spoken and transliterated messages. In the second experiment, eight to ten expert receivers of cued speech were presented with visual stimuli excised from the transliterated messages and asked to transcribe the stimuli. Intelligibility of the visual speech signals was measured as the percentage of words correctly received. Results show that accuracy is inversely related to lag time and accounts for roughly 25% (on average) or more (when transliterator experience is controlled) of the variance in intelligibility. [Work supported by NIH /NIDCD Grant 5-R03-DC007355.]

1aSC13. Online processing of acoustic cues used in speech perception: Comparing statistical and neural network models. Joseph Toscano (Dept. of Psych., Univ. of Iowa, Iowa City, IA 52242, joseph-toscano@uiowa.edu) and Bob McMurray (Univ. of Iowa, Iowa City, IA 52242)

Most phonological contrasts are signaled by multiple acoustic cues, yet it is unclear how these cues are combined during speech perception. Formal computational modeling offers a useful tool for studying this process. Two computational approaches are presented here. The first is a mixture of Gaussians (MOG) model that forms categories and combines cues based on their statistical distributions [Toscano and McMurray, Proceedings of the Cognitive Science Society (2008)]. The second is a neural network model that combines statistical learning and dynamic online processing [McMurray and Spivey, Proceedings of the Chicago Linguistic Society (1999)]. Both the MOG and the network use the statistical distributions of speech sounds to form categories. The MOG offers transparency in that its categories correspond directly to distributional statistics measured from phonetic data. However, it does not capture the online processing observed in behavioral experiments that suggest that the speech system makes preliminary commitments before all cues are available [McMurray, Clayards, Tanenhaus, and Aslin (submitted)]. The network offers an approach that may allow us to observe this processing. Thus, while the MOG may better clarify the relationship between acoustics and phonological categories, the network may better model the process of speech perception. [Work supported by NIH.]

1aSC14. Categorization and generalization of multiple dimensions in nonspeech stimuli. Grant McGuire (Dept. of Linguist., Univ. of California at Santa Cruz, 1156 High St., Santa Cruz, CA 95064-1077)

Nonspeech (NS) stimuli have seen considerable use for exploring linguistic perception questions. One problem with such studies is extrapolating results found using abstract sounds to specific linguistic phenomena. This paper reports a series of experiments using noise-tone pairs, which have all the advantages of NS sounds but also have sufficient similarity to speech sounds to make connections with linguistic data. These sounds were used in several perceptual learning experiments exploring categorization and generalization. Subjects were trained to categorize using a single dimension or two integrated dimensions and then asked to label different sets of novel stimuli to assess generalization and dimensional reliability. Preliminary results demonstrate that learned dimensions were preferred for categorizing novel sounds and that generalization of integrated dimensions depends strongly on the relationship between the two dimensions. These results are discussed in terms of linguistic categories and theories of perceptual learning.

1aSC15. From production to acoustics to perception: The case of fricatives. Bob McMurray (Dept. of Psych., E11 SSH, Univ. of Iowa, Iowa City, IA 52240, bob-mcmurray@uiowa.edu) and Allard Jongman (Univ. of Kansas, Lawrence, KS 66044, jongman@ku.edu)

Although the acoustics of speech mediate production and perception, few studies have examined the entire chain. English fricatives present an interesting case because they are realized by multiple cues. Previously, 20 speakers were recorded producing the eight English fricatives with six vowels. Twenty-one cues were measured [Jongman *et al.*, J. Acoust. Soc. Am. **108**, (2000)]. Fricatives were presented (with and without the vowel) to 20 listeners for identification. Listeners were better at identifying fricatives in context (M=91%) than isolation (M=75%), and sibilants (M=94%) than nonsibilants (M=72%). Acoustic measurements were analyzed with a regression model of cue-parsing [McMurray *et al.* (unpublished)]. The model approximated listeners' performance in context (M=87%). However, this was only true when speaker-specific variance could be parsed out; significant speaker variation was observed in every cue. Thus, vowel context may aid perception by helping listeners identify the speaker. While parsing speaker and vowel from the cues was helpful, this was not true for voicing. Voicing and place cues overlapped, but knowing the voicing did not improve identification of place (and vice versa). Finally, when cue weightings based

on perception and articulation were computed independently, subtle differences emerged, suggesting that perception is similar, but not isomorphic, to production

1aSC16. Perceptual sensitivity to within-category acoustic variation by language impaired and typically developing adolescents. Cheyenne Munson, W. Dan McEchron, Bob McMurray (Dept. of Psych., Univ. of Iowa, E11 SSH, Iowa City, IA 52242, cheyenne-munson@uiowa.edu), and J. Bruce Tomblin (Univ. of Iowa, Iowa City, IA 52242)

Perceptual deficits have been suggested as an underlying cause of language impairment, and language impaired (LI) listeners sometimes show less categorical perception than normal listeners (Tallal & Piercy, 1974, Neuropsychologia, 12; Sussman, 1993, J. Speech Language Hearing Research, 36). This suggests that LI listeners might be extrasensitive to acoustic details. This issue was addressed using the visual world paradigm. Participants (32 TD, 37 LI, 14–20 years old) saw pictures on a computer screen, including a minimal pair differing in VOT (e.g., *beach/peach*). Auditory stimuli were six 9-step VOT continua constructed from natural speech. Participants heard a stimulus and clicked the referent. Eye movements were monitored to measure lexical activation. Both groups showed identical categorization functions (mouse clicks). Looks to competitor objects (peach when they clicked beach) increased as a function of VOT for both groups with no group interaction on either side of the boundary (B: $p<.001$; P: $p<.001$). Results were identical when the step closest to the boundary was removed, indicating sensitivity to *within-category* acoustic details. However, the LI group showed a heightened competitor activation overall (P: $p<.01$; B: $p=.06$). Thus, problems inhibiting lexical competitors, not perceptual differences, may be an underlying cause of language impairment.

1aSC17. Retuning speech sound categories: An eyetracking study. Kaori Idemaru and Lori Holt (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213)

Speech categories are defined by multiple probabilistic acoustic cues. Fundamental frequency (*F*0) and voice onset time (VOT) are correlated in the English stop voicing contrast, for example. However, such correlations are often imperfect—especially in cases of non-native or disordered speech. The present experiments investigate listeners' ability to adjust perceptual cue weighting in online perception in response to changes in the cue correlations experienced across time. Native-English listeners heard minimal-pair words beginning with stop consonants varying along a VOT series. The *F*0 of the words was gradually shifted over the course of the experiment from the canonical English correlation (higher *F*0 for voiceless stops) to the opposite pattern (lower *F*0 for voiceless stops). Categorization was assessed via explicit responses while eye gaze data were simultaneously recorded using the visual world paradigm. Both data types revealed that the influence of *F*0 on voicing categorization changed in response to changes in *F*0/VOT correlation. Some listeners use of *F*0 reversed such that higher *F*0 led to more voiced responses; other listeners discontinued use of *F*0 in voicing categorization. These patterns suggest that listeners are continually monitoring the input for regularity and retuning acoustic cue use in an online manner to accommodate these regularities.

1aSC18. Characteristics of listener sensitivity to talker-specific phonetic detail. Rachel M. Theodore and Joanne L. Miller (Dept. of Psych., 125 NI, 360 Huntington Ave., Boston, MA 02115-5000)

Listeners are sensitive to talker differences in phonetic properties of speech, including voice-onset-time (VOT) in word-initial stop consonants. Earlier findings from our laboratory [R. M. Theodore and J. L. Miller, J. Acoust. Soc. Am. **123**, 3934 (2008)] indicate that learning how a talker produces one voiceless stop (e.g., /p/ in *pain*) transfers to another voiceless stop (e.g., /k/ in *cane*), providing support for feature-based processing of VOT at the level of individual talkers. Here we examined possible constraints on such processing by asking whether transfer would also occur when the learning and transfer words were not minimal pairs. In familiarization phases, listeners heard two talkers produce *pain*. Critically, word-initial VOTs were manipulated such that one talker produced *pain* with relatively short VOTs and the other talker produced *pain* with relatively long VOTs. In test phases, listeners were presented with a short-VOT and a long-VOT variant of *coal* produced by each talker, and were asked to select which variant was most representative of the talker. Results showed that the listeners selected the VOT variant of *coal* in line with their previous exposure to *pain*, indicating that feature-based processing of talker-specific VOT is robust.

1aSC19. Categorization, category goodness, and psychophysical difference in perceptual assimilation. Jenna Silver (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611), Ratree Wayland (Univ. of Florida, Gainesville, FL 32611), and James D. Harnsberger (Univ. of Florida, Gainesville, FL 32611)

Two experiments were conducted to investigate the relative importance of three factors that influence the perceptual assimilation (i.e., identification and discrimination) of non-native speech contrasts: the categorization of non-native speech sounds with native speech sounds, the goodness of fit in such categorization judgments (or category goodness), and the phonetic differences between non-native stimuli. These factors were examined in a series of categorial *AXB* discrimination and forced-choice identification tests (which included goodness ratings) involving the perception by 80 native speakers of American English of 2 Hindi voicing contrasts, /tS/-/dZ/ and /k/-/g/, produced by three talkers in initial position in three different vowel contexts. The discrimination scores were correlated with (1) categorization scores based on the proportions of different responses in identification, (2) mean difference in goodness ratings, and (3) the results of an analysis of multiple acoustic cues for the Hindi stimuli (representing phonetic difference). Overall, the categorization factor showed the strongest correlation with discrimination scores ($r=0.87^{**}$), relative to phonetic difference ($r=0.57^{*}$), which was accounted for exclusively by measures of voice onset time. For non-native contrasts that were identified with single native categories, category-goodness difference was a more important factor in discrimination ($r=0.61^{**}$) relative to phonetic difference ($r=0.50^{*}$).

1aSC20. Perception of Japanese mora nasal /N/ and mora obstruent /Q/ by native Japanese, English, and Thai speakers. Takeshi Nozawa (Program in Lang. Education, College of Economics, Ritsumeikan Univ., 1-1-1 Nojihigashi, Kusatu, Shiga 525-8577, Japan, t-nozawa@ec.ritsumei.ac.jp) and Yuriko Furukawa (Osaka Univ., Mino, Osaka 562-8558, Japan)

Japanese mora nasal /N/ and more obstruent /Q/ are produced at different places of articulation, being homorganic with the following consonant. Four native speakers of Japanese produced /N/ and /Q/ in /CVNCV/ and /CVQCV/ contexts, where the consonant after /N/ and /Q/ is always a stop. Their utterances were recorded and digitized. The word final /CV/ was edited out, and the stimuli with the structure of /CVN/ and /CVQ/ were created. Twelve native speakers of Japanese, American English, and Thai were recruited as listeners. The listeners were told to identify the syllable-final nasal (/m/, /n/ or /ŋ/, and the missing stop (/p/, /t/, /k/, /b/, /d/, or /g/). Generally, the Thai listeners outperformed the other two groups of listeners, and the Japanese listeners performed most poorly. The Thai listeners good performance may be partly explained by the fact that in Thai postvocalic stops are normally unreleased, and the Thai listeners were less dependent on the release burst to identify the place of articulation of a postvocalic stop. The Thai listeners tended to identify voiced stops as voiceless stops of the same place of articulation. This may be because voiced stops cannot occur postvocalically in Thai.

1aSC21. Phonetic discrimination is affected by native language phonology. Jessica Maye (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, j-maye@northwestern.edu)

English and Japanese speakers were tested on discrimination of phonetic contrasts that vary in their phonemic status according to phonological context. The presence versus absence of /u/ in English stimulus pair *ebzo-ebuzo* is nonphonemic in Japanese due to phonological syllable structure constraints, but in word-initial position (stimulus pair *zobe-uzobe* the same contrast is phonemic. Similarly, the contrast between Japanese /d/ versus /r/ is nonphonemic in the foot-medial position in American English due to an allophonic flapping process, but in word-initial position the two sounds are perceived as the English phonemic contrast /d/ versus /l/ (Japanese stimulus pairs *gudo-guro* and *dogu-rogu*. Participants were tested in a four-token oddball task on all four stimulus sets in a blocked design. Each participant group (Japanese and English) showed significantly poorer discrimination of the non-native contrast when it appeared in the phonological context in

which it is nonphonemic in their native language (p's < 0.001). This result was equally robust even for stimuli in which the contrastive portion of the stimulus (*zo-uzo*, *do-ro*) was spliced from the nonphonemic context and pasted into the phonemic context. These data indicate that phonological context plays an integral role in the perception of phonetic signals.

1aSC22. Articulatory gestures influence the perception of speech. Henny Yeung (Dept. of Psych., Univ. of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada, hhyeung@psych.ubc.ca), Mark Scott, Bryan Gick, and Janet Werker (Univ. of Birtish Columbia, Vancouver, BC V6T 1Z4, Canada)

A central claim of the motor theory of speech perception [Liberman *et al.* (1967)] is that speech perception involves motor representations. We report evidence that articulatory movements influence perception. Subjects made forced-choice identifications of naturally recorded /aba/ and /ava/ tokens, while silently making articulatory gestures in time with presentation of the auditory tokens. These articulatory gestures either agreed with the auditory token (e.g., articulating "aba" while hearing /aba/) or disagreed (e.g., articulating "ava" while hearing /aba/). Subjects more frequently misidentified auditory /aba/ as /ava/ when articulated gestures disagreed, as compared to a base line condition (i.e., simply listening). Two further conditions suggest that simple priming of /ava/-percepts is unlikely. First, subjects articulated "afa" instead of ava while hearing /aba/. If error rates are specifically related to articulatory gestures, then the similarity in gestural movements between /f/ and /v/ should result in similar error rates. A priming account would not make such a prediction. In line with the motor explanation, subjects did have an equivalent error rate in afa and ava conditions. Second, minimal interference was observed when subjects only imagined themselves saying /ava/. These results support the notion that activation of motor movements can influence the perception of speech.

1aSC23. Investigating the consonant-vowel boundary. II. Perceptual contributions of glimpse windows. Daniel Fogerty and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, dfogerty@indiana.edu)

Discrete segmental units do not occur in fluent speech because of coarticulation. However, separation of the meaningful sounds in language into the categories of consonants and vowels is one of the most fundamental principles of how sound is structured. Our previous research has shown that for sentence presentations, vowels have a distinct perceptual advantage over consonants in determining sentence intelligibility. TIMIT sentences were used to investigate perceptual contributions of consonants and vowels across the consonant-vowel (CV) boundary, by shifting the CV boundary by specific proportions of the vowel, such that consonant duration increased while vowel duration decreased. Glimpse windows are defined as the speech signal preserved between noise replacements. The perceptual effect of windows either locked to specific segmental information or placed randomly was examined. Results from glimpse windows locked to segmental information confirmed a 2-to-1 vowel advantage for intelligibility at the traditional CV boundary and suggest that vowel contributions remain robust against deletions of the signal. When glimpses were presented randomly, summed duration of glimpses predicted performance. However, performance remained lower than when glimpse windows of equivalent duration were locked to vowels. Specific segmental information appears to differentiate performance between consonant and vowel conditions. [Work supported by the NIH.]

MONDAY AFTERNOON, 10 NOVEMBER 2008 LEGENDS 9, 1:00 TO 3:10 P.M.

Session 1pAAa

Architectural Acoustics: Acoustics of Single Family Residences

Richard D. Godfrey, Cochair
448 N. Pearl St., Granville, OH 43023

Nancy S. Timmerman, Cochair
25 Upton St., Boston, MA 02118-1609

Chair's Introduction—1:00

Invited Papers

1:05

1pAAa1. New home buyer noise reduction listening study. Harry A. Alter (19 Beechtree Ln., Granville, OH 43023)

In 2007 Owens Corning Science Technology in Granville, Ohio built a comparative wall assembly listening facility called the Acoustic Research Experience Laboratory (AREL) Annex. Within this listening facility, approximately 100 perspective new home buyers were selected to listen and evaluate various sound clips when played through wall assemblies at varying levels of noise reduction. Jurors responded to perceptual and lifestyle questioning. The building of the AREL Annex and listening study results will be discussed.

1:25

1pAAa2. Effects of residential audible distractions on the performance and perception of home office workers. Alicia J. Wagner and Lily M. Wang (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, aliciajean01@gmail.com)

This research examines the effects that typical residential audible distractions have on task performance and subjective perception of home office workers when transmitted across various residential wall constructions. Previous studies have investigated how speech, music, and office equipment noises may deteriorate one's performance, but few have presented these distractions in such a way as to imitate the home office environment or used audible distractions that are characteristic of a residential setting. In this study, subjects performed math, verbal, and typing tasks over 1 h while exposed to four types of audible distractions: (1) pop music, (2) television, (3)

conversation, and (4) a potpourri of kitchen, pet, and children's noises. A short questionnaire was also administered to determine subjective perception under the different noise signals. Statistical analyses of the results indicate that none of the audible distractions caused a significantly different effect on task performance. The loudest distraction (music), however, was subjectively rated to be most distracting, while the signal with the most variation in time (potpourri) was considered to be the most annoying.

1:45

1pAAa3. Auralization as a tool for acoustical design of single family residences. Carl Rosenberg, Jonah Sacks (Acentech Inc., 33 Moulton St., Cambridge, MA 02138, crosenberg@acentech.com), Erin Dugan, and David McDonald (USG Corp., Libertyville, IL 60048)

Acoustical requirements play an increasing role in residential building design. In an effort to understand these changes USG Corp., a manufacturer of building materials and developer of acoustical solutions for residential and commercial markets, engaged Acentech Inc. to develop aural demonstrations, or auralizations, of common acoustical issues in single family residences. Using computer modeling and audio processing techniques, a number of convincingly realistic simulations of familiar residential scenarios were developed. These auralizations allowed USG to better understand acoustical issues and recommend appropriate products and systems. This paper shares the experience of this project.

2:05

1pAAa4. Incorporating acoustically challenging functions into single family residences. Russ Berger (RBDG, 4006 Belt Line Rd., Ste. 160, Addison, TX 75001, russ@rbdg.com)

Recording and broadcast studios, screening theaters, music performance venues, and other acoustically challenging functions have been successfully incorporated into single family residences. Several examples will be discussed illuminating concerns beyond normal sound isolation and acoustical performance requirements.

Contributed Papers

2:25

1pAAa5. Sound in a single family house. Sergio Beristain (Lab Acoust., ESIME, IPN. IMA President, P.O. Box 121022, 03001 Mexico, D.F., Mexico. sberista@hotmail.com)

Apart from road and air traffic noise, and some noise from neighbors, the most important noise in single family residences usually is the noise generated within the residence itself, such as music, general activities, operating machinery for the house and people comfort, and the sound isolation characteristics of the different partitions within the house. Mexican single family residences are usually made out of brickwork, but some inner walls are nowadays changing from brickwork to rocksheet and similar materials, which reduces noise attenuation from kitchens, washing rooms, and living spaces. Home theaters are usually installed at will, and the resulting sound has limitations. A case is presented where family relations started to deteriorate due to the low isolation capabilities of the inner walls.

2:40

1pAAa6. A measurement survey of the acoustic conditions in home offices. Megan J. Christensen and Lily M. Wang (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, lwang4@unl.edu)

As home offices are becoming more prevalent in our society, it is critical to ensure that the acoustic conditions in those environments are suitable for optimal home office worker performance. In this paper, a measurement survey is presented of existing conditions in several home offices within the metropolitan Omaha, Nebraska area. The data that were gathered include background noise levels (both with and without ventilation noise), air-borne sound attenuation of the office envelope, and logged data of the office sound levels during a typical 48-h period over the work week. Summaries of the findings will be discussed and are useful in understanding how residential design and construction should be modified to improve the acoustic conditions of home offices.

2:55

1pAAa7. Acoustical considerations for luxury motor homes. Jennifer Shaw and Charles Moritz (Blachford Acoust. Lab., 1445 Powis Rd., West Chicago, IL 60185, jshaw@blachfordinc.com)

Although not considered a typical single-family home, a large number of luxury motor home owners spend at least 6 months per year living in their coach. A survey of class A motor home owners confirmed that for these owners, the coach becomes a home-away-from-home and requires many of the same amenities of a real home, including low interior noise levels. The survey noted that when parked at a campsite or motor home resort, residential noise issues were a significant concern, particularly noise from a neighbor's generator. Acoustical issues found in motor homes, along with the results of the survey, will be compared with updated results from previously reported motor home testing [J. T. Kunio and C. T. Moritz, SAE 2003 Trans., **112**, 1800–1810 (2003)] to understand how these concerns may be addressed.

Session 1pAAb

Architectural Acoustics: Multifamily Structure—Advances and Legal Issues

Angelo J. Campanella, Chair
Campanella Associates, 3201 Ridgewood Dr., Columbus, OH 43026-2453

Chair's Introduction—3:20

Contributed Papers

3:25

1pAAb1. Who pays? Defining construction defects and assigning blame in borderline acoustical assemblies. John LoVerde and Wayland Dong (Veneklasen Assoc., 1711 16th St., Santa Monica, CA 90404, jloverde@veneklasen.com)

It is often the case when performing field noise isolation testing that some assemblies will meet the relevant statutory requirements and other nominally identical assemblies will fail. Sometimes this can be attributed to errors in construction materials and/or methods. However, when an assembly is "borderline," that is, only just meets the requirements based on historical or laboratory data, it can be expected that a certain percentage will fail even when constructed without error. This is due to the variation in construction procedures and materials as well as the inherent uncertainties in acoustical measurements. As a result, it is often not clear whether any particular assembly failed because of construction error or inadequate design and, by extension, whether the contractor or the designer bears the responsibility and resultant financial liability. The definition of construction defects in acoustical assemblies and the resultant division of responsibility are discussed and illustrated with actual cases.

3:40

1pAAb2. Performance issues of resilient channels in wall systems. Stephen W. Payne, Jr. (USG Corporate Innovation Ctr., 700 N. Hwy., 45 Libertyville, IL 60048)

There has been much recent discussions on the issue of proper installation of resilient channels. Of particular concern is the issue of the "shorting out" of the resilient channels, wherein the gypsum board attachment screws passes through the resilient channel and into the structural member. This paper will present the test results obtained during a series of tests to evaluate the significance of this issue. In addition data will be presented from a further study on the significance of the geometry of the resilient channel design.

3:55

1pAAb3. Characterization of airborne sound transmission and impact isolation in floor/ceiling assemblies with structural cementitious floor sheathing and cold formed steel framing. Timothy D. Tonyan (Systems Development, Structural Technologies Group, U.S. Gypsum Co., 21925 W. Field Pkwy., Ste. 245, Deer Park, IL 60010), Stephen W. Payne, Jr. (USG Corp. Libertyville, IL 60048), and Robert Elfering (Shiner + Assoc., Chicago, IL)

There is growing demand for floor/ceiling assemblies that combine noncombustible structural cementitous floor sheathing panels with cold-formed steel (CFS) framing. The combination being light in weight and possessing high stiffness makes these floor systems particularly advantageous in seismic zones, where reduction in building mass can translate into structural efficiencies and cost savings. However, as a result of their relatively low mass appropriate design strategies must be used to control airborne sound transmission and impact noise. This paper presents research conducted to characterize the acoustical behavior of floor/ceiling assemblies using structural cementitious floor sheathing combined with CSF framing. A variety of isolation and damping strategies are described and tested. The influence of the use of different underlayments and acoustical mats on STC and IIC values in the floor/ceiling assemblies is documented and evaluated. Isolation approaches, including the use of drywall suspension systems for ceiling attachment, are also presented. The results demonstrate that, by using the appropriate damping and isolation strategies, floor/ceiling assemblies combining structural cementitious floor sheathing with CFS framing can provide excellent acoustical performance. Recommendations for further research are also provided.

4:10

1pAAb4. Acoustic and vibratory characterization of in-room footfall noise: Part 2. Robert M. Tanen, Jonathan C. Silver, Michelle C. Vigeant, and Robert D. Celmer (Acoust. Program + Lab., Mech. Eng. Dept., Univ. of Hartford, 200 Bloomfield Ave., W. Hartford, CT 06117, celmer@hartford.edu)

Prior to Phase I of this study [J. Acoust. Soc. Am. **22**], existing footfall literature was primarily focused on transmission between spaces, i.e., IIC. This study measured the sound power and vibratory spectra produced by footfalls using both human subjects and standard tapping machines. One tapper had rubber-tipped drop weights and the other had cored samples of shoe soles used by human participants. Within a reverberation room, 12 floor profiles were tested with the tapping machines. Human subjects were tested with three previously tested floors and a new floor profile. Fourteen male and female subjects walked on the floor surfaces while wearing three different types of footwear: leather-soled shoes (hard), rubber-soled shoes (medium), and sneakers (soft). Sound power spectra and vibratory signatures for each condition were measured using the procedures of ISO 3741. The current study verified repeatability from Phase I to Phase II, produced sound power footfall data for vinyl flooring applied directly to concrete, and developed improved correction curves to model tapping machines as human footfalls. Similarities between the rubber-tipped tapper and women's leather shoes and a comparison of sound power and corresponding vibratory signatures are discussed. [Worked supported by The Paul S. Veneklasen Research Foundation.]

4:25

1pAAb5. Sound quality of laminate flooring systems. James Wilson, III and Kenneth Cunefare (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The laminate flooring industry identifies the ability of a laminate floor to recreate the sound of traditional hardwood floors as an important metric for their perceived quality and acceptance. The market describes the inability of laminate flooring to reproduce the natural sound of wood as a drawback to their market appeal. Therefore, if a laminate flooring composite can be offered to the market, which recreates the natural wood floor acoustic experience, the floor will offer additional value to the consumer. Products on the market today exist, which do improve the sound quality of laminate flooring composites. However, there is no unified standard to justify or prove marketing claims made by manufacturers. Prior work in the area focused on the spectral content of the signal. In the area of psychoacoustics, many additional metrics exist to describe the human perception of sound, which aids in describing sound quality. These metrics can be utilized to characterize the

sound of traditional hardwood floors and laminate systems, beyond what has been done in previous work. By employing these metrics and then correlating them to subjective sound jury perception, it is possible to better predict the sound quality of a laminate floor.

4:40

1pAAb6. Conceptual design considerations for the acoustical design of an animal research facility. Byron Davis and Kevin Richardson (VACC, 490 Post St., Ste. 1427, San Francisco, CA 94102, byron@vaconsult.com)

Conceptual design considerations from the acoustical design of an animal research facility are presented in this paper. Animals in laboratory settings are experimentally stressed to observe outcomes. Stressors external to the experiment confound data and reduce research productivity. Among these experimental contaminants is acoustical noise, which has well-documented extra-auditory effects in animals. Both steady-state and transient sounds can affect development, sleep, reproduction, behavior, and other parameters important to animal models. Reasonably comprehensive sensitivity data exist in the medical and laboratory literature; however, there are few treatments of the acoustical design of vivarium and animal laboratory facilities. Acoustical design of these facilities is challenging due to the broad array of sensitivities (including ultrasonic frequencies), scarcity of high-frequency data for noise sources, contamination control and cleanliness requirements, and surface durability concerns. Acoustical design criteria for animal holding areas and experimental test rooms are considered. The criteria are based not only on sound pressure levels but also on frequency content, isolation of incompatible species, expected vocalizations, and minimization of disruption. Noise sources include mechanical/electrical/plumbing equipment, maintenance activities, and other animals. The design efforts include categorization of animal groups, noise isolation concepts, acoustical design, and noise control measures.

4:55

1pAAb7. Room sound absorption provided by underbench sound absorbing material. Pamela Harght and Robert Coffeen (School of Architecture and Urban Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

The concept that sound absorbing material added beneath "hard" benches and pews will be useful in the reduction in reverberation in religious worship facilities, and other places of assembly are occasionally advanced by facility owners and users who do not desire to employ bench and pew seat upholstery or seat cushions. This paper presents data on underbench sound absorption calculated from measurements made in a reverberant space with and without the addition of a sound absorbing material to the underside of wood benches.

MONDAY AFTERNOON, 10 NOVEMBER 2008 LEGENDS 8, 1:25 TO 4:30 P.M.

Session 1pAO

Acoustical Oceanography and Underwater Acoustics: Acoustics and Inversions on the Continental Shelf and Canyons

James F. Lynch, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543-1541

Jennifer L. Wylie, Cochair
Rosenstiel School of Marine and Atmospheric Sci., 4600 Rickenbacker Causeway, Miami, FL 33149

Chair's Introduction—1:25

Contributed Papers

1:30

1pAO1. Observed intensity and horizontal field coherence variability of low-frequency pulse transmissions on the continental shelf. Timothy F. Duda, Jon M. Collis, Ying-Tsong Lin, and James F. Lynch (Woods Hole Oceanograph. Inst., AOPE Dept., MS 11, Woods Hole, MA 02543)

Rapid coastal environmental evolution leads to highly variable acoustic fields. To quantify such variability, one component of the Shallow-Water 2006 (SW06) program on the shelf east of New Jersey was time series measurement of sound transmitted from fixed sources to joined horizontal and vertical line arrays. Transmission paths were both cross-shelf and along-shelf (across and along dominant internal-wave crests). Data were collected for over one month. Intensity time series of 100–400-Hz pulses was found to have strong variability at periods from hours to over a day, consistent with long-wavelength internal-tide effects. Such effects can arise from adiabatic mode and/or coupled mode propagation. Separation of fluctuations into slow and rapid contributions allows calculation of a time-varying horizontal coherence-length statistic. For along-wave crest transmission, this was highly variable, typified by values ranging from a few acoustic wavelengths to over 40 wavelengths, typically 10–25. The slow coherence-length fluctuations had signatures of periodic (tidal) mode-refraction episodes (with short scale) during active intervals, caused by internal-wave ducting. Conditions were more steady at other times. Across-crest transmissions showed shorter than expected scale lengths of tens of wavelengths with more subtle tidal dependence. [Work supported by the Office of Naval Research.]

1:45

1pAO2. Acoustic hindcasts of array performance using a combination of submesoscale hydrodyamic and acoustic models. Steven Finette, Roger Oba, Thomas Hayward, Colin Shen (Naval Res. Lab., Washington DC 20375-5320), Patrick Gallacher, Alex Warn-Varnas, and Steve Piacsek (Naval Res. Lab., Stennis Space Ctr., MS 39529-5004)

It is well known that submesoscale ocean processes such as nonlinear internal gravity waves can have a significant effect on both the amplitude and the phase of an acoustic field propagating through this type of ocean environment. We report here on hindcasts computed with a numerical model combination consisting of a submesoscale hydrodynamic solver to compute a set of three-dimensional (3-D) environmental (sound speed) volumes evolving in time and a 3-D wide-angle parabolic equation code for acoustic field computation within each environmental volume. The data set was chosen from the ASIAEX 2001 experiment in the South China Sea, during a period of strong internal wave activity. A nonlinear wave packet was simulated propagating up the shelf and passing through both the acoustic source and receiver positions. The hindcasts computed the time evolving beam response on a horizontal array, located approximately 19 km from a 300 Hz

low-frequency modulat source. Comparison of experimental and modeled beamformed results was made using three different submesoscale codes to compute the dynamic ocean environment. Each code was exercised using the judgment of a different oceanographer, given the same set of measured bathymetry and water column properties. [Work supported by the Office of Naval Research.]

2:00

1pAO3. Phase fluctuations and horizontal refraction of low-frequency sound signals in shallow water in presence of internal waves. Mohsen Badiey, Jing Luo (College of Marine and Earth Studies, Univ. of Delaware, Newark, DE 19716, badiey@UDel.Edu), Boris Katsnelson, Alexander Tskhoidze (Voronezh State Univ., Voronezh 294006, Russia), James Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and James Moum (Oregon State Univ., Corvallis, OR 97331-5503)

Fluctuations of the sound pulse phase front propagating approximately perpendicular to the direction of motion of a train of intensive internal waves in the experiment Shallow Water 2006 (SW06) are studied. Acoustic data received by a horizontal/vertical line array (WHOI-Shark array) during 3 h are analyzed. Low-frequency modulated signals with a carrying frequency of 300 Hz were transmitted, propagating approximately along coastal line at the source-receiver distance of 20 km. During this time, a train of intense internal waves (Rosey) was moving toward the New Jersey coast. Internal waves were being monitored by the ship radars (aboard R/V Sharp and R/V Oceanus) and by the thermistor chains. It is shown that phase fluctuations of the acoustic pulses arise with the appearance of internal waves at the source-receiver acoustic track and they correspond to horizontal refraction of the sound pulses. For the low-frequency sound pulses, modal decomposition of received signals is used both for the processing of experimental data and for the theoretical analysis. Experimental data are compared with theoretical estimation of horizontal refraction. [Work was supported by ONR code 321OA and by CRDF, Grant BP3C10].

2:15

1pAO4. Anisotropic properties of long-time intensity fluctuations of midfrequency signals in presence of intense internal waves in shallow water. Boris Katsnelson, Valery Grigorev (Dept. of Phys., Voronezh State Univ., 1 Universitetskaya sq, Voronezh 394006, Russia, katz@phys.vsu.ru), and James Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Intensity fluctuations of sequence of midfrequency pulses radiated for 5 h in SW06 experiment were considered. Frequency band of pulses is 2–10 kHz; they were radiated by the source (R/V Knorr) with the interval about 15 s and received by four hydrophones (SHRUs) placed at the distances from 4 to 14 km and different directions (SW, SSW, NE, and NNE) from the source. During this time period, remarkable train of internal waves (amplitude of displacement up to 15 m) passed toward coastal line at the velocity of 0.6 m/s. Angles between wave front of train and acoustical tracks were about 65–80 deg. Mechanism providing sound fluctuations in this case is mode coupling (for the low-frequency signals) or additional scattering of rays by internal waves (in high- or midfrequency situation). It was shown that temporal fluctuations of the sound intensity of received pulses initiated by internal waves have spectra, depending on the direction of propagation of internal wave relative acoustic track. Results of comparison of experimental data with theoretical estimations demonstrate good consistency. [This work was supported by ONR and RFBR.]

2:30

1pAO5. Variability of the water column sound speed profile and its effect on acoustic propagation during the Shallow Water 2006 Experiment. Megan S. Ballard and Kyle M. Becker (Appl. Res. Lab. and Graduate Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

Water column sound speed variability can have a significant effect on acoustic propagation. However, measurements of water column properties, either by ship based systems or moored instruments, are often neither spatially nor temporally colocated with acoustic measurements. Variability of the water column over the acoustic propagation path must be approximated by extrapolation. During SW06, a low-frequency sound source, broadcasting pure tones, was towed along radials from a vertical array comprised of both hydrophones and oceanographic sensors. Concurrently, and from the same ship, a towed CTD chain was used to measure the water column as a function of both time and space. Changes in sound speed as high as 15 m/s were recorded over a 5-km aperture. The CTD chain, in combination with environmental sensors on the VLA, allowed for continuous monitoring of the water column sound speed profile at both the source and receiver locations. It is shown that horizontal wavenumber predictions for the acoustic field are biased by the sound speed measured at the source compared to at the receiver. In addition, predictions of the acoustic field are improved when the range dependence of the sound speed profile is considered. [Work supported by NDSEG and ONR.]

2:45

1pAO6. Temporal fluctuations and coherencies of broadband signals observed during Shallow Water 2006. Jennifer Wylie and Harry DeFerrari (Univ. of Miami, 4600 Rickenbacker Cswy, Miami, FL 33149, jwylie @rsmas.miami.edu)

A 50 h acoustic propagation experiment conducted during SW06 included receptions of broadband signals at three ranges with single hydrophone receiver units. 100–1600 Hz centered broadband signals were propagated through identical shallow channels during periods of low- and high-internal wave energies. Here, the transition from nearly perfect coherent and stable mode/ray arrivals to the formation of short live micromodes/ray arrivals is presented. The relative influence of volume scattering and bottom scattering is seen to shift with frequency; low frequencies see a smooth reflective bottom but have interaction with internal waves induced volume fluctuations, whereas higher frequencies are increasingly more sensitive to bottom scattering. Discrete mode arrivals give way to a continuum of numerous short lived arrivals with increasing frequency. For still higher frequencies, coherence times are shorter still and at some point the phase coherence is so reduced that signal processing gain is lost and the signals are no longer detectable. Ultimately, the performance of higher-frequency sonar and underwater communication systems will be limited by these effects. The limits of phase coherent gain for the 800 and 1600 Hz signals are estimated for this range site.

3:00—3:15 Break

3:15

1pAO7. Investigation of an unusual noise phenomenon with horizontal and vertical hydrophone array data and three-dimensional propagation modeling. Georges A. Dossot, James H. Miller, Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett Bay Campus, Narragansett, RI 02882), James F. Lynch, Arthur E. Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

Acoustic data from an experiment conducted during Shallow Water 2006 (SW06) showed an unexplained broadband noise phenomenon. While the R/V Knorr towed a J-15 acoustic source, which emitted a 93 Hz continuous-wave signal, the R/V Endeavor performed Scanfish measurements in the vicinity to characterize internal wave phenomena. The WHOI Shark horizontal and vertical hydrophone array detected the 93 Hz signal, but eventually the signal was overpowered by broadband low-frequency noise. The broadband noise may be associated with either research vessel but is uncharacteristically intense given that both vessels were greater than 25 km away from the Shark array. The complex acoustic environment due to the shelfbreak front, an internal wave packet, and the bathymetry of the continental shelf may have caused ducting of the noise field. Both three-dimensional acoustic modeling and array processing techniques will be used to characterize the unexplained noise levels. *In situ* data from temperature sensors and Scanfish measurements provide environmental information needed to accurately model the sound speed field. These methods can also be applied to other SW06 experiments that were tailored to investigate the role of internal waves in acoustic propagation—such as those carried out on the R/V Sharp. [Work sponsored by the Office of Naval Research.]

3:30

1pAO8. Nonlinear internal wave parameter extraction from images synthetic aperture radar. C. Chaya Boughan (Rensselaer Polytechnic Inst., Troy, NY 12180, boughc@rpi.edu), Timothy F. Duda, James F. Lynch, Arthur E. Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

Enhancement of satellite synthetic aperture radar (SAR) ocean images is needed for improved predictions of nonlinear internal wave effects on acoustic amplitude, phase, and coherence. Their effective use in acoustic applications can be hampered by the false negative problem of SAR images that fail to reveal internal wave presence. Our aim is a largely automatic extraction of nonlinear internal wave features such as wave amplitude, wave front width, wave front separation, and longwave correlation length. A complex discrete wavelet transform is used to denoise contrast-enhanced SAR images, and a fingerprint recognition algorithm is applied to extract edges. The image is segmented into connected components that are analyzed to obtain estimates for nonlinear internal wave parameters. To evaluate the process, estimates are compared with temperature and other data from moorings positioned approximately 100 miles east of the New Jersey coast. Extensions of the procedure may give an acoustically useful description of nonlinear internal waves in SAR images. Assimilation of parameter estimates with volume data offers a new approach to appraising internal wave effects on propagation. [Work supported by the ONR.]

3:45

1pAO9. Sensitivity issues in acoustic mode tomography. Tarun K. Chandrayadula and Kathleen E. Wage (Dept. of Elec. and Comput. Eng., Mailstop 1G5, George Mason Univ., Fairfax, VA 22030)

Ocean acoustic mode tomography relies on a high signal to noise ratio and an accurate measure of the time of arrival for the mode signals. The low-order modes measured during tomography experiments such as the 2004 long range ocean acoustic propagation experiment (LOAPEX) are sensitive to environmental and experimental conditions. For example, internal waves can cause mode coupling and time of arrival perturbations. The source and the VLA move due to currents and tides, which results in Doppler spreading and frequency selective fading. Uncertainties in the signal affects the accuracy of tomographic methods. Previous work [Chandrayadula et al., J. Acoust. Soc. Am. 121, 3053 (2007)] mainly focused on building a channel model for the internal wave effects. This paper extends the previous work by building a more complete channel model that also includes the effects due to residual source/receiver motion. The statistics from the channel model are used to study the effect of uncertainties on the mode signals measured during LOAPEX. [Work supported by ONR Ocean Acoustics Graduate Traineeship Award.]

4:00

1pAO10. Acoustic wave scattering from submerged turbulent wakes. Tokuo Yamamoto (Appl. Marine Phys. Div., RSMAS, Univ. of Miami, Key Biscane, FL 33149)

Acoustic wave scattering in submerged turbulent wakes has been solved on the assumption of small and smooth turbulent fluctuations. The turbulent fluctuations are modeled by the Kolmogorov spectrum. The turbulence in the submarine wakes is modeled by a MIT Ocean Model by pat.gallacher @nrlssc.navy.mil. The effect of turbulent velocity fluctuation and the effect of turbulent temperature fluctuation are roughly equal on the scattering of acoustic wave from turbulent wakes (1 Hz–1 MHz). Acoustic scattering by a turbulent wake is very strong in the forward direction while it is small in the backward direction. The scattering from internal waves is modeled using the Garret–Monk spectrum. Since the frequency spectrum (1 Hz–1 MHz) of wake turbulence and that of internal waves (0.0138 Hz–10 MHz) do not overlap the two scattering mechanisms do not interact. The two mechanisms of scattering can be treated separately and then added linearly. [Work supported by ONR Code 321 OA.]

4:15

1pAO11. An overview of the 2005 YFIAE: Yellow Sea Oceanic Front and Internal Waves Acoustic Experiment. Ning Wang (Dept. of Phys., Ocean Univ. of China, 238-Songling Rd., Qingdao 266003, China, wangyu@public.qd.sd.cn), Jin Zhong Liu (Weifang College, Shandong, China), Da Zhi Gao, Wei Gao, and Hao Zhong Wang (Ocean Univ. of China, Qingdao 266003, China)

An overview of the Yellow Sea Oceanic Front and Internal Waves Acoustic Experiment 2005 conducted in the South Yellow Sea is presented. The experiment was a multi-institutional effort which acquired high quality environmental and acoustic data. Two of its goals were to observe internal waves including linear random and nonlinear and cold water mass of South Yellow Sea to understand and describe their effects on acoustic signals. The talk contains three primarilary experimental results. First, a comparison of data from VLA and thermistor string record suggests that nonlinear internal waves cause coherent variations of arrival time structure in frequency bands 600–800 Hz. Second, normal mode coupling due to the South Yellow Sea Front (SYSF) is observed during the experiment and is used to monitor the boundary motions of the SYSF. Finally, to interpret the "frequency dispersion" in inverted effective sound speed of sea bottom, a new fast algorithm for the inversion of bottom acoustic parameters is proposed.

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Interaction with Tissues

Michael L. Oelze, Chair
Univ. of Illinois at Urbana-Champaign, Electrical and Computer Engineering, Urbana, IL 61801

Contributed Papers

1:30

1pBB1. Evaluation of a singular-spectrum analysis algorithm for detecting two types of brachytherapy seeds commonly used to treat prostate cancer. Jonathan Mamou, Sarayu Ramachandran, and Ernest J. Feleppa (Riverside Res. Inst., F. L. Lizzi Ctr. for Biomedical Eng., 156 William St., New York, NY 10038)

Brachytherapy for prostate-cancer treatment involves permanent implantation of radioactive seeds into the gland. Reliable imaging of implanted seeds is needed to correct dosimetry errors in the operating room. A singular-spectrum analysis algorithm previously was able to detect palladium seeds over various angles between the seed and ultrasound-beam axes. We subsequently evaluated algorithm performance with iodine seeds. The algorithm extracts pairs of eigenvalues from the autocorrelation matrix of seed echo signals. Selected eigenvalues are used to compute a *P*-value for seed presence. The algorithm was applied to echo signals obtained using a 5-MHz transducer to scan seeds implanted into an acoustically transparent gel pad and a piece of *ex vivo* beef. The angle of the seed axis with respect to the beam axis was varied from normal to the beam axis to 20 deg from normal. Algorithm performance was denoted by a score computed from *P*-values. Scores for both seed types in beef varied from 70 to 40 dB over the range of angles studied. Scores computed from *B*-mode images were approximately 30 dB lower indicating superior contrast with the SSA algorithm. The SSA algorithm was successful in detecting palladium and iodine seeds over a range of angles to the ultrasound beam.

1:45

1pBB2. Derating of nonlinear high intensity focused ultrasound fields to predict millisecond boiling in tissue. Michael Canney (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, mcanney@u.washington.edu), Olga Bessonova, Vera Khokhlova (Moscow State Univ., Moscow 119992, Russia), Michael Bailey, and Lawrence Crum (Univ. of Washington, Seattle, WA 98105)

The most commonly used method for derating high intensity focused ultrasound (HIFU) fields from water to tissue is based on multiplying the acoustic intensity measured in water by an exponential factor to compensate for attenuation in the tissue path assuming linear wave propagation. Yet, in nonlinear HIFU fields, the intensity provides little information about either heating or negative and positive pressure amplitudes, which are important in predicting bioeffects. In this work, a new derating method is presented and tested for a 2 MHz high gain focused ultrasound source. Focal waveforms are experimentally measured and modeled after propagation through both water and tissue paths at output intensities of up to 24 000 W/cm^2. The focal waveforms measured after propagation through tissues were made equivalent to those obtained in water by increasing the pressure amplitude at the source. From the change in source amplitude pressure, the attenuation of the tissue was determined. The focus was then shifted to within the tissue sample, and the measured attenuation was used to calculate the time to reach 100°C. Calculations were in excellent agreement with the time measured to attain boiling in the tissue, which was only several milliseconds. [Work supported by NIH DK43881 and NSBRI SMS00402.]

2:00

1pBB3. Viscoelastic response of cylindrical vessels surrounded by gelatin and excited using impulsive ultrasound radiation force. Matthew Urban (Dept. of Physiol. and Biomed. Eng., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, urban.matthew@mayo.edu), Daniel Rosario (Cornell Univ., Ithaca, NY 14853), Miguel Bernal (Mayo Clinic College of Medicine, Rochester, MN 55905), Wilkins Aquino (Cornell Univ., Ithaca, NY 14853), and James Greenleaf (Mayo Clinic College of Medicine, Rochester, MN 55905)

The objective of this study was to assess the potential for noninvasively identifying viscoelastic material properties in arterial vessels using ultrasound radiation force and modern computational inverse problem techniques. A rubber tube embedded in gelatin was excited at different points along the axis using impulsive ultrasound radiation force. The resulting mechanical waves were measured using a laser vibrometer and pulse-echo ultrasound. The inverse problem was cast as an optimization problem, in which the discrepancy between the measured dynamic response and the computed finite element model response was minimized with respect to viscoelastic material parameters. The model was very sensitive to the tube's viscoelastic parameters and relatively insensitive to those of the gelatin. The viscoelastic properties of the tube obtained from the inverse problem were compared to the results obtained from a dynamic mechanical analyzer (DMA). For the overlapping range of frequencies (130–190 Hz), the results of the inverse problem gave an equivalent modulus that range from 7.3–8.5 MPa, while the DMA values range from 6.0–9.2 MPa. These results have important implications for using ultrasound radiation force in noninvasive characterization of the viscoelastic properties of arteries *in vivo*. [This work was supported in part by Grant No. EB002640 from NIH.]

2:15

1pBB4. Study of Lamb wave dispersion in porcine myocardium. Ivan Nenadic, Matthew Urban, and James Greenleaf (Ultrasound Res. Lab., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905)

Diastolic dysfunction is the impaired ability of the left ventricle (LV) to passively fill during diastole and maintain its stroke volume at physiologic filling pressures and is associated with impaired relaxation and reduced compliance of LV. Knowledge of the quantitative values of the myocardial viscoelastic properties would improve the clinician's ability to evaluate diastolic dysfunction in patients. To gain better understanding of the viscoelastic properties of the LV as a function of geometry, we studied Lamb wave propagation in different directions in excised porcine free-wall myocardium. The phantom consisting of a dissected porcine LV free-wall embedded in gelatin was fixed in a water tank. A mechanical shaker was used for harmonic excitation over a range of 40–500 Hz and a 7.5 MHz ultrasound transducer was used for motion measurements. Motion amplitude and phase at different depths of the sample were analyzed to obtain dispersion curves in four orthogonal directions. The data were fitted to the Lamb wave equations and shear elasticity and viscosity coefficients were estimated. The values of μ_1 and μ_2 were 13.75 $\pm$ 0.96 and 5.86 $\pm$ 0.75 Pa s. Our results suggest that the viscoelastic characteristics of the porcine LV free-wall in orthogonal directions were similar.

2:30

1pBB5. Development of a lung tissue fatigue and failure model. Mark S. Wochner, Yurii A. Ilinskii, Mark F. Hamilton, and Evgenia A. Zabolotskaya (Appl. Res. Labs., The Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

The time-dependent nature of mammalian lung damage due to acoustic excitation at the lung resonance has been documented in recent experiments [Dalecki *et al.*, J. Acoust. Soc. Am. **116**, 2560 (2004)]. A finite-element based model of human lung response to low-frequency underwater sound has been developed to calculate induced shear stresses and strains, but until now we have not attempted to simulate progressive damage. Unlike for many man-made materials, there are very little experimental data from which fatigue parameters can be calculated for soft tissue. This presentation discusses our investigation into methods of modeling the cyclic fatigue known to occur in mammalian lungs exposed to low-frequency underwater sound. Attempts will be made to correlate reported thresholds for lung damage to computational results obtained by our model. Discussions of the limitations of such an approach and future applications will be discussed. [Work supported by ONR]

2:45

1pBB6. A mathematical model of very soft tissue for acoustics applications. Elisabetta Sassaroli, Brian E. O' Neill, and King C.P. Li (The Methodist Hospital Res. Inst., 6565 Fannin St., Houston, TX 77030, esassaroli@comcast.net)

A theoretical model of very soft tissue is presented. The soft tissue is considered to be composed of two continuum homogeneous media: the fluid medium and the solid medium. The average pressure, the velocity in the fluid, and the average displacement in the tissue can be obtained by averaging the equations of continuum mechanics over a scale that is large compared with the mean spacing between the solid and liquid regions but small compared to the length scale of the sample considered and also smaller than the ultrasound wavelength. The biphasic model presented here is inspired by the one originally proposed by Pride *et al.* who used it to model fluid-saturated porous materials, but here it is altered to suit the conditions for ultrasound propagation in a very soft tissue. The phase-averaged equations of continuity and momentum, as well as the stress-strain relations in the quasistatic limit, are discussed. In linear approximation, the averaged momentum equation for the liquid phase gives rise to a generalized Darcy's law.

3:00

1pBB7. Computational models of aberration in ultrasound breast imaging. Yi-Ting Shen and James C. Lacefield (Dept. of Elec. and Comput. Eng. and Robarts Res. Inst., Univ. of Western Ontario, London, Ontario, Canada)

Two methods are proposed for simulation of distributed aberration. One method models aberration as a superposition of five parallel time-shift screens with 8–11 irregularly shaped strongly scattering inclusions. The second method employs an anatomically realistic three-dimensional model of breast anatomy that includes lobular ducts, periductal and intralobular loose fibrous tissue, interlobular dense fibrous tissue, Cooper's ligaments, fat, and skin. Simulations of two-dimensional linear ultrasound propagation in the two model media and digitized breast tissue specimens were performed using a first-order *k*-space ultrasound simulator [Tabei *et al.*, J. Acoust. Soc. Am. **111**, 53–63 (2002)]. The initial field was a planar pulse wavefront with a 7.5-MHz center frequency and a 5-MHz −6-dB bandwidth. Propagation was computed over 25-mm paths. Both of the proposed models reproduce two characteristics of aberration observed in simulations using digitized breast specimens that are not included in conventional aberration models: non-Gaussian first-order statistics of arrival-time fluctuations and sharp changes in root-mean-square arrival-time fluctuation as a function of propagation distance. The two models, respectively, represent a relatively simple and a detailed approach in simulating realistic challenging aberration for in-silico testing of adaptive focusing techniques. [Research supported by an NSERC Discovery Grant.]

3:15

1pBB8. Time-domain three-dimensional Green's functions for power law media. James Kelly (Dept. of Elec. and Comput. Eng., Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824), Mark Meerschaert, and Robert McGough (Michigan State Univ., East Lansing, MI 48824)

Frequency-dependent loss and dispersion are typically modeled with a power law attenuation coefficient, where the power law exponent ranges from 0 to 2. Typically, these effects are modeled in the frequency domain and then time-domain results are obtained via inverse fast Fourier transform because analytical solutions in the time domain were previously not available. To address this problem, analytical three-dimensional Green's functions are derived in power law media for exponents between 0 and 2 by utilizing stable law probability distributions. For exponents equal to 0, 1/3, 1/2, 2/3, 3/2, and 2, Green's function is expressed in terms of Dirac delta, exponential, Airy, and hypergeometric functions. For exponents strictly less than 1, Green's functions are causal and expressed in terms of the Fox function. For exponents strictly greater than 1, Green's functions are also expressed in terms of Fox and Wright functions and are noncausal. For exponents equal to 1, Green's function is expressed as a stable distribution and is noncausal. However, numerical computations demonstrate that for observation points only one wavelength from the radiating source, Green's function, is effectively causal for power law exponents greater than or equal to 1.

Session 1pMU

Musical Acoustics: Dynamical Approaches in the Study of Music Perception and Performance II

Edward W. Large, Chair
Florida Atlantic Univ., Ctr. for Complex Systems, 777 Glades Rd., Boca Raton, FL 33431

Invited Papers

1:00

1pMU1. Kinematics and kinetics of music-induced movement. Petri Toiviainen (Finnish Ctr. of Excellence in Interdisciplinary Music Res., Dept. of Music, PL 35(M), Univ. of Jyvaskyla, Finland, ptoiviai@campus.jyu.fi)

Music listening is often associated with spontaneous body movements, frequently synchronized with the musical beat. While there exists an extensive body of work on synchronization of tapping, spontaneous movements to music have been investigated to a much lesser extent. The present study investigated the kinematic and kinetic aspects of spontaneous movements using a high-resolution motion-capture system. Various kinematic variables were estimated from the data, while body-segment modeling was utilized to obtain estimates of kinetic variables, such as forms of mechanical energy as well as instantaneous power produced during the movements. Although the participants produced a wide variety of movement patterns, some commonalities between them were found. On the kinematic level, it was found that musical beat was most clearly represented by movements in the vertical direction. On the kinetic level, the instantaneous internal power of the body showed clear peaks at the instants of musical beat. The results indicate that, regardless of the wide variety of spontaneous movement patterns, musical beat tends to be associated with bursts of instantaneous muscular power. This could suggest that the perception of the temporal structure of music is associated with imitation-based corporeal representations.

1:30

1pMU2. Conductors' temporal gestures: Spatio temporal cues for visually mediated synchronization. Geoff Luck (Dept. of Music, Univ. of Jyvskyl, P.O. Box 35(M), 40014 Jyvskyl, Finland, luck@cc.jyu.fi)

A growing body of research utilizies computational feature-extraction methods to quantify both music and music-related movement. The use of signal processing and visualization techniques, for example, is widespread in the music information retrieval community. At the same time, similar techniques are being applied to the analysis of movement data acquired using motion capture systems and are being increasingly used to quantify both temporal and expressive movements of musicians and others engaged in music-related activity. One such activity is conducting, and this talk will focus on one specific aspect of this activity: how a conductor conveys the beat to musicians using gesture alone. A number of studies involving a variety of response paradigms, both movement and music feature extraction techniques, and both laboratory and live settings will be presented. Results of these studies suggest that perception of the beat is less related to spatial cues than it is to temporal cues, with periods of high acceleration along the trajectory of a gesture being the best predictor of perceived beat location. In other words, conductors communicate the beat primarily by varying the speed of movement along the trajectory, not by changing the direction of movement.

Contributed Paper

2:00

1pMU3. Analysis of movements for drumstick control using surface electromyograms. Takuya Fujisawa, Naoki Iwami (Graduate School of Sci. and Technol., Ryukoku Univ., Japan), Masafumi Kinou, and Masanobu Miura (Ryukoku Univ., Japan)

Skills in controlling drumsticks correctly and appropriately in drumming are required to play rhythms without mistakes at appropriate dynamic levels of sound. Results obtained by analyzing drummers' movements from visual information, such as motion captures or camera recordings, have been reported in past studies, but analysis using biological information has been neither reported nor suggested to be employed. The aim of this study was to investigate the movements of drummers' hands and fingers by recording their surface electromyograms (EMGs) when drumming. Three amateur drummers participated in an experiment to record drummers' surface EMGs in playing single-stroke under three different conditions, such as "no-drum single-stroke," "no-drumstick single-stroke," and "normal single-stroke" at three different tempi of 80, 100, and 120 bpm. They were asked to play 16 measures of 4 beats under the tempi denoted above, and employed score in this experiment is simple, where four quarter-notes are simply allocated in a measure. Results of the experiment show that the average of surface EMG for playing no-drum single-stroke was significantly higher than that for normal single-stroke, indicating that drummers played single-stroke under lowest load in terms of muscle control. [Work supported by the HRC, Ryukoku Univ.]

2:15—2:35 Break

Invited Papers

2:35

1pMU4. Dynamics of thalamocortical circuits for sound processing revealed by magnetoencephalography. Bernhard Ross and Takako Fujioka (Rotman Res. Inst., Baycrest Ctr., Univ. of Toronto, 3560 Bathurst St., Toronto, ON M6A 2E1, Canada)

Music perception and cognition involves multi-modal processing within a wide range of neural networks working in concert. Rhythmic brain activities, or neural oscillations, are thought to play an important role in such long-range communication. How are related networks established and dynamically reconfigured in order to adapt to the ever changing auditory environment? Oscillations in the 40-Hz range (gamma band) in thalamocortical connections are proposed as a key mechanism. A common problem to delineate the behavior of 40-Hz oscillatory activity, however, is the small effect size and unknown time, courses when using noninvasive megnetoencephalography (MEG) recording. To overcome this problem, auditory stimulation with sounds containing a strong 40-Hz rhythm can be used to drive neural networks into a state of high synchrony. These areas are successfully identified by beamforming techniques, which transform MEG signals to voxel-based source images. Phase lags between primary auditory cortices and thalamus and auditory association areas suggest the information flow across the regions. Moreover, changes in the sound stimulus were observed as temporal changes of synchrony reflecting dynamic reconfiguration of neural networks. The relevance of these observations for detecting changes in sound localization will be demonstrated.

3:05

1pMU5. A hybrid model of timbre perception. Hiroko Terasawa and Jonathan Berger (CCRMA, Dept. of Music, Stanford Univ., The Knoll, 660 Lomita, Stanford, CA 94305, hiroko@ccrma.stanford.edu)

Timbre is a fundamental attribute of sound. It is important in differentiating between musical sounds, speech utterances, and characterizing everyday sounds in our environment as well as novel synthetic sounds. A hybrid model of timbre perception, which integrates the concepts of color and texture of sound, is proposed. The color of sound is described in terms of an instantaneous (or ideally timeless) spectral envelope, while the texture of a sound describes the temporal structure of the sound, as the sequential changes of color with an arbitrary range of time-scale. The computational implementation of this model represents a sound's color as the spectral envelope of a specific window, and its texture as the granularity (or microtexture) of the corresponding window. The temporal structures across windows from both color and texture parts of the model serve as the texture of a sound in a larger time-scale. In support of the proposed theory a series of psychoacoutic experiments was performed. The quantitative relationship between the spectral envelope and subjective perception of complex tones used Mel-frequency cepstral coefficients as a representation. A perceptually tested quantitative representation of texture was established using normalized echo density.

Contributed Papers

3:35

1pMU6. Auditory roughness profiles and musical tension/release patterns in a Bosnian ganga song. Pantelis Vassilakis (DePaul Univ., School of Music/Libraries, 2350 N. Kenmore Ave., JTR 207, Chicago, IL 60614) and Roger A. Kendall (Univ. of California, Los Angeles, CA 90095-1657)

Within western musical tradition, auditory roughness constitutes one of the perceptual correlates of dissonance. In a previous study [P. N. Vassilakis (2006), "The worlds of music: Culture-dependent emotional reactions to an improvisation on the mijwiz," *Proceedings of the 51stSEM:*, University of Hawaii, Manoa, HI, pp. 197– (2006) an application that calculates roughness profiles of musical pieces [P. N. Vassilakis, "SRA: An online tool for spectral and roughness analysis of sound signals," J. Acoust. Soc. Am., **120**, 3677 (2006)], based on a previously published roughness calculation model [P. N. Vassilakis, "Auditory roughness as means of musical expression," *Selected Reports in Ethnomusicology* **12**, 119–144 (2005)], was used to examine if and how tension/release patterns within a stylized improvisation on the Middle Eastern mijwiz relate to auditory roughness changes. An extension of that study, the present experiment, examines the relationship between roughness and tension/release patterns in a Bosnian ganga song. Patterns were obtained by a Bosnian ganga singer/scholar and by American-raised musicians in a perceptual experiment. Similarly to the mijwiz study, cultural background differences were associated with tension/release judgment differences as well as with differences between tension/release patterns and auditory roughness time profiles. The results provide further evidence that the concepts of musical tension and release and their relationship to auditory roughness are culture-specific.

3:50

1pMU7. Optimization of piano tunings by minimizing perceived beat loudness. David J. Carpenter and Richard L. Tutwiler (Graduate Program in Acoust., The Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, djc315@psu.edu)

In order to create an ideal piano tuning that best matches the inharmonicities and spectra of each instrument, tuning frequencies are optimized using a new beat loudness perceptual model that considers the time dependence of the frequency and amplitude of individual partials. These optimized tunings are compared to prior inharmonicity-based methods of tuning calculation. First, several pianos are prepared to an initial tuning using traditional methods. The resulting tones are recorded and analyzed in MATLAB. Each partial's frequency and amplitude are estimated at discrete time intervals over the duration of the tones. Beat rates are derived for the consonant musical intervals by employing phase-based high resolution frequency estimation techniques on coincident partials. A beat loudness model is developed based on established perceptual theories of consonance and equal loudness contours. The total perceived beat loudness is then minimized by applying optimization and global search methods that alter the tuning of each note. The resulting tunings achieve more consonance than prior tuning methods. Comparisons show subtle frequency variations in the tuning curve that adapt to each piano's unique inharmonicity, amplitude, and decay rate of individual partials.

Session 1pNS

Noise and Animal Bioacoustics: Advances in Measurement and Noise and Noise Effects on Humans and Non-Human Animals in the Environment II

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Fluid Mechanics and Eng., Einsteinufer 25, 10587 Berlin

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Chair's Introduction—1:00

Invited Papers

1:05

1pNS1. Assessing noise impacts on wildlife under the National Environmental Policy Act. Sheyna Wisdom (URS Corp., 2700 Gambell St., Ste. 200, Anchorage, AK 99503, sheyna_wisdom@urscorp.com)

Under the National Environmental Policy Act, authors must address environmental impacts of various anthropogenic actions on wildlife. One such impact of increasing awareness and concern is effect noise on wildlife, both during construction and operation of the project. However, biologists often have difficulty in understanding the fundamentals of acoustics and noise analysts often have difficulty in understanding the biological implications of increased noise on wildlife. As a result, inappropriate weighting metrics (such as A-weighted decibel) or time descriptors (e.g., community noise equivalent level) are often used erroneously to assess noise impacts on wildlife. Noise exposure thresholds on wildlife exist for marine mammals and fish, as mandated by the National Marine Fisheries Service. However, no such thresholds exist for terrestrial wildlife. This talk provides specific examples of how noise impacts on wildlife have been assessed using GIS-based technology, industry-accepted noise propagation models, and peer-reviewed literature in the absence of management guidelines. Examples include assessing construction noise impacts on the California coastal gnatcatcher in southern California, aircraft noise impacts on sage grouse in central California, and helicopter disturbance on caribou in Alaska.

1:25

1pNS2. On the need for context-sensitive noise impact criteria in rural communities. Richard Horonjeff (81 Liberty Square Rd. 20-B, Boxborough, MA 01719, rhoronjeff@comcast.net) and Grant Anderson (76 Brook Trail, Concord, MA, 01742)

With the growing population, infrastructure elements previously the exclusive province of urban and suburban communities are pushing their way into rural settings. In their wake comes increased noise nuisance and the need for context-sensitive community noise standards and guidelines. Major differences between urban and rural soundscapes make difficult the stretching of urban guidelines into rurality. Urban noise impact considerations are typically those of minimally increasing the pre-existing anthropogenic din; rural considerations center on maintaining the absence of din altogether. Most quantitative community noise standards are loudness based, comparing source loudness with that of the ambient. However, the desires of rural communities suggest that an audibility-based metric would better suit their needs. Current noise standards explicitly considering rural areas do so by recognizing the lower ambient sound levels in rural areas but still relying on loudness-based metrics. Thus, the need exists to examine the situation from an audibility perspective if places of solitude are to be preserved. This paper identifies the soundscape values of rural populations and compares them with those of urban dwellers. It then identifies the essential elements of noise metrics needed for rural soundscape preservation and also identifies the legal challenges encountered in protecting rural soundscapes.

1:45

1pNS3. Applying the Occupational Safety and Health Administration (OSHA) ultrasonic noise ceiling values. Martin Lenhardt (Dept. of Biomedical Eng., P.O Box 980168, Virginia Commonwealth Univ., Richmond, VA 23298-0168)

The Occupational Safety and Health Administration (OSHA) voted in 2003 to accept the American Conference of Governmental Industrial Hygienists (ACGIH) increased threshold level values (TLVs) for airborne ultrasound from the more stringent levels set previously as a result of ultrasonic sickness studies reported in the 1960s. The impedance mismatch between the air and the body, which prevents most ultrasonic energy absorption, was the rationale. The TLVs were increased by 30 dB, unless solid or liquid coupling is possible, allowing the unintended transfer of acoustic energy into the worker. The TLVs were set based on two lines of evidence: (1) detectability of directly coupled ultrasound and damage reported from ultrasonic exposure. The risk of ultrasonic exposure underwater is also addressed in direct coupling to an ultrasonic source. The measurements needed are airborne sound pressure level up to 100 kHz, waterborne sound pressure up to 100 kHz, and high frequency vibration. Only a 2–5 dB threshold shift for 13–17 kHz for over three years has been reported after airborne exposure. The eye was found to be a window into the skull for ultrasonic energy, reducing the impedance mismatch to 33 dB, weakening the argument that high skin impedance is an acceptable barrier.

2:05

1pNS4. Effects of sounds from seismic exploration on the calling behavior of bowhead whales. Susanna B. Blackwell (Greeneridge Sci., Inc., 1411 Firestone Rd., Goleta, CA 93117 susanna@greeneridge.com), Christopher S. S. Nations, Trent L. McDonald (WEST, Inc., Cheyenne, WY 82001), Charles R. Greene, Jr. (Greeneridge Sci., Inc., Goleta, CA 93117), Aaron Thode (Scripps Inst. of Oceanogr., La Jolla, CA 92037), and A. Michael Macrander (Shell Exploration and Production Co., Anchorage, AK 99503)

The westward migration of bowhead whales (*Balaena mysticetus*) was studied during summer and autumn of 2007 to examine the effects of airgun sounds on the whale calling behavior. Whale calls were recorded by 35 directional autonomous seafloor acoustic recorders (DASARs), placed in groups of seven recorders at five locations in the Beaufort Sea covering an east-west span of 280 km. The directional capability of DASARs allowed triangulation of an estimated whale position for about 130 000 calls. Call detection rates and call locations were compared to the timing of seismic operations—which took place in the center of the study area—and the estimated received levels of airgun sounds at the whale call locations. The analyses showed that seismic operations led to a significant decrease in call detection rates used as a proxy for calling rates. Within about 30 km of the seismic activities, received sound pressure levels from airgun pulses at call locations were greater than 140 dB *re* 1 μPa for about 20%–40% of calls. Quantile regression analyses showed that seismic activities were correlated to statistically significant shifts in the whales' distance from the shore, either offshore or inshore. [Study funded by Shell Exploration and Production Company, Alaska.]

Contributed Papers

2:25

1pNS5. The presence of infrasound in our everyday life. Kimberly Lefkowitz, Arno S. Bommer, and Robert D. Bruce (CSTI Acoust., 15835 Park Ten Pl. Ste. 105, Houston, TX 77084, kim@cstiacoustics.com)

There has been a great deal of research in the past ten years pertaining to infrasound. The effect on humans and animals of high levels of infrasound in both water and air has been studied. The effect of infrasound on structures was examined. In some situations, there has been paranoia over the effect of infrasound and, in other cases, infrasound has been overlooked completely when examining an acoustical problem. This study addresses an important element that remains to be studied, the prevalence of infrasound in a variety of locations. Infrasound was measured in a house, at a bus stop, at a typical office environment, and a variety of other situations that a typical person would be exposed to during the course of a week. These levels are then compared to less typical sites such as refineries, dredging areas, manufacturing plants, and community areas where noise complaints have been lodged. These comparisons give a preliminary understanding of peoples' exposure to infrasound.

2:40

1pNS6. Subjective evaluation of community noise in Canada's National Capital Region and its relation to waking levels of salivary biomarkers. David Michaud, Stephen Keith, Jason Tsang (Health Canada, 775 Brookfield Rd., Ottawa, ON K1A 1C1, Canada), and Anne Konkle (Environ. Health Sci. Bureau, Health Canada, Ottawa, ON K1A OK9, Canada)

Some research has suggested an association between long-term exposure to traffic noise and relative risk of cardiovascular disease (Babisch *et al.*, 2005; Willich *et al.*, 2006). It has been assumed that noise may act as a nonspecific stressor. Acute exposure to noise can evoke physiological and behavioral changes reminiscent of a stress response in rodents (Michaud *et al.*, 2003), but it is equivocal that this occurs in humans chronically exposed to traffic noise. This pilot project examined annoyance to community noise and salivary biomarkers known to be influenced by stressor exposure. A face-to-face interview subjectively assessed community noise for 60 residents (30 males, 30 females; mean age 41.3, SD=14.98). Traffic sound levels will be determined and respondents categorized into high (>65 dBA, Leq24) and low (<50 dBA) noise areas. Participants also provided saliva samples upon awakening, 30 min after awakening, and prior to bedtime. Concentrations of salivary biomarkers of alpha-amylase and cortisol were spectrophotometrically determined using commercial enzyme-linked immunosorbant assays. Two-way (high-noise versus low-noise) mixed-model analyses of variance will examine differences in questionnaire and salivary responses. Sex differences will be evaluated with independent *t*-tests, and polynomial regression analyses will relate salivary biomarker levels to high- or low-noise areas.

Session 1pPA

Physical Acoustics: Henry Bass Session: Frontiers in Acoustics

Richard Raspet, Cochair
Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

James M. Sabatier, Cochair
Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—1:25

Invited Papers

1:30

1pPA1. Early studies of vibrational relaxation phenomena. F. Douglas Shields (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, dshields@olemiss.edu)

Bass received the Silver Medal in Physical Acoustics in 2006 in recognition for the prodigious amount of work he did, largely in the area of sound absorption in the atmosphere in 1971 and 1972. When he came to Ole Miss straight from graduate school in 1970, he was 27 years old with a wife and three children. He had worked his way through college while supporting his family and served a 2 year tour of active duty in the Signal Corp before entering graduate school. He continued his service in the Army Reserve and retired in 1993 with the rank of Lieutenant Colonel. During the first three years he was at Ole Miss, he published 17 refereed papers and gave 5 papers at ASA meetings. This talk traces the path his research followed during this period in moving from his dissertation study of rotational relaxation as a function of temperature in a series of polar polyatomic gases to the successful explanation of the variation of sound absorption in the atmosphere with humidity, temperature, and frequency. Also discussed are his experiments during this period with the spectraphone and his work with Hans Bauer in developing the theory for sound amplification from controlled excitation reactions (SACER).

1:45

1pPA2. Acoustical absorption in the atmosphere at high altitudes. Louis Sutherland (Consultant in Acoust., 27803 Longhill Dr., Rancho Palos Verdes, CA 90275)

Before 1978, acoustical absorption in the atmosphere was modeled semiempirically based on numerous experimental studies, especially those of Harris and Tempest. In 1969, this author presented a semiempirical model to explain the divergence of the absorption data from theoretical values predicted from classical absorption and molecular relaxation of just moist oxygen. Joe Piercy published a correction of my error in neglecting molecular relaxation of moist nitrogen. This activity stimulated formation of an ANSI working group on atmospheric absorption chaired by Joe Piercy and included the late Henry Bass and myself as members. The result was ANSI Std. S1.25-1978 (1995) which was stated as being valid only up to 20 km. Later, I had the distinct honor to collaborate with Bass in a paper [J. Acoust. Soc. Am. **115** (2004)] which provided a computational model for atmospheric acoustical absorption up to 160 km. Unique elements of this high altitude atmospheric absorption model are reviewed, including the molecular relaxation loss from carbon dioxide. This high altitude atmospheric absorption model was possible only with the key role played by Bass in its development. The Society has lost a true scientist, scholar, and gentleman of the highest caliber.

2:00

1pPA3. Tornadoes, thunder, and underwater sound from lightning. Anthony A. Atchley (Graduate Program in Acoust., 201 Appl. Sci. Bldg., University Park, PA 16802, atchley@psu.edu)

Outdoor sound propagation was a cornerstone of Bass' research throughout much of his career. During the late 1970s, propagation of sound from natural sources became of particular interest. This talk focuses on three aspects of this phase of Bass' research: (1) efforts to localize lightning from an analysis of recordings made with a distributed ground microphone array, (2) recording the underwater sound produced by lightning striking the surface of the ocean, and (3) development of a tornado alert detector based on the acoustic signature of tornadoes.

2:15

1pPA4. Vibrational relaxation effects on rise times of weak shocks. Richard Raspet (Dept. of Phys. and Astronomy and Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

After I finished my Ph.D. on the "Application of Monte Carlo techniques in the analysis of laser induced microfission explosions," I was unemployed and trying to find a job. Hank had a contract from the Air Force to investigate the generation and propagation of thunder. As part of this work, Hank had assigned an undergraduate student the task of adapting a finite wave propagation program developed by Anderson at the University of Texas to incorporate Hank's newly developed sound attenuation algorithm for the ANSI Standard S1-26, "Sound absorption in the atmosphere." The student had graduated without making much progress and Hank offered me a summer job to work on the problem. I managed to find the single error in the program and as a result, Hank and I wrote a paper [J. Acoust. Soc. Am.]. This was the start of my career in acoustics and a lifelong collaboration with Hank. We and several co-authors published eight papers based on the study of the interaction of vibrational relaxation with finite wave distortion. This paper will present a summary of the findings of this line of research over a 24 year period.

2:30

1pPA5. The influence of Hank Bass on ground effect research between 1981 and 1995. Keith Attenborough (Dept. of Design, Development, Environment and Mater., The Open Univ., Milton Keynes MK7 6AA, UK, k.attenborough@open.ac.uk)

In 1981 Bolen and Bass published a paper on ground effect, which included a pioneering effort to deduce ground impedance from propagation data. This study was a forerunner to recent work on the ANSI standard for measuring ground impedance. My first interaction with Bass in the same year stemmed from a mututal interest in observations of buried geophone responses to airborne sound sources. Our initial experiements involved understanding the response of buried microphones to airborne sound and the relationship of ground properties to the phenomenon of ground effect. Subsequently, work with Bass and colleagues inspired treatments of the ground as a layered poroelastic medium. As a result of Bass' initiative, the first 12 Long Range Sound Propagation Symposia (LRSPS) was held in Mississippi. This helped to introduce researchers interested in atmospheric acoustics to the propagation codes used in underwater acoustics. The LRSPS initiative, together with Bass' involvement with the NATO RSG11 study group, resulted in benchmarks for testing the various atmospheric propagation codes that were available in 1995.

2:45

1pPA6. The interactions of Henry E. Bass with the National Research Council in Canada. Gilles Daigle, Tony Embleton, David Havelock, Joseph Piercy, and Michael Stinson (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

Interactions between researchers at NRC and Bass span a period of more than 30 years during which many research areas of mutual interest were explored. This paper presents an overview of Bass' research in atmospheric sound propagation during this period and the synergistic collaborations that developed with his friends and colleagues at the NRC. Early work began with the study of relaxation in gases and absorption of sound in the atmosphere. This expanded to include sound propagation above a finite impedance ground, the effects of refraction and atmospheric turbulence, and the development of fast numerical codes to predict long-range sound propagation. The close collaboration and enduring friendships that we valued so much were forged over the years through many interactions, such as the numerous joint field experiments trudging through muddy fields, working in subzero temperatures, or suffering the blistering desert heat. Throughout, Bass was a driving force, and his contributions will fuel research for the community at large for some time to come.

3:00—3:15 Break

3:15

1pPA7. Work of Hank Bass in the mid-eighties to develop a comprehensive model of sound propagation through the atmosphere. Michael White (U.S. Army ERDC/CERL, P.O. Box 9005, Champaign, IL 61826-9005)

The success of stealth technologies in the mid-eighties simultaneously raised the importance of other possible modes of discovery, one of them being acoustic detection. Bass initiated work on a comprehensive model for predicting sound transmission through the atmosphere in order to better assess the propagation. The model first considered atmospheric absorption and ground reflection of sound and would render predictions for single frequencies and spectra. Variability in received signals seemed to be explained by the strong effects of atmospheric temperature refraction, wind advection, and turbulence, but few direct comparisons of measurement to theory existed. To address this, Hank formed several collaborations using ray-tracing, fast-field, and parabolic equation methods for modeling refraction and advection, leading to benchmarks for outdoor sound propagation in 1993. Measurements of the short-term distributions of phase and amplitude over a few hundred meter distance revealed separate scattering regimes, with each being accessible by theory and numerical computation.

3:30

1pPA8. Coupling of airborne sound into the earth. James Sabatier (NCPA, Univ. of Mississippi, University, MS 38677)

While a student at the University of Southwestern Louisiana in 1982, I was attracted to the Physical Acoustics Research Group at the University of Mississippi after reading a list of physical acoustics publications left by Professor Gordon Baird on a graduate student recruitment trip. I made a brief visit to UM and met Hank, Bolen, and Crum. Hank offered me an assistantship ten times greater than the one at USL. I recall him saying the U.S. Army had a seismic sensor to detect tanks, but when helicopters flew overhead, the sensors

alarmed. Aspects of this work were described in the paper, of Bass and Bolen ["Coupling of airborne sound into the earth: Frequency dependence," J. Acoust. Soc. Am. **67**, (1980)]. During graduate years at UM, I met Keith Attenborough and was introduced to the work of Keith, Hank, and Lee on microphones buried in soils and rigid frame porous media acoustics. Under their direction, my graduate acoustics education was spent applying Biot poroelasticity to first foot of the ground. In this talk, I will summarize the acoustic-to-seismic work that was done with Hank.

3:45

1pPA9. Hank Bass and the origins of the National Center for Physical Acoustics. Lawrence Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The National Center for Physical Acoustics was established by an Act of Congress due principally to the efforts of Henry Bass. This presentation will review the history behind the NCPA and the various roles that Hank played in making this center a reality.

4:00

1pPA10. Optoacoustics in liquids and gasses. Stanley Cheyne (Dept. of Phys. Astronomy, Hampden-Sydney College, Hampden-Sydney, VA 23943)

Optoacoustics, the production of sound from light, was studied experimentally and theoretically during the 1980s at the University of Mississippi. Bass's primary interest in optoacoustics was to understand the physical processes at the molecular level. In one paper [Ali *et al.*, "Spectrophone measurements in sulfur hexafluoride," IEEE **UFFC-33**, 615 (1986)], a detailed analysis of the energy transfer processes from IR laser radiation to the dissipation of the generated sound wave was investigated. In a second paper [Thompson *et al.*, "Optoacoustic observation of internal relaxation in liquid CS_2," J. Acoust. Soc. Am. **85**, 2405 (1989)], it was shown that a liquid with relatively slow relaxation times affected the temporal shape of the optoacoustic signal. Finally, in a third paper [S. A. Cheyne and H. E. Bass, "Observation of optoacoustic amplitude in CS_2 at high input energies," J. Acoust. Soc. Am. **88**, 1842 (1990)], it was shown that the amplitude of the optoacoustic signal changed nonlinearly with input energy due to molecular interactions.

4:15

1pPA11. The wave propagation in porous material: A continuum from Biot solids to thermoacoustic heat engines. William Arnott (Dept. of Phys., Univ. of Nevada Reno, MS 220, Reno, NV 89557, patarnott@gmail.com)

Wave propagation in poroelastic materials has applications ranging from ground motion induced by atmospheric sound to thermoacoustic heat engines. Bass was involved with research in this entire continuum, including both basic and applied problems. This paper will especially cover the wave propagation formulation of thermoacoustics as it was applied by Bass and others to understand and predict the behavior of thermoacoustic prime movers. Some recollections of Bass as a statesman, scientist, and enjoyable character to be with will be presented. My first meeting with Bass and some of his guidance will also be discussed.

4:30

1pPA12. Henry Bass' contributions to the infrasound renaissance: Notes from the field. Milton Garces (Infrasound Lab., Univ. of Hawaii at Manoa, 73-4460 Queen Kaahumanu Hwy., Kailua Kona, HI 96740-2638, milton@isla.hawaii.edu)

The deployment of the international monitoring system global infrasound network at the turn of the 21st century inspired a renaissance in innovation, development, and application of infrasound technology. In the United States, Hank Bass was responsible for defining and directing the scientific agenda, and he skillfully navigated the turbulent waters of national and international policy through a decade of administration changes. In addition, he created and nurtured a cohesive community of collaborating partners in academia, industry, and government. This presentation honors Hank's contributions to the field of infrasound within the context of past, present, and future research efforts in the U.S. and abroad.

Session 1pPP

Psychological and Physiological Acoustics: Potpourri (Poster Session)

Rahul Shrivastav, Chair
Univ of Florida, Communication Sciences and Disorders, 336 Dauer Hall, Gainesville, FL 32611-7420

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbers papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1:30

1pPP1. Nonzero threshold loudness: A circular argument in Zwislocki (1965) and Moore, Glasberg, and Baer (1997). Lance Nizami (1312 Grayson Pl., Decatur, GA 30030, nizamii2@aol.com)

ANSI S3.4-2005 updates the calculation of loudness to reflect empirical evidence that loudness is nonzero at detection threshold. Hellman [Acoustics Today (2007)], in reviewing the changes, cited the work of Zwislocki [*Handbook of Mathematical Psychology* (1965) Vol. III] and Moore *et al.* [J. Audio. Eng. Soc. **45** (1997)] as theoretical support. Zwislocki proposed that detection threshold reflects an internal noise that acts like an external masker, and that (using generalized notation here for clarity) the tone+noise loudness for an rms tone pressure x is a function f of the sum of noise contribution(s) c and a tone contribution $g(x)$. The Zwislocki $g(x)$ was zero at $x=0$ and increased monotonically thereafter. Altogether, tone+noise loudness is $f(g(x)+c)$. Zwislocki then assumed that listeners can perceptually separate tone from noise. He subtracted noise loudness from tone+noise loudness to get tone loudness, $[f(g(x)+c)-f(c)]$. That exceeds zero for whatever x is deemed the tone-detection threshold. Unfortunately, with $g(x)>0$ for $x>0$, a nonzero threshold-tone-loudness was predetermined. Moore *et al.*, using auditory filter output power as x, produced a congruent equation for tone loudness in quiet, repeating the circularity. Circularity was inevitable, because detection threshold is defined using a percentage-correct performance that indicates nonconstant loudness.

1pPP2. Tone-glide direction identification revisited. Lawrence Feth and Evelyn Hoglund (Speech and Hearing Sci., Ohio State Univ., Columbus, OH 43210)

Dawson and Feth (2004) reported that listeners showed an apparent difference in sensitivity for the direction of frequency change when listening to virtual frequency glides or frequency modulated (FM) tones. Their work was a partial replication and extension of the work by Gordon and Poeppel (2002) using a one-interval fixed-block paradigm. In each interval, the listener was asked to report whether the frequency was rising or falling using buttons labeled up and down. Psychometric functions were reported but the original data were not examined for interval response bias. This study addresses possible bias in two ways. The original data were reexamined for response bias, and direction identification was tested using FM glides in the single-interval adjustment-matrix (SIAM) procedure. The SIAM procedure [Kaernbach, (1990)] is designed to control response bias in a one-interval paradigm. Contrary to the previous work, data collected with the SIAM procedure do not show a consistent pattern of sensitivity toward change in either direction. Additionally, psychometric functions using this procedure are steeper than those of the prior studies. Implications of these results in comparison with previous work will be discussed. [Work supported by a grant from NIH/NIDCD R01-DC006879].

1pPP3. Informational masking of virtual frequency glides. Gayla Poling and Lawrence Feth (Dept. of Speech Hearing Sci., The Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210, poling.75@osu.edu)

Informational masking (IM) experiments use multiple-burst maskers to change masker uncertainty. The effect of IM is believed to be mediated by interference in the central auditory processing of the signal. Virtual frequency (VF) glides are produced by amplitude modulating two sinusoids separated by one or more ERB at a given test frequency. The amplitude of the low-frequency tone is initially ΔI dB more intense than the higher-frequency tone, and then falls linearly as the amplitude of the higher frequency rises to be ΔI dB greater at the end. Listeners report hearing a weak rising pitch for such signals although no signal energy moves through the frequencies between the tones. Our hypothesis is that central auditory system processing of these signals is similar to that for frequency-modulated (FM) sinusoids, despite clear differences in their peripheral excitation patterns. This study was designed to investigate the effects of masker uncertainty on dynamic signals (FM and VF glides) for different ERB separations for the anchor frequencies. Preliminary results demonstrate a smaller release from IM for the VF glides, which may be explained by the salience of the VF signals compared to the FM tones. [Work supported by NIH/NIDCD RO1-DC006879.]

1pPP4. The effects of off-frequency listening on the pitch of narrowband complexes. Veronica Eckstein-Lilley and Bruce Berg (Dept. of Cognit. Sci. Univ. of California Irvine, 3151 Social Sci. Plaza, Irvine, CA 92697, veckstei@uci.edu)

A pitch-matching paradigm was used to investigate the filtering properties of the auditory system. Listeners adjusted the frequency of a pure tone to match the pitch of a five-tone complex. Each complex had five equal amplitude sinusoidal components with equal frequency spacing (20 Hz). Five of six total conditions constituted a signal added in phase to any one of the five components, while the sixth condition added no signal. Two different signal levels were tested. Pitch matches were best described by the envelope weighted average of the instantaneous frequency, but only if the stimulus was first passed through the skirt of an auditory filter [V. Eckstein and B. G. Berg, J. Acoust. Soc. Am. **120**, 3126 (2006)]. The results appear to be more consistent than those previously obtained with three-tone complexes. By using additional components and adding a signal at the edges of the signal complex, this study depicts a more complete inspection of the auditory filter shape, reduces pitch ambiguity, and eliminates nonlinearities. [Work supported by NSF.]

1pPP5. An acoustic comparison of nonlinguistic sounds to sentences spoken in American English. Corine Bickley (Gallaudet Univ., 800 Florida Ave. NE, Washington, DC 20002, corine.bickley@gallaudet.edu) and Yell Inverso (Pennsylvania College of Optometry, Elkins Park, PA 19027)

Acoustic characteristics of nonlinguistic (nonspeech) sounds (NLSs) were measured for duration and spectral variation, and compared to acoustic characteristics of spoken sentences (TIMIT database). The NLS, included

samples produced by animal, human, mechanical, and natural sources. The acoustic comparison examined stoplike onsets, fricativelike intervals, vowel-like intervals, and syllablelike variations in amplitude. The NLSs were identified by two groups of listeners: listeners with normal hearing and users of cochlear implants. Results of the listening tests have been reported previously by Inverso *et al.* (2007) and Inverso (2008). All of the NLSs were identified accurately by listeners with normal hearing, but not by the users of cochlear implants. The current analysis focuses on the ways in which the NLSs are similar to and other ways in which they are different from sentences spoken by a variety of talkers. It was found that speechlike variation in amplitude, both in terms of duration and event onset/offset, was a strong cue for listeners with cochlear implants; that is, NLSs that contained distinct events that were similar in duration and amplitude variation to syllables in speech were identified more accurately than ones that did not.

1pPP6. Statistical inference in the perceptual learning of non-native speech category. Yuan Zhao (Dept. of Linguist., Stanford Univ., Stanford, CA 94305, yuanzhao@stanford.edu)

The study investigated the influence of category bias and category variance in English speakers' learning of two lexical tone categories (high versus low). In Experiment 1, all subjects completed an *AX* discrimination pretest of adjacent tone pairs on a ten-step pitch continuum. They were then trained to associate a "low tone" with the meaning "pink" and a "high tone" with the meaning "blue" via a picture-sound association task. Subjects received either a biased training with a 5:1 ratio between high and low tones or an unbiased training with a 1:1 ratio between the two categories. It was then followed by an identification post-test. The results showed that poor pitch discriminators categorized significantly more sounds into the biased category. Good discriminators categorized significantly more sounds into the other category. Therefore, good discriminators have a novelty preference, while poor discriminators have a familiarity preference. In Experiment 2, subjects received either a small or a big variance training. The results showed that poor discriminators performed significantly better if trained with a big-variance input. However, good discriminators' categorization performance was significantly better if trained with an input of a small variance. The study suggests that statistical cues are utilized differently among subjects of different auditory sensitivities.

1pPP7. Acoustic discrimination and categorization of formant dynamics for novel and familiar stimuli. Daniel Fogerty (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, dfogerty@indiana.edu)

Perception of a formant transition continuum from /b/ to /d/ was tested for a high-fidelity speech continuum and an analogous sinewave speech continuum. Endpoint stimuli were recorded by a male talker, matched for fundamental frequency and duration, and interpolated using a STRAIGHT morphing algorithm to produce a continuum of 11 steps. The first three formants were replaced with sinewaves to create a sinewave speech continuum. This study used a between-subject paradigm, which enabled comparison of highly familiar natural stimulus categories to novel sinewave stimulus categories while preserving fundamental acoustic cues for perception (the first three formants). Categorization training and discrimination blocks were interleaved during testing to track the effect of category learning on discrimination. Results for sinewave speech failed to demonstrate significant short-term category learning or a change in discrimination performance across three blocks. However, discrimination and categorization performance were fundamentally different between continua. While discrimination for natural speech was best near the category boundary, participants who heard sinewave speech appeared to use detailed differences in pattern dynamics, resulting in best discrimination near specific stimuli characteristic of the dynamic pattern. Results will be discussed as they relate to categorical perception and category learning. [Work supported by the NIH.]

1pPP8. Measuring listening effort independent of task difficulty. Mini Shrivastav, Rahul Shrivastav, and James Harnsberger (336 Dauer Hall, Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611, mnarendr@csd.ufl.edu)

The effects of task difficulty on many listening tasks are often confounded by those of the effort expended during the task. It is important to have valid measures of effort independent of task difficulty. The present experiment was aimed at doing so by measuring various markers of effort for listening conditions of varying difficulty but with equal performance. Task difficulty in a tone-detection experiment will be manipulated by changing the level of a broadband noise added to the background so that the signal-to-noise ratio changes from favorable to unfavorable. The participants will be a group of young normal-hearing individuals. Effort will be measured using several tools such as self-reported questionnaires and various physiological markers such as galvanic skin response, heart rate, and blood oxygen level. For each background noise condition, the percentage of correct detection performance will be calculated. Conditions that result in greater than 90% accuracy will be considered as having equal performance. Any difference in physiological measures between these conditions will then likely reflect differences in listening effort. The study will provide preliminary candidates for developing an objective and real-time measure of listening effort.

1pPP9. A functional model of the slope sensitivity of the primary auditory cortex neurons in awake cats. Kenji Ozawa, Yoshikazu Koike, Hiromi Wakagi, Yu Sato, and Sohei Chimoto (Univ. of Yamanashi, 4-3-11 Takeda, Kofu 400-8511, Japan, ozawa@yamanashi.ac.jp)

Our measurement of single-spike responses to tone-burst stimuli showed that there are heterogenic response types of the primary auditory cortex (A1) neurons in awake cats: phasic neurons sensitive to stimulus amplitude-rise-slope at the onset of stimuli, sustained cells sensitive to the stimulus slope, and the steady state level at later time-phase of stimuli. In this study, in order to explain the behavior of the slope-sensitive phasic cells, an auditory model has been developed. The model consists of six stages: the inner hair cell (IHC), the primary auditory nerve (AN), the cochlear nucleus (CN), the inferior colliculus (IC), the medial geniculate body (MGB), and the A1 neuron. The model was implemented as a MATLAB program on a personal computer using Meddis's model for IHC [R. Meddis, J. Acoust. Soc. Am. **79**, 702-711 (1986)] and Maki's model for AN, CN, and IC [Maki *et al.*, J. Acoust. Soc. Jpn. **60**, 304-313 (2000)], while the models of MGB and A1 neuron were newly developed. The model output successfully replicated the physiological data in terms of response rise-time, maximum driven rate, and onset-latency corresponding to the slopes of input stimuli. [Work supported by KAKENHI 20300076.]

1pPP10. The sound field reproduction method based on the spacial covariances and its reproduction efficiency. Hiroki Hagiwara, Yoshinori Takahashi (Kogakuin Univ., 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan, hagiwara@b09.itscom.net), Mikio Tohyama (Waseda Univ., Tokyo 169-8050, Japan), and Kazunori Miyoshi (Kogakuin Univ., Tokyo 163-8677, Japan)

This paper describes a sound field reproduction method based on the spatial covariances. Wave surface control and convoluting the head related transfer function are major techniques for sound field reproduction. On the contrary, Takahashi and Tohyama, coauthors of this report, considered that the spatial impressions of a sound field, which we perceive through our ears, are reproduced by preserving the relative relationship between the observation points even if the wave surface is not completely controlled. Moreover, they proposed a new sound field reproduction method that can control the point-to-point covariance in a sound field [19th ICA, RBA15-012]. In this work, we demonstrated about the differences between the covariance based method and the conventional methods, such as stereophonic playback and boundary surface control based on the Kirchhoff–Helmholtz equation. Then, we discussed the performances of several methods on the point of view of reproductivity and device cost.

1pPP11. Improvement of scale factor for time-domain audio watermarking based on low-frequency amplitude modification. Harumi Murata, Akio Ogihara, Motoi Iwata, and Akira Shiozaki (Dept. of Comput. Sci. and Intelligent Systems, Osaka Prefecture Univ., 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan, murata_h@ch.cs.osakafu-u.ac.jp)

The objective of this work is to improve the sound quality for the audio watermark method based on amplitude modification. We improve the sound quality by modifying the scaling curve smoothly. The audio watermarking method based on amplitude modification has been proposed by Lie as a prevention technique against copyright infringement. The watermark information is embedded into audio signals in the time domain. One-bit watermark

information is embedded by modifying the differences of average-of-absolute-amplitude from three sections in a group of samples. At the section boundary points, the scaling factors are forced to 1.0 to ensure continuity. The factors are then progressively increased or decreased to the stable value near the center of section. However, when the adjacent sections were both increased or both decreased, signal discontinuities happen in this boundary point and cause the "click" sounds that are perceivable to human ears. In this paper, we aim to make the audio waveform continuous and smooth with the scaling factors of the section boundary points that are not set to be 1.0 but are determined according to the stable values of the adjacent sections. By the proposed method, sound quality can be improved in objective and subjective evaluations.

1pPP12. Applying acoustic ray tracing from the full wave equation solution to determine the nonair conduction pathways into the human head. Jared McNew, Alessandro Bellina (Dept. of Elec. and Comput. Eng., Univ. of Illinois, B428C Beckman Inst., 405 N. Mathews, Urbana, IL 61801), and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

Flight deck crews are subject to high intensity sounds for prolonged periods of time. Even while using current hearing protection devices such as ear plugs, ear muffs, and helmets, hearing loss is still possible. It is likely that this hearing loss is due to the cochlear stimulation through pathways other than the usual air conduction pathways. In order to develop new hearing protection devices to prevent hearing loss due to these nonair conduction pathways, ray tracing from the full wave equation solution is being used to determine and visualize these dominant pathways. Numerical techniques for performing ray tracing are presented along with the validation of the method by comparing with Snell's law. For small changes in sound speed, angle of transmission errors are less than 20%. As the frequency of the incident pulse is increased, these errors are reduced as would be expected. Results of ray tracing performed on simulation data from two concentric fluid spheres as an approximation to the human head are also presented. Using this model, an insertion loss of 26 dB is observed. [Work supported by AFOSR FA9550-06-0128.]

1pPP13. Simulation of bone-conducted sound pathways to the inner ear. William O'Brien (Bioacoustics Res. Lab., Dept. of ECE, Univ. of Illinois, Urbana, IL 61801)

A project to determine bone-conducted sound pathways to the inner ear is sponsored by the Air Force in an effort to reduce hearing loss in personnel working in high-noise environments. The study includes a computer simulation, in three dimensional (3D), of an acoustic pulse wave propagating through and around a human skull. Numerical 3D simulations are benchmarked with known analytic results in order to verify accuracy. The pulse center frequency can be varied in order to identify how frequencies affect the vibrations of the skull. The program is used to validate an experimental mapping of the skull in which the response in the inner ear, with and without hearing protection, is measured as function of transducer input location at different external head locations and center frequency. An FETD algorithm written in MPI and executed on highly parallelized clusters is used to achieve these results efficiently. The mesh, on which the code operates, is a uniform grid of eight node brick elements. Thus the program can directly operate on any digitized image (in two dimensional) or a list of digitized images (in 3D) in which each image is a single slice of a 3D volume. [Work supported by AFOSR FA9550-06-0128.]

1pPP14. Numerical simulation of mechanisms of sound conduction through the human head with a fast elasto-acoustics integral equation solver. Elizabeth Bleszynski, Marek Bleszynski, and Thomas Jaroszewicz (Monopole Res.,739 Calle Sequoia, Thousand Oaks, CA 9160)

The purpose of this paper is to describe our progress in development and applications of numerical simulation tools designed to (a) investigate mechanisms of energy transfer to cochlea through air- and bone-conduction pathways and to (b) assess effectiveness of suitable noise protection devices. Our numerical simulations employ a recently implemented acoustoelastic integral equation solver, capable (through the use of nonlossy fast Fourier transform based matrix compression algorithm and parallelization on distributed-memory systems) of accurate large-scale numerical simulations with anatomically realistic models of the human head, discretized with several millions of tetrahedral elements. Recently added new solver features (including treatment of shear waves, described in terms of node-based linear elements) enable us to analyze detailed aspects of acoustic/elastic wave interaction with the human head with significant accuracy and efficiency. In particular, they facilitate examination of the relative amounts of energy transferred to the middle and inner ears directly through the outer ear and indirectly via excitation of elastic waves in the skull and soft tissues. [This work is supported by AFOSR.]

1pPP15. Influence of static force on bone conduction threshold measurement. Lynn Brault, Woojae Han, Charissa Lansing, Ron Chambers (Dept. of Speech Hearing Sci., Univ. of Illinois at Urbana-Champaign, 901 S. Sixth St., Champaign, IL 61820), Jared McNew, Alessandro Bellina, and William D. O'Brien (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The influence of static force on bone conduction threshold measurement behavioral bone conduction thresholds are evaluated in clinical settings using a standard headband. The standard headband is designed to apply 5.4 N of static force to the oscillator (ANSI S3.6-1996 R2004), but in practice the static force varies with head size over a range of several Newtons. The influence of the varied force on behavioral thresholds is uncertain. These thresholds have clinical utility in determining the type and severity of hearing loss and research utility in modeling bone-conducted sound pathways. Behavioral bone conduction thresholds for a group of normal-hearing young adult listeners (ages 18–30) were obtained in 1-dB steps as a function of several force levels (2 N, 5 N, variable force) using custom-calibrated and standard headbands for mastoid and forehead oscillator placements across a frequency range from 250 to 8 kHz in sixth-octave bands. No significant differences in behavioral thresholds were obtained as a function of the static force levels tested, although the results supported previous findings that bone conduction thresholds are more sensitive for mastoid than forehead placement. The impact of static force levels on performance for speech understanding with bone conduction listening systems requires further study. [Work supported by AFOSR FA9550-06-0128.]

Session 1pSC

Speech Communication: Perception of Words, Sentences, and Indexical Information (Poster Session)

Sarah H. Ferguson, Chair
Univ. of Kansas, Dept. Speech-Language Hearing Sciences and Disorders, 1000 Sunnyside Ave., Lawrence, KS 66045

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbers papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1pSC1. Subjective ratings of sentences in clear versus conversational speech. Sarah Hargus Ferguson and Emily E. Kerr (Dole Ctr., Dept. of Speech-Lang.-Hearing: Sci. and Disord., Univ. of Kansas, 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045)

Investigations of the relationship between speech acoustics and intelligibility face a significant methodological challenge, especially when using materials produced by multiple talkers and/or in multiple speaking styles. Ideally, acoustic analyses comparing talkers or styles will use identical materials for each. In contrast, intelligibility studies using meaningful stimuli must use either different materials or different listener groups to avoid learning effects. The latter solution becomes prohibitive when the number of talkers is large, such as in the Ferguson clear speech database (2004). While talkers recorded different materials in each speaking style, materials were the same for all talkers. For word intelligibility, familiarization prior to testing can prevent learning effects. For sentences, however, different listeners would be needed for each talker, resulting in 41 listener groups. An alternative solution is to use subjective rather than objective measures. This project explored the feasibility of this approach. Young listeners with normal hearing rated the clarity of sentences produced by eight talkers from the Ferguson database. Rated clarity was significantly higher for clear versus conversational sentences and varied significantly among talkers. Further analyses will compare clarity to perceptual and acoustic measures such as vowel intelligibility, speaking rate, and vowel space characteristics.

1pSC2. Acoustic correlates of speaker discrimination in English. Chandan Narayan (Inst. for Res. in Cognit. Sci. Univ. of Penn., 3401 Walnut St. Ste. 400A, Philadelphia, PA 19104) and Jiahong Yuan (Univ. of Pennsylvania, Philadelphia, PA 19104)

This study investigated speaker discrimination in utterances varying in syllable length and speaker gender taken from the TIMIT corpus of American English. Twenty native English speakers presented one-, two-, and three-syllable utterances (within speaker gender) in a two-alternative forced-choice task. Perception results were analyzed in light of both source level ($F0$ and long-term average spectrum of LPC residuals) and formant level measurements ($F1$–$F4$). Results showed that male speakers were discriminated better than female speakers. Source features ($F0$ and LTAS of LPC residuals) significantly predicted listener response, while higher spectral information ($F1$–$F4$) had little effect. The varying importance of vocal source and vocal tract characteristics in speaker discrimination is discussed.

1pSC3. Effects of talker and token variability on perceptual learning of dialect categories. John K. Pate and Cynthia G. Clopper (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210, jpate@ling.osu.edu)

Dialect classification is difficult for naive listeners, but perceptual learning tasks using sentence-length utterances have been shown to produce modest improvements in performance. The goal of the current study was to explore perceptual learning by naive listeners in a speeded dialect classification task with shorter word-length utterances. In a series of experiments, participants were trained in a two-alternative forced-choice dialect classification task (Cleveland versus Cincinnati) with feedback and were then tested in the same task with novel talkers and novel words without feedback to assess learning. Variability in the stimulus materials in the training phase, including the number of talkers from each dialect and the number of different tokens produced by each talker, was manipulated across experiments to determine how variation in the input affected perceptual learning of dialect categories. The results revealed that training materials consisting of utterances produced by multiple different talkers from each dialect with multiple different tokens produced by each talker led to a significant improvement in dialect classification performance compared to a baseline condition without training. These findings suggest that dialect classification performance can improve with training on short word-length utterances, but that robust dialect category learning requires high variability stimulus materials.

1pSC4. Channel segregation improves perception of speech with temporally desynchronized bands. Michael Kiefte (School of Human Commun. Disord., Dalhousie Univ., Halifax, NS B3H 1R2, Canada, mkiefte@dal.ca), Christian E. Stilp, and Keith R. Kluender (Univ. of Wisconsin, Madison, WI 53706)

Stilp *et al.* [J. Acoust. Soc. Am. **122**, 2971 (2007)] investigated the intelligibility of temporally desynchronized bands of speech and concluded that listeners's resilience to temporal asynchrony may possibly be due to differences in fundamental frequency (f_0) across syllables which may help in segregating bands. This hypothesis is tested in two experiments in temporally desynchronized bands but with spectral manipulations designed to either aid or inhibit stream segregation. Seven-syllable sentences were synthesized at three different speaking rates and processed by four nonoverlapping 1/3-octave filters. Onsets of the lowest- and highest-frequency bands were parametrically delayed. In Experiment 1, f_0 in delayed bands was uniformly elevated using pitch-synchronous overlap add synthesis. In the control condition (no f_0 manipulation), intelligibility was nonmonotonic with delay across speaking rates with local minima corresponding to the duration of one syllable. The two-speaker manipulation decreased spectral similarity across bands making intelligibility more resilient to temporal distortion. In Experiment 2, f_0 contours were flattened. Intelligibility was uniformly poorer than in the control condition; the increased spectral similarity compromised listeners' ability to segregate information from band pairs at various delays. Acoustic measures of potential information, absent explicit linguistic information, reinforce the strong relationship between spectral predictability and intelligibility. Overall, results provide support that channel segregation plays an important role in the perception of temporally desynchronized bands. [Supported by NIDCD]

1pSC5. Interaction between speech coding and semantic predictability in speech perception. Nirmal Srinivasan and Thomas D. Carrell (Dept. of Special Education and Commun. Disord., Barkley Memorial Ctr., Univ. of Nebraska-Lincoln, Lincoln, NE 68583, nirmal@bigred.unl.edu)

The present experiments investigated how humans process speech presented in the form low-passed natural sentences (5000 Hz), cell phone coded sentences, and computer synthesized sentences. Spin sentences [Kalikow *et al.*, (1977) and Bilger *et al.*, (1984)] were presented in a background of multispeaker babble using these three coding representations. Word accuracy and visual-motor task performance were measured. Intelligibility was based

on the participants' accuracy in repeating the final word of the sentence. Attention was based on the participants' performance on the simultaneous pursuit rotor task [Srinivasan and Carrell (2007)]. In the pursuit rotor task, a stylus was used to maneuver a cursor on a computer display with the goal of keeping the cursor aligned as closely as possible to a moving target. Word accuracy was higher for predictable sentences than for unpredictable sentences for all three encodings. However, pursuit rotor performance showed a complex set of interactions. For example, participants performed better on the pursuit rotor task for unpredictable sentences than predictable sentences, but only for the cell phone speech not for natural or synthetic speech. The findings suggest that the perception of cell phone speech does not degrade the same way that natural or synthetic speech does.

1pSC6. Effect of level of presentation on scaled speech intelligibility of speakers with dysarthriafurther data. Yunjung Kim (Dept. of Commun. Sci. and Disord., Louisiana State Univ., Baton Rouge, LA 70803), Gary Weismer, Raymond Kent (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, Madison, WI 53705), and Joseph Duffy (Mayo Clinic, Rochester, MN)

This study seeks to examine the effect of level of presentation on scaling of speech intelligibility of speakers with dysarthria. Previous data [Kim *et al.* (2007)] reported the somewhat surprising result that different presentation levels do not cause a significant change in speech intelligibility scores of speakers with dysarthria, but do for healthy speakers. The current study reports further data using the same speech stimuli, however, with a main focus on the across-conditions effect of presentation level, since the previous data were limited to within-conditions effects. Four conditions of level of presentation were generated by varying absolute level of presentation (high and low) and normalizing each utterance for peak vowel intensity (adjusted and nonadjusted). In this presentation, we will report how scaled speech intelligibility is affected by the level of presentation of speech samples produced by speakers with various types and severity of dysarthria and by healthy speakers. The results of this study will be helpful to understand the rationale of speech therapy for dysarthria, especially considering the frequent use of speech intelligibility and speech therapy techniques designed based on intensity variation of speech. [Work supported by NIH DC00319 and internal fund from Louisiana State University.]

1pSC7. Evaluation of two voice stress analyzers. Harry Hollien and James Harnsberger (IASCP, 46 Dauer Hall, Univ. of Florida, Gainesville, FL 32611)

The purpose of this study was to evaluate two commonly used voice stress analyzers: NITV's computer voice stress analyzer (CVSA) and Nemesysco's layered voice analysis (LVA) system. In both cases, a speech database was used, which contained materials recorded (1) in the laboratory, while highly controlled deceptive and shock induced stress levels were systematically varied and (2) during a field procedure. Subjects were 24 males and females (age range 18–63 years) drawn from a population representative of the United States. All held strong views on an issue and were required to make sharply derogatory statements about it. The systems were then evaluated in a double blind study using two sets of examiners: (1) two UF scientists trained/certified by the manufacturers and (2) either three experienced CVSA operators or two LVA instructors provided by the manufacturer(s). The results for both devices showed that the "true positive" (or hit) rates ranged from chance to somewhat higher levels—50% to 65 %—for all conditions/types of materials (stressed-unstressed, truth, or deception). However, the false positive rate was just as high—often higher. Sensitivity statistics demonstrated that these systems operated at about chance levels. [Work supported by Counterintelligence Field Agency, DoD.]

1pSC8. Getting phonetic structure from an aphonetic signal. Joanna H. Lowenstein and Susan Nittrouer (Dept. of Otolaryngol.–Head and Neck Surgery, Ohio State Univ., 4331 Cramblett, 456 W. 10th Ave., Columbus, OH 43210, lowenstein.6@osu.edu)

Several studies have shown that adults can use a global structure such as that found in sine wave and amplitude envelope signals for speech perception. Recent work has demonstrated that children can perceive speech processed as sine wave signals as well as the adults, but have more difficulty perceiving vocoded signals. If speech perception can occur even when signals are abstracted down to the level of sine wave speech, why then does the human vocal tract produce a signal with such complexity? Perhaps the redundancy encoded in the complexity of the speech signal is useful for other tasks, such as phonemic awareness and short-term recall. This study presented three lists of eight near-rhyming words in sine wave speech, speech in 0-dB noise, and unprocessed speech to adults and 8-year olds as a short-term serial recall task. Adults performed the same on the unprocessed speech and speech in noise, and performed more poorly on the sine wave speech. 8-year olds, on the other hand, performed identically in all conditions, but made more errors overall than the adults. Adults who were not able to understand sine wave speech had poorer overall recall scores on the other two tasks. [Work supported by NIDCD Grant No. DC-00633.]

1pSC9. Effects of acoustic transformation on cross-modal speech information and audiovisual gain II. James W. Dias and Lorin Lachs (Dept. of Psych., California State Univ. Fresno, 2576 E. San Ramon ST11, Fresno, CA 93740)

Auditory and visual perceptual processes interact during the identification of speech sounds. Some evaluations of this interaction have utilized a comparison of performance on audio and audiovisual word recognition tasks. A measure derived from these data, *R*, can be used as an index of the perceptual gain due to multisensory stimulation relative to unimodal stimulation. Recent evidence has indicated that cross-modal relationships between the acoustic and optical forms of speech stimuli exist. Furthermore, this cross-modal information may be used by the perceptual mechanisms responsible for integrating disparate sensory signals. However, little is known about the ways in which acoustic and optic signals carry cross-modal information. The present experiment manipulated the acoustic form of speech in systematic ways that selectively disrupted candidate sources of cross-modal information in the acoustic signal. Participants were then asked to perform a simple word recognition task with the transformed words in either auditory-alone or audiovisual presentation conditions. It was predicted that audiovisual gain would be relatively high for those transformations in which the relative spacing of formants was preserved but would be nonexistent for those transformations that destroy the relative spacing of formants. The results are discussed in terms of existing theories of audiovisual speech perception.

1pSC10. Channel and language effects on reaction time in a voice identification. Kyna Fasnacht and Ruth Huntley Bahr (Dept. of Commun. Sci. Disord., Univ. of South Florida, 4202 E. Fowler Ave., PCD 1017, Tampa, FL 33620)

Previous research has indicated that listeners have difficulty with bilingual speaker identification when voice samples are compared across languages [Bahr and Frisch (2002)]. Communication channel can negatively impact voice identification, but its effect on bilingual speaker identification has not been well defined. Reaction time has also been found to increase for speech processing when speaker accent is strong [Schmid and Yeni-Komshian (1999)]. Therefore, it is the purpose of this investigation to assess the effect of channel differences and language to determine how these factors influence recognition accuracy and processing time. A paired comparison listening task was used to evaluate monolingual listeners' performance when channel (laboratory quality, landline phone, and cell phone transmissions) and language (English versus various dialects of Spanish) were varied. Reaction time was recorded to investigate changes in processing demands. Results indicated that listeners were less accurate when channel conditions varied. Difficulties with language/dialect were more noticeable in conversation. These findings were evident in both the accuracy and reaction time data. The role of dialectal variation in the forensic speaker identification process will be described and related to findings from the stereotype literature, which describes "an other race effect," i.e., all dialects of Spanish sound the same.

1pSC11. Using eye-tracking and long-term repetition-priming to examine talker-effects in spoken word recognition. Conor T. McLennan (Dept. of Psych.-CB175, Cleveland State Univ., 2121 Euclid Ave., Cleveland, OH 44115, c.mclennan@csuohio.edu) and Paul A. Luce (Univ. at Buffalo, SUNY, Buffalo, NY 14260)

Our research examines the circumstances in which talker variability affects listeners' perception of spoken words. In our previous work, we used the long-term repetition-priming paradigm in which listeners were presented

with two blocks of spoken words. Some words were repeated from one block to the next and some words heard during the second block had not been presented during the first block. We found that repeated words were responded to more quickly than new words (a repetition-priming effect). Moreover, repeated words spoken by the same talker in both blocks were responded to more quickly than repeated words spoken by different talkers (a talker-effect). Crucially, talker-effects emerged relatively late during perceptual processing. In the current study, we extended our previous work by using the eye-tracking paradigm and by investigating whether an attentional manipulation would modulate the role that talker variability plays during the perception of spoken words. Because eye-tracking has been shown to tap into processing quite early, we hypothesized that talker-effects would only be obtained if listeners' attention were drawn to the talkers during the first block. Results will add to our knowledge of the circumstances in which talker variability affects the perception of spoken words. This research was supported in part by research Grant No. 5 R03 DC 7316-4 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

1pSC12. Alternative data analysis techniques when using the eye-tracking paradigm to investigate spoken word recognition. Conor T. McLennan (Dept. of Psych., Cleveland State Univ., CB175 Cleveland, OH 44115, c.mclennan@csuohio.edu)

Researchers have been using the eye-tracking paradigm as a tool for investigating the representations and processes involved in listeners' perception of spoken words for over a decade now (thanks in large part to the work by Tanenhaus and colleagues). Moreover, this paradigm is becoming increasingly widespread, with the number eye-tracking researchers, laboratories and publications increasing rather substantially over the past few years. Although the eye-tracking paradigm has been used to investigate a number of different issues related to spoken word recognition, including lexical competition, parallel activation in bilinguals, and ambiguity resolution, most (if not all) of these studies use participants' eye fixations as the unit of measurement when performing statistical analyses (e.g., fixations on a picture that corresponds to the spoken word heard during a trial). Fixations are a perfectly fine source of data; indeed we have learned a lot about spoken word recognition from studies performing analyses of fixation data. However, there are different ways to use fixations, as well as various alternative data analysis techniques, that researchers using the eye-tracking paradigm to investigate spoken word recognition may wish to consider. [Work was supported in part by the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (Grant No. 5 R03 DC 7316-4).]

1pSC13. Investigating lexical influences on the accuracy of speechreading words presented in isolation and in sentence context. Edward Auer, Jr. and Rebecca Reed (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, auer@ku.edu)

Isolated spoken words that are perceptually similar to many other words in the mental lexicon are typically more difficult to recognize than words that are perceptually unique. This effect has been demonstrated for spoken words presented in both auditory and visual presentation conditions. Here, this effect of perceptual similarity on visual recognition (speechreading) of spoken words was compared for words presented in isolation and in sentential contexts. Sentences were developed containing key words that varied in predicted perceptual uniqueness. 123 (41-easy, 41-moderate, 41-difficult) spoken sentences were presented visual-alone to 40 participants for identification. 301 keywords spoken in isolation were presented visual-alone to 23 participants for identification. In both presentation conditions, identification accuracy varied as a function of the perceptual similarity of the target word to other words in English. Specifically, identification accuracy decreased as the number of perceptually similar words increased. Furthermore, the effect of perceptual similarity interacted with presentation condition (isolated words versus sentence). The results suggest that the typical advantage afforded by perceptual uniqueness is reduced when words are presented in a sentence context. Implications for future modeling studies and for predicting the intelligibility of sentence length materials will be discussed. [Work supported by an NIH/NIDCD R01DC04856.]

1pSC14. Examining the relationship between lexical access and the perceived strength of foreign-accents. Amee P. Shah (Dept. of Health Sci., Cleveland State Univ., 2121 Euclid Ave., MC 423, Cleveland, OH 44115, a.shah101@csuohio.edu) and Conor T. McLennan (Cleveland State Univ., Cleveland, OH 44115)

Foreign-accented speech is one source of variability that listeners may face during the perception of spoken language. Although a number of previous studies have examined the role that foreign-accented speech plays in listeners' ability to process spoken language, we are interested in examining the converse of this relationship. That is, previous work has demonstrated that accent manipulations affect listeners' ability to process spoken language. For example, the speed and accuracy with which listeners access spoken words produced with a foreign-accent is compromised (at least initially) relative to the same words produced with a native-accent. However, the present study, which builds on our previous work [Shah and McLennan (2007)], examines whether manipulations known to affect listeners' ability to access spoken words will affect listeners' subjective judgment of foreign-accented speech. More specifically, because spoken words are typically easier to process when they have been heard recently (repetition priming), we investigated whether listeners would respond more quickly in an accent-rating task to primed words relative to unprimed words and/or whether primed words would be perceived as having been produced with a weaker foreign-accent. Results provide new information regarding the relationship between lexical access and the perception of foreign-accented speech.

1pSC15. Effects of noise, cognitive demand and lexical characteristics on word recognition by normal hearing native listeners. Astrid Zerla Doty, Catherine L. Rogers (Dept. of Commun. Sci. and Disord., Univ. of South Florida, 4202 E. Fowler Ave., PCD1017, Tampa, FL 33620), and Judith Becker Bryant (Univ. of South Florida, Tampa, FL 33620)

Although listeners can adapt to many challenging listening conditions, the combined effects of multiple co-occurring challenges on speech understanding are less well known. The present study examines the combined effects of background noise, cognitive demand, and lexical characteristics on isolated word recognition. Young normal-hearing native English speakers heard six lists of 24 words, each composed of 12 lexically "easy" and 12 lexically "hard" target words in an open-set word identification task. Easy words had high target word frequency, few phonological neighbors, and low-neighborhood frequency; conversely, hard words had low frequency, many neighbors, and high-neighborhood frequency. Word lists were presented in quiet, in a moderate degree of background noise, and with or without a competing digit recall task. In the digit recall task, listeners saw three or six digits on the monitor prior to presentation of the word list and were asked to recall the digits at the end of the word recognition task. Word recognition scores and the size of the "easy" word benefit will be compared in quiet, in noise, and with and without the digit recall task. The accuracy of digit recall will also be compared across conditions of background noise and number of digits presented.

1pSC16. Time course of talker-specific learning in spoken word recognition. Lynne C. Nygaard, Sabrina K. Sidaras, and Jessica E. D. Alexander (Dept. of Psych., Emory Univ., Atlanta, GA 30322, lnygaar@emory.edu)

The current study examined the effects of talker-specific perceptual learning on the time course of spoken word processing to assess the point at which effects of talker familiarity emerge during spoken word recognition. Listeners learned to identify six talkers' voices (three males, three females) over three days of training. At test, listeners either completed an immediate or a delayed word-shadowing task. Items at test were novel words produced by the six familiar talkers heard during training and by a set of six unfamiliar talkers. A separate group of controls completed the test phase only. The results indicated that effects of talker familiarity were found in both the immediate and delayed shadowing tasks. Listeners were faster to shadow words produced by familiar than unfamiliar talkers both when responding immediately to the word and when responding after a cued delay. When

tested with unfamiliar talkers, the trained listeners did not differ from untrained controls. These findings suggest that effects of talker familiarity emerge relatively early in spoken word processing and persist as word recognition unfolds. Perceptual learning of talker-specific characteristics appears to influence the time course of lexical processing by facilitating the rapid recovery of the linguistic structure of spoken words.

1pSC17. Representation of multiple variant forms in spoken word recognition: Friends, not enemies. Larissa Ranbom, Eleni Pinnow, and Cynthia Connine (Dept. of Psych., Binghamton Univ., P.O. Box 6000, Binghamton, NY 13902, lranbom1@binghamton.edu)

Prior research [Ranbom and Connine (2007)] showed that alternative variant forms of a word are represented in the lexicon as a function of variant frequency. Two experiments investigated whether these multiple representations conspire or compete during spoken word recognition of nasal flap variants. In experiment 1, nasal flap productions of high (i.e., "counter") and low (i.e., "enter") variant frequency words, as well as a set of nonword controls (i.e., "penter"), were presented in a phoneme monitoring task (detection of a /t/ sound). The results showed "higher-t" detection rates for words than for nonwords, and higher-t detection for high than for low-variant frequency words. Experiment 2 used truncated versions of the stimuli from experiment 1 and showed comparable t-detection rates across all conditions. The results suggest that coactivated multiple representations of phonological variants conspire during spoken word recognition.

1pSC18. Effects of lexical status within phonetic categories. Sam Soleimany and James Sawusch (Dept. of Psych., Univ. at Buffalo, Park Hall, Buffalo, NY 14260-4110, sss26@buffalo.edu)

Previous work has shown an interaction between lexical knowledge and phonetic categorization. For example, Ganong (1980) showed that ambiguous initial phonemes were identified so that the resulting syllable was a word. While previous studies have emphasized lexical influences at the phonetic category boundary and their time-course, the present work examines lexical influence effects within categories. Changes in within category discrimination for word and nonword tokens were examined with an ABX discrimination paradigms. Signal detection analysis was employed to determine whether changes in rating responses within a phonetic category were attributable to changes in sensitivity, bias, or both. While changes in identification and discrimination between phonetic categories may be attributable to post-perceptual biases, changes within categories may not be. Results will be discussed in regard to their impact on theories of word recognition and whether lexical influences are prelexical, postlexical, or both.

1pSC19. Speech perception in a language-trained chimpanzee (Pan troglodytes). Lisa Heimbauer, Michael Beran, and Michael Owren (Lang. Res. Ctr., Georgia State Univ., 3401 Panthersville Rd., Decatur, GA 30034, lheimbauer1@student.gsu.edu)

After decades of research, the question of whether humans perceive spoken language using a specialized "speech mode" remains unresolved. Studies in nonhumans suggest that animals perceive phonemic contrasts much as humans do, but involve subjects trained for thousands of trials on single discriminations. This work reports initial speech perception results from Panzee, a chimpanzee (*Pan troglodytes*) reared by humans speaking to her as they would to a child and also training her to use graphical wordlike "lexigrams." Panzee comprehends approximately 126 spoken words, documented through a procedure in which a digitally presented spoken word is matched to one of four lexigrams presented on a monitor. First experiments have compared performance with natural digitized versions of 24 spoken words to synthetic LPC-based replicas and to whispered versions. Using a different subset of eight test words within each of three 96-trial sessions showed comparable mean performance for natural (83.3%), synthesized (82.5%), and whispered (78.5%) versions. Percent-correct performance on 24 trials representing the first time a given test word was heard was also comparable for synthesized (79.2%) and whispered (78.5%) sounds. The possibility that Panzee is showing speech-mode perception will be tested in experiments with noise-vocoded and sine-wave speech. [Work supported by NICHD.]

Note: Payment of additional registration fee required to attend tutorial.

MONDAY EVENING, 10 NOVEMBER 2008 LEGENDS 8, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Aircraft Noise Prediction

Lily M. Wang, Chair
Univ. of Nebraska, Lincoln, Architectural Engineering, 1110 S. 67th St., Omaha, NE 68182-0681

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Aircraft noise prediction. Joe Posey (NASA Langley Res. Ctr, M.S. 461, Hampton, VA 23681, joe.w.posey@nasa.gov)

Noise has been an issue for airport communities and passengers since the advent of commercial air transport almost a century ago. Impressive gains have been made in aircraft noise control, but expectations are rising and air traffic will at least double in the next 20 years. Also, lighter composite structures pose a challenge for interior noise control. Furthermore, societal expectations for mobility, new technology, demands for carbon footprint minimization, and other environmental imperatives will lead to revolutionary aircraft designs in the future. This tutorial lecture will consider noise from present and future subsonic jet aircraft, rotorcraft, propeller aircraft, and supersonic transports. Noise prediction capabilities will be identified, along with an overview of noise control technology. Acousticians are preparing to predict and control community and interior noise for arbitrary configurations using models based more on first principles, whereas the state-of-the-art is largely semiempirical. It is imperative that acousticians be included in the early stages of the design process and for all design team members to have some exposure to noise control principles to avoid wasting resources on designs that should be nonstarters from the noise perspective.

Session 2aAAa

Architectural Acoustics: Fostering Productive Architect—Acoustical Consultant Relationships

Scott D. Pfeiffer, Cochair
Threshold Acoustics LLC, 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604

Molly K. Norris, Cochair
Threshold Acoustics LLC, 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604

Chair's Introduction—8:30

This session features a panel discussion with three architects who have worked extensively in the area of performing arts, as well as other areas of architecture which have found the integration of the input of acoustics consultants beneficial. They offer three different perspectives on productive architect-acoustician relationships and what collaboration techniques they have found successful on a variety of large and small scale projects. The panel will be made up of Alan Joslin of Epstein Joslin Architects in Cambridge, Larry Kearns of Wheeler Kearns Architects in Chicago and Mark Sexton of Kruek + Sexton Architects in Chicago.

Contributed Papers

8:35

2aAAa1. An architect's perspective on acoustics. Gary Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611, gsiebein@siebeinacoustic.com), Adam Bettcher, Joshua Fisher, Threcia Robinson, and Reece Skelton (Sebein Assoc., Inc., Gainesville, FL 32607)

The basis for conceptual thinking in architecture is simultaneously aesthetic, humanistic, and scientific. An architectural theory was developed that includes philosophical, aesthetic, and technical thinking about acoustics as generative ideas for form. This theory has its roots in defining an acoustic landscape or soundscape: both interior and exterior, where the soundscape is formed by the architectural ideas and the soundscape reciprocally defines the architecture. The ensuing dialogical exploration of form, space, material, occupation, and phenomenological qualities provides a basis for the mutual expression or poesis of architectural and acoustic designers. The theory was used as the basis for graduate studio design classes at the University of Florida and is presented through the work of several students.

8:50

2aAAa2. Bypassing metrics and designing holistically use of site-specific auralizations. Joshua Cushner (Arup Acoust., 155 Ave. of the Americas, New York, NY 10013)

Acoustics as a discipline has developed many objective metrics to define subjective acceptance criteria. The most common metrics used by acousticians tend to be generalized in order to communicate effectively with clients and architects at the expense of accuracy. These generic criteria can be given undue importance on a particular project when a more site-specific investigation may indicate the criteria to be inappropriate. Simply stated, standard acoustic criteria can serve to create an artificial barrier to otherwise effective solutions and strain the consultant-architect and consultant-client relationship. However, resistance from deviating from standard criteria is understandable when few nonacousticians (and even some acousticians) can appreciate the qualitative aspects of a given criteria. Providing site-specific auralizations can serve to bridge the communication gap with architects and clients by moving from a criteria-based design to a performance based design. Through the process of site evaluation and design auralization, analyses can be provided which place the client or architect directly in the subject environment, where they are able to make value-based judgments, irrespective of generic criteria. Results from recent project studies show this approach to be useful to enhance communication with the architect and provide the acoustician with another useful analysis tool.

9:05—9:10 Break

9:10—10:10 Panel Discussion

Session 2aAAb

Architectural Acoustics: Acoustical Challenges of Implementing Curtain Wall Construction

Scott D. Pfeiffer, Cochair
Threshold Acoustics LLC, 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604

Molly K. Morris, Cochair
Threshold Acoustics LLC, 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604

Chair's Introduction—10:25

Invited Papers

10:30

2aAAb1. Spertus Institute of Jewish Studies: Sound isolation of unique curtain wall design. Polyana Frangetto (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, pfrangetto@kirkegaard.com)

The case study describes the challenges encountered in designing and building the curtain wall window system of the new Spertus Institute—ten-story building in Chicago, IL, designed by Krueck + Sexton Architects, opened in November 2007. The multifaceted glass facade, overlooking Michigan Avenue into Grant Park and Lake Michigan, is formed by complex folds in three dimensions and supported by a customized structural system. Noise isolation to the urban environment and between diverse program spaces along the building's glass front such as exhibition galleries, library, administrative offices, multimedia classrooms, museum, children's center, board room, space for community events and celebrations, and café is a high priority. Innovative design details were developed to preserve the vertical sound isolation from floor to floor along the jagged slab edges and to address the horizontal sound isolation between the sloped panes of glass. This paper presents in detail the program functions, design considerations, options explored, and test results of the final solution.

10:50

2aAAb2. Luxury isolation: Curtain wall construction and the impact on the expectation of acoustic isolation. Carl Giegold and Molly Norris (Threshold Acoust., 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604, cgiegold@thresholdacoustics.com)

Curtain wall construction is an aesthetically popular choice among architects for exterior glazing. It is often employed in luxury settings where the expectation of acoustic isolation is high, but the ability to achieve it is low. This is particularly an issue with regard to room-to-room isolation at curtain wall mullions. The author will present a case study involving the use of curtain wall construction in a high rise luxury hotel and including the issues with warranty and an alternative approach to improving the performance of these elements.

Contributed Papers

11:10

2aAAb3. Methods for reducing flanking airborne noise transmission through mullions of curtain wall systems. John LoVerde and Wayland Dong (Veneklasen Assoc., 1711 16th St., Santa Monica, CA 90404, jloverde@veneklasen.com)

Over the course of design of high-rise structures incorporating curtain wall design, there is a common limitation of the sound isolation that can be obtained at the intersection of demising partitions at the curtain wall. Approximately 45 years ago, Veneklasen Associates (VA) developed details to control the isolation. Through field testing and observations, this detailing was optimized. Approximately 15 years ago, VA constructed the intersection details in a laboratory to confirm the results from our experience. In the past 15 years, these details have further been advanced to create sound isolation in line with the desire of the user, including multifamily residential dwellings. VA shall present the historical detailing and modifications to the detailing and correlate this information with subjective and objective assessments of these constructions.

11:25

2aAAb4. Descriptors of aural connectivity: Architectural enclosure features and acoustical qualities. Selen Okcu, Craig Zimring (College of Architecture, GaTech, Atlanta, GA 30332-0155 gth678y@mail.gatech.edu), and Erica Ryherd (GaTech, Atlanta, GA 30332)

In intensive care unit (ICU) settings, auditory monitoring of alarm and nonalarm sounds is one of the key nursing functions. In the ICU sound environments there are many different noise sources, some producing more than one tone to communicate different conditions. In a sound environment with many different sound layers offering different sound qualities and sound levels nurses are expected to distinguish the sounds from each other and identify critical ones. Earlier studies showed that effective auditory monitoring requires recognition, localization, and immediate reaction to critical sounds to improve patient safety, nurse work efficiency, and reduce nurse burnout. This study explores the nurse task specific auditory monitoring experience by documenting the nurse auditory tasks and systematically mapping related sounds. This is analyzed as *aural connectivity*, a network measure that reflects the overall pattern of where users can recognize and

localize the critical sounds in architectural enclosures. It is hypothesized that enclosure types shape sound environments that affect user auditory experience. This paper defines the metrics for enclosures and associated acoustical qualities by looking at two ICU settings and discusses the preliminary results regarding the relationship between qualities of sound environment and architectural enclosures.

11:40

2aAAb5. Defining perceptual requirements of real-time acoustic auralizations. Linda Gedemer and Jonas Braasch (Dept. of Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

Acoustic modeling evolves from static sources and receivers to dynamic sources and receivers as computational speeds allow for faster renderings. Current research has focused on creating highly accurate dynamic models with the hope of producing accurate auralizations but still struggles with the trade-off between accuracy and available processing speed. The question remains as to how accurate do these models need to be if the receiver (listener) moves dynamically through the acoustic model. The focus of this paper is to address the accuracy required of dynamic models and/or auralizations from a perceptual standpoint. From this, the necessity for creating models that are perhaps more accurate than required will be considered.

TUESDAY MORNING, 11 NOVEMBER 2008 LEGENDS 2, 8:00 TO 11:50 A.M.

Session 2aAB

Animal Bioacoustics: Marine Mammal Acoustics in Honor of Sam Ridgway I

Whitlow W. L. Au, Chair
Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

Chair's Introduction—8:00

Invited Papers

8:05

2aAB1. Why honor Dr. Sam H. Ridgway. William Evans (Dept. of Biological Sci., Univ. of Notre Dame, Notre Dame, IN 46556, evans1930@sbcglobal.net)

Based on my 45 year association with Ridgway, I have summarized the accomplishment of this outstanding scholar and scientist. I discuss the transition from a clinical veterinarian to an accomplished researcher and innovator. Our understanding of animal bioacoustics, especially marine mammals, has been greatly influenced by his research. This is not only true in the U.S. but also in the international community. This includes the development and use of specialized techniques in advancing our understanding of the dolphin bioacoustics. His development and use of innovative methods in understanding the dolphin sound production and the impact of high level noise on dolphin hearing are significant contributions.

8:15

2aAB2. Anatomical, electrophysiological, and histological studies of dolphin auditory system: Establishment of a theory of hearing for dolphins. James McCormick (Dept. Anesthesiology, Wake Forest Univ. School of Medicine, Winston-Salem, NC 27157-1009, dolphin-flt-int@triad.rr.com)

Wever and I at Princeton University wanted to study hearing in dolphins utilizing Wever's cochlear-potential recording method, which directly reflects the performance of the ear's mechanical system. A method of humane anesthesia for dolphins for electrophysiological studies did not exist. Preliminary studies by Ridgway with halothane anesthesia published in 1965 seemed promising and motivated me and Wever to collaborate with Ridgway. Subsequently, Ridgway and I tested and perfected the use of halothane anesthesia for dolphins. This led to a coast to coast collaboration on dolphin hearing between Ridgway at Naval Missile Center Point Mugu and me and Wever at Princeton. Ridgway shared his expertise in Marine Mammal Medicine not only with us, but with the world, and in the process he became an expert marine mammal neuroscientist. Our three way collaboration demonstrated that the dolphin hears by bone conduction, with the middle ear ossicles acting as an inertial mass relative to the sound received in the lower jaw of the dolphin—the external auditory meatus and ear drum being vestigial.

8:35

2aAB3. Sam's early work on the sea turtle. Richard Fay (Parmly Hearing Inst., Loyola Univ. Chicago, Chicago, IL 60626)

I first met Sam Ridgway in 1967 or 1968 before he had become "the dolphin doctor." He was a visiting scientist at the E.G. Wever's Laboratory at Princeton, NJ and I was a new Ph.D. student. He had already done the classic experiments on the dolphin ear with Wever, McCormick, and Palin (1970, 1971, 1972, and 1974), never to be repeated in the same way. He was there to study hearing in the sea turtle (*Chelonia mydas*) using cochlear potentials. This was also the last of these kinds of experiments on these charismatic animals. Wever was most interested at the time in the reptile ear and had a swimming pool installed at the laboratory for them. The publication that resulted was that of Ridgway *et al.* ["Hearing in the giant sea turtle, *Chelonia mydas*," Proc. Natl. Acad. Sci. U.S.A. **64**, 884–890 (1969)]. It stands as the best information we have on sea turtle hearing, with an indication of sensitivity and a frequency response function. The only better information would be a behavioral audiogram, which is yet to be done. This early chapter in Sam's career is vitally important today with respect to the setting of sound exposure criteria for fish and turtles.

8:55

2aAB4. Auditory-visual equivalence learning by marine mammals. Ronald J. Schusterman, Kristy Lindemann, and Colleen Reichmuth (Long Marine Lab., Univ.of California, 100 Shaffer Rd., Santa Cruz, CA 95060)

Forty years ago Sam Ridgway began pioneering research on the brain and sensory systems of marine mammals. Our paper honoring Sam reviews studies on the abilities of dolphins and California sea lions to integrate and classify information coming from different sensory channels in cross-modal matching-to-sample tasks. Following echoic recognition of an object, a dolphin that is permitted to investigate objects visually but not echoically can readily learn and remember that earlier reflected sound cues are related to the currently reflected light cues. A California sea lion has shown that it can also integrate auditory-visual information into something resembling equivalence classes. In these studies the sea lion was taught to relate an array of individual auditory signals (e.g., "ringing bell" and "wailing siren") to several individual members of two previously established visual equivalence classes ("numbers" and "letters"). Following the training phase it was later determined that the sea lion had knowledge of untrained transitive relationships between the auditory signals and the remaining members of the two visual equivalence classes. Such findings suggest that these marine mammals have neural pathways connecting auditory and visual associative brain areas in the construction of long lasting integrated or global representations.

9:15

2aAB5. Dolphin detection of targets as a function of angular presentation and target strength. Patrick Moore (U.S. Navy Marine Mammal Program, SPAWARSYSCEN San Diego, Code 71520, 53560 Hull St., San Diego, CA 92152), Lois Dankiewicz (SAIC, San Diego, CA 92110), David Kastak (Univ. of California, Santa Cruz, Santa Cruz, CA 95060), and Dorian Houser (Biomimetica, Santee, CA 92071)

As part of an ongoing study, bottlenose dolphin (*Tursiops truncatus*) echolocation beam steering was assessed as a dolphin performed a target detection task (4- and 7.62-cm spheres) while echolocating with its head held in a fixed orientation and monitored via underwater camera. Targets were placed approximately 4.0 m in front of the dolphin [zero degree position (*P*0)] and at 34 deg to the left and right of *P*0 (PL and PR). Echolocation clicks and echoes were digitized using a high-speed 32 channel analog-to-digital card, sampling 29 calibrated hydrophones with analog filter-amplifiers arranged in a diamond-shaped array supported in a hemispherical web in front of the animal. Using a yes-no response and a modified method of constants procedure with 0.50 target presence/absence probability, detection ability (d) and response bias (B) of the dolphin were computed for both targets at all positions. Detection thresholds (d) for the 4- and 7.62-cm spheres at *P*0, PL, and PR were 4.65, 2.08, and 2.62, and 4.08, 2.81, and 3.2, respectively. Beam characteristics were computed and beam steering was verified for this dolphin. These results also show angular detection at displacements about 8 deg farther from *P*0 than previous reports for spheres half as large.

9:35

2aAB6. Using auditory evoked potentials to measure marine mammal hearing. Paul Nachtigall (Marine Mammal Res. Program, Hawaii Inst. of Marine Biology, Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734, nachtiga@hawaii.edu)

The use of auditory evoked potential (AEP) measurements for examining the hearing of marine mammals was accelerated by the need to rapidly test hearing following sound exposures during temporary threshold shift experiments. AEP measures have allowed quick measurements of hearing following intense noise exposures of sounds ranging from filtered white noise to 53-C sonars. The refinement of AEP hearing measurements with envelope following response measures leads to comparisons between behavioral and AEP hearing thresholds. These comparisons have allowed an expansion of basic hearing measurements including measures of new species such as the white-beaked dolphin and polar bear, measures of increased numbers of the same species in order to better examine population audiogram estimates and variability, and testing of hearing of animals in the field. AEP measures have also allowed the measurement of hearing during echolocation. A false killer whale has been shown to hear its own outgoing signal 40 dB down from the same sort of signal presented directly in front of it, actively control the level of its hearing of echolocation returns in an automatic gain control of hearing, and hear sounds other than echolocation sounds 20 dB differently when it is echolocating depending on target condition.

9:55

2aAB7. Studying white whale hearing and sound production at depth: Sam H. Ridgway acoustic research in the open sea. Donald A. Carder (SPAWAR Systems Ctr. San Diego, Div. 71500, 53560 Hull St., San Diego, CA 92152-5001)

When cetacean hearing sensitivity at depth versus surface arose among anthropogenic ocean sound issues, Sam Ridgway advanced an approach using the first hearing tests of trained cetaceans in the open sea to provide an answer. Two trained white whales, *Delphinapterus leucas*, made 885 voluntary dives to a platform at depths from 5 to 300 m, stationing up to 12 min at a time to test the effects of depth on hearing and sound production. As depth increased, hearing sensitivity did not diminish, as in the aerial ears of humans and other terrestrial mammals tested in pressurized chambers. As whale hearing is unchanged by depth to at least 300 m, zones of influence on them are unaffected by depth. The whales were also trained to emit a pulse train, and then whistle to the presence of a small target. Echolocation pulses did not vary statistically in amplitude or frequency at depths between 5 and 300 m. However, the peak frequencies at the open ocean test site were at 4–40 kHz, where in San Diego Bay there were usually dual peaks at 30–80 and 100–120 kHz. Response whistles decreased in amplitude, and frequency emphasis shifted higher with a depth from 5 to 300 m.

10:15—10:30 Break

10:30

2aAB8. Temporary threshold shift: Sam Ridgway's contribution to an important and not-so-temporary line of marine mammal research. Carolyn E. Schlundt (EDO Professional Services, 3276 Rosecrans St., 1st Fl., San Diego, CA 92110, carolyn.melka@edocorp.com), James J. Finneran, and Donald A. Carder (Space and Naval Warfare Systems Ctr., San Diego, CA 92152)

When the National Research Council formed a committee in 1992 to investigate the current knowledge and need for research on the potentially harmful effects of low-frequency sound on marine mammals, Sam Ridgway embarked on a course of study that is now entering its second decade. Following the committee recommendations and with the goal of providing long-term protection of the species, Ridgway initiated the first formal investigations of temporary threshold shift (TTS) in dolphins and white whales. Key to that effort was Ridgway's development of a vocal response paradigm based on the method of free response, which allowed thresholds to be quickly estimated after noise exposure. Initial studies focused on short duration tones. Subsequent studies included single, intermittent, and impulsive sounds and have produced TTSs large enough to track growth and recovery. The resulting data reveal the effects of exposure frequency, sound pressure, duration, and temporal pattern. Recent success using auditory evoked potentials to measure TTS will enable many more subjects/species to be tested. As a result of Ridgway's pioneering work, direct data now exist to predict the effects of noise on marine mammals and to develop acoustic exposure guidelines.

Contributed Papers

10:50

2aAB9. An acoustic finite-element model to study sonar interactions with marine mammals. Gonzalo Feijoo and Kenneth Foote (Appl. Ocean Phys. and Eng. Dept., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

A computer model based on the finite-element method (FEM) is being developed to study the interaction of sonar signals with marine mammals. This model solves the Helmholtz equation in a computational box that includes the animal and the surrounding medium, water. The FEM code has been validated with analytical solutions for the scattering of a plane wave by a fluid sphere over a range of parameters and frequencies of interest. Then it has been applied to a specimen, a 142-cm-long common dolphin (*Delphinus delphis*), and internal pressure and displacement fields were computed. The animal is represented in the computer model by a set of tissue groups whose acoustic properties, density, and sound speed are taken from the literature. The geometry of each tissue group was constructed from segmented computerized tomography images. Results are presented for harmonic signals in the 1–10 kHz frequency range. [Work supported by NOPP through ONR award No. N000140710992.]

11:05

2aAB10. Statistical analysis of acoustic propagation in beaked whale habitats. Steven E. Crocker, Eryn M. Wezensky, Glenn H. Mitchell (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841-1708), and Colin D. MacLeod (Univ. of Aberdeen, Aberdeen AB24 2TZ, United Kingdom)

Research and regulatory communities have raised questions about the influence of oceanographic conditions on propagation of midfrequency sound produced by naval sonars prior to beaked whale mass stranding events. Resolving these questions requires detailed study of acoustic properties at locations known to support beaked whale populations potentially at risk of exposure. To address these concerns, this study investigated sound propagation at two known beaked whale habitats: the North Atlantic Frontier to the west of Scotland and Sagami Gulf in eastern Japan, by computing the range dependent propagation loss for an omnidirectional 3.5 kHz signal with a source level of 235 dB *re* 1 μPa at 1 m. From this, a method was developed to reduce the three-dimensional sound field to yield a source-receiver range at which there was a high probability that a given received sound pressure level would not be exceeded. To investigate the uncertainty in data for sound speed profiles and seabed acoustic properties, the sensitivity of this method to typical variations of these factors was evaluated. Understanding the physical oceanographic properties affecting sound propagation can provide critical science for development of conservation management and mitigation practices.

11:20

2aAB11. Beaked whale detection utilizing characteristics of foraging echolocation clicks. Stephen Martin (SPAWAR Systems Ctr. San Diego, 53560 Hull St., San Diego, CA 92152-5001, steve.w.martin@navy.mil)

A technique is presented for the automatic acoustic detection of beaked whale (family *Ziphidae*) foraging echolocation clicks. The technique currently employs a three-stage process. The first stage is an acoustic detection screener that uses a submillisecond detection window. The second stage analyzes for click frequency modulation characteristic of two species of beaked whale (*Mesoplodon densirostris* and *Ziphius cavirostris*). The third stage performs temporal filtering to reduce false positives using information such as individual beaked whale interclick interval characteristics. Results from nine test files (90 min of data from US Navy instrumented test ranges AUTEC and SCORE) show good performance on eight of the nine test files. The ninth test file (short finned pilot whale) exhibited 930 (assumed) incorrect calls out of 9893 screener detections (9.4%). Over 70 000 min of data from the US Navy Pacific Missile Range Facility were also processed with third stage detections ranging from 3423 to 84 240 beaked whale click detections (depending on temporal processing utilized) from over 13×10^6 screener detections.

11:35

2aAB12. Ship strike acoustics: A paradox with a parametric solution. Edmund Gerstein, Joseph Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), and Steven Forsythe (U.S. Naval Undersea Warfare Ctr. Div. Newport, Newport, RI)

Marine mammals are vulnerable to ship collisions when they are near the surface. Here acoustical laws of reflection and propagation can limit their ability to hear and locate the noise from approaching vessels. Defining the physics of near surface acoustical propagation as it relates to ship noise and hearing is central to understanding and mitigating ship strikes. Field data from controlled ship passages through vertical hydrophone arrays demonstrate a confluence of acoustical factors that poses detection challenges including (i) downward refraction, (ii) spreading loss, (iii) Lloyd's mirror effect, (iv) acoustical shadowing, and (v) masking of approaching ship noise by ambient noise and distant ships. A highly directional, dual-frequency parametric sonar has been developed to mitigate these challenges and to alert marine mammals of approaching vessels. The system projector is a planar array, comprised of 45 elements, band centered to transmit a high carrier frequency along with a lower side band signal. The nonlinearity of water is used to demodulate the mixed high-frequency carrier into a lower-frequency waveform audible to both manatees and whales. The bow mounted arrays project a narrow beam directly ahead of vessels and "fill in" acoustical shadows in an effort to alert marine mammals of the approaching danger. [Work funded by the DOD Legacy Resource Management Program (Navy).]

Session 2aAO

Acoustical Oceanography and Underwater Acoustics: Attenuation Coefficient of Sediments From Low- to Mid-Frequencies

James F. Lynch, Chair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Chair's Introduction—7:55

Invited Papers

8:00

2aAO1. Sound attenuation coefficients in sandy and silty seabottoms from low- to mid-frequency field measurements in shallow water. Ji-Xun Zhou, Xue-Zhen Zhang (Georgia Inst. of Technol., Atlanta, GA 30332-0405 and Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080, China, jixun.zhou@me.gatech.edu), and Nicholas P. Chotiros (The Univ. of Texas at Austin, Austin, TX 78713-8029)

The debate on the sound speed dispersion and the frequency dependence of sound attenuation in sediments has persisted for decades, mainly due to the lack of sufficient experimental data in the low- to mid-frequency speed/attenuation transition band and the limitations of existing models. This paper analyzes and summarizes the LF measurements, conducted at 20 locations in different coastal zones around the world, that have resulted in the identification of nonlinear frequency dependence of sound attenuation in sand-silt seabottoms for a frequency range of 50–2000 Hz [Zhou and Zhang, J. Acoust. Soc. Am. **117**, 2494 (2005); **119**, 3447 (2006)]. The resultant LF sound speed and attenuation can be described equally well by the Biot-Stoll model, the Chotiros BICSQS model, and the Buckingham VGS model. However, a combination of the LF field-inverted sound speeds and attenuations with data from the SAX99 and SAX04 measurements over a broadband of 50–400 000 Hz cannot simultaneously be matched by any of these models with one set of adjustable input parameters; these models either underestimate broadband dispersion or overestimate LF attenuation. Comments on the LF field-derived data, such as "speed-attenuation coupling," are made. The possible reasons for the broadband data-model mismatch are discussed. [Work supported by the ONR.]

8:20

2aAO2. Attenuation in marine sediments: *In situ* measurements versus propagation-loss inversions. Michael J. Buckingham (Scripps Inst. Ocean., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, mbuckingham@ucsd.edu)

Acoustic attenuation in marine sediments may be measured either *in situ*, from the difference in signal level at two embedded probes, or from an inversion of transmission loss in the water column. Most *in situ* measurements are performed above 10 kHz and return an attenuation that scales as the first power of frequency, whereas transmission-loss inversions are performed at lower frequencies, returning a near-square-law variation with frequency. The difference between the *in situ* and transmission-loss estimates could be due to a combination of factors, including bottom roughness, which may lead to an enhanced attenuation from the transmission-loss inversions. The nonoverlapping frequency bands of the two techniques could also account for the differences in the estimated scalings of attenuation with frequency. Thus, the attenuation could scale roughly as the square of the frequency below a few kilohertz, transitioning to a linear dependence above 10 kHz. Such behavior would be consistent with the recently developed VGS theory of acoustic propagation in saturated granular materials: the effective viscosity of the molecularly thin layer of pore fluid between contiguous grains dominates the low-frequency square-law behavior of the attenuation and the strain hardening controls the high-frequency linear dependence on frequency. [Research supported by ONR.]

8:40

2aAO3. Attenuation of shear and compressional waves in surficial sediments. Michael Richardson (Marine Geosiences Div. Naval Res. Lab. SSC, MS 39529-5004)

Recent *in situ* measurements of shear and compressional wave attenuation in sediments provide an appropriate data set to develop empirical relationships among seafloor acoustic and physical properties. Compressional wave attenuation increases with increasing mean grain size and bulk density and with decreasing porosity, whereas the relationships among shear wave attenuation and these physical properties display the opposite trends. The ratio of shear to compressional attenuation ranges from 25 to 60, with the highest values occurring in fine-grained, high-porosity sediments. Empirical relationships defined by regressions of compressional and shear wave speeds on sediment physical properties have high confidence levels. However, confidence levels for the prediction of attenuation are poor due to a combination of measurement technique, spatial variability, and presence of strong near-surface gradients.

9:00

2aAO4. On the exponential power law for low frequency attenuation in shallow water. William M. Carey (Dept. of Mech. Engrg., College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

A review of measured transmission over sandy bottoms has found the attenuation $\alpha(f)=\alpha(f_o)\cdot(f/f_o)^n$, $1.6\leq n\leq 1.87$ for 50 Hz$<f<$1 kHz where the reference frequency f_o is nominally 1 kHz. Since plausible theory suggests that n should be equal to 2, why are observed n values different? Calculations incorporating scattering, layering, and gradients were found not to explain this observation. The discrepancy can be explained when an additional attenuation mechanism (generated lower velocity shear waves carry energy out of the waveguide) is considered. Since the intrinsic attenuation, $\alpha_i(f)$, and the apparent attenuation due to shear wave conversion, $\alpha_{\mathrm{swc}}(f)$, are small, the observed attenuation can be $\alpha(f)\approx\alpha_{\mathrm{swc}}(f)+\alpha_i(f)$. If the removal of energy by shear wave conversion is comparable to the intrinsic attenuation at the lower frequencies but less than the intrinsic attenuation at the higher frequencies, the inferred frequency dependence can be less than quadratic. Inversions with fluid sediments account for this effect by the use of a site-specific n factor that accounts for the loss of energy from the compressible field due to shear wave conversion. Furthermore the knowledge of $c_s(f,z)$ is required and may necessitate new methods to measure this property of sediments.

9:20

2aAO5. The physical mechanism (viscosity related) of low-frequency acoustic wave attenuation in sandy/silty sediments. Allan D. Pierce (Dept. of Mech. Eng., Boston Univ., Boston, MA 02215, adp@bu.edu)

Considered saturated sediments contain diverse rock pebbles (characteristic size of 0.1 mm). The weight of higher pebbles holds lower pebbles in contact sufficiently that acoustically induced solid displacements vary slowly over several adjacent pebbles. Apart from contact areas, each is nearly surrounded by water at a nearly uniform pressure. An appropriate first approximation predicts that the elastic stress tensor in the pebbles is diagonal, with components equal to the negative of the acoustic pressure in the neighboring fluid. The assumptions of Mallock and Wood apply: the mass weighted local average velocity is proportional to the negative gradient of the pressure in the water. The no-slip condition at the interfaces tends to force the water to move with the pebbles, but the finite viscosity allows the fluid at small distances from the interfaces to move at a different velocity than the pebbles. The apparent driving force for the oscillations of the interstitial water relative to the pebbles is associated with the inertia of the water and is proportional to the difference in densities. The derived approximate wave equation predicts attenuation proportional to frequency squared, proportional to the square of the difference of the densities, and inversely proportional to viscosity. The derived dimensionless proportionality constant is consistent with recent experiments.

9:40

2aAO6. Low- to mid-frequency model of attenuation and dispersion. Nicholas P. Chotiros, Marcia J. Isakson (Appl. Res. Labs., The Univ. of Texas at Austin, TX 78713-8029, chotiros@arlut.utexas.edu), Ji-Xun Zhou (Georgia Inst. of Technol., Atlanta, GA 30332-0405), and David P. Knobles (The Univ. of Texas at Austin, Austin, TX 78713-8029)

A model of attenuation and sound speed in ocean sediments in the low- to mid-frequency range is presented. It is based on a combination of the Biot–Stoll and contact squirt and shear drag model [Chotiros and Isakson, J. Acoust. Soc. Am. **116**, 2011–2022, (2004)] and the frame virtual mass model [Chotiros and Isakson, J. Acoust. Soc. Am. **121**, EL70 (2007)]. This combination can match measured sound speeds below Wood's equation lower bound and it addresses issues related to the Kramers–Kronig relationship. It is able to match both the sound speed dispersion and attenuation measurements from a large number of sites, including the sediment acoustics experiments (SAX99 and SAX04) in the Gulf of Mexico, the shallow water experiment on the Atlantic coast, and the Yellow Sea [Zhou, J. Acoust. Soc. Am. **78**, 1003–1009 (1985)]. [Work supported by ONR, Ocean Acoustics.]

10:00—10:15 Break

Contributed Papers

10:15

2aAO7. Measurements of attenuation of sound in marine sediments at low frequencies. Ross Chapman and Yong-min Jiang (Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper describes an experimental technique for measuring the attenuation of sound in marine sediments at low frequencies. The method makes use of the signals from sub-bottom reflectors that are received on a vertical hydrophone array at close ranges in shallow water environments. The signal path geometry is determined from an inversion of the travel time differences of the sea bottom and sub-bottom reflected paths to estimate the sediment sound speed and the depth of the sub-bottom reflector. The method is applied to data from experiments carried out at a site on the New Jersey continental shelf in the Shallow Water 06 experiment. The sediment type and the structure of the sediment column were ground truthed by independent measurements at the site. The chirp signal was transmitted in two frequency bands, from 100–900 and 1500–4500 Hz, and the data were match filtered to obtain the multipath signals at the array. A prominent signal from a sub-bottom interface known as the *R*-reflector was resolved in the data for each frequency band. Results are presented for the average attenuation over the depth to the reflector. [Work supported by ONR Ocean Acoustics.]

10:30

2aAO8. Wide-band attenuation measurements in New Jersey Shelf sediments. Altan Turgut (Acoust. Div., Naval Res. Lab., Washington, DC 20375, turgut@nrl.navy.mil)

The frequency dependence of attenuation in marine sediments is investigated using *in situ* measurements from a wide-band acoustic probe system during the Shallow Water 2006 (SW06) experiment. Direct-path pulse propagation measurements within the 10–80 kHz frequency band show linear frequency dependency of attenuation in dB/m at two silty-sand sites on the New Jersey Shelf. The measurements also show evidence of mild velocity dispersion within the 10–80 kHz frequency band. The measured attenuation and velocity dispersion curves are in good agreement with those predicted by an extended Biot theory for sediments with a distribution of pore sizes [Yamamoto and Turgut, J. Acoust. Soc. Am. **83**, 1744–1751 (1988)]. Attenuation roll-off measurements (dB/m kHz) were also conducted within a 3–8 kHz frequency band using a chirp-sonar subbottom profiler. The comparable levels of attenuation roll-off are observed from the chirp sonar and acoustic probe measurements. [Work supported by ONR.]

10:45

2aAO9. Laboratory measurements of sound speed and attenuation in water-saturated artificial sediments as a function of porosity. Theodore F. Argo, IV, Matthew D. Guild, Preston S. Wilson (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292), Charles Radin, Matthias Schröter, and Harry L. Swinney (Univ. of Texas at Austin, Austin, TX 78712-0264)

Several models have been proposed to describe the sound speed and attenuation of acoustic waves in water-saturated granular sediments. The difference between the predictions of competing models can be only a few percent, and it is difficult to obtain both sufficient knowledge of the sediment physical parameters and sufficiently small measurement uncertainty to adequately verify competing models. Furthermore, the inhomogeneity of natural sediments causes additional measurement and interpretation uncertainty. Therefore, well-controlled laboratory measurements are useful for model validation and also facilitate exploration of the effects of sediment physical parameter variation. A method for controlling the porosity of an ensemble of water-saturated granular particles [Schröter *et al.*, Phys. Rev. E **71**, 030301(R) (2005)] was used to prepare columns of artificial sediments, composed of monodisperse 280-μm-diameter glass spheres and distilled water of varying porosity. Acoustic time-of-flight measurements were used to determine the phase speed and attenuation of the material for frequencies from 250 to 800 kHz and for porosities from 0.37 to 0.43. The sound speed and attenuation measurements will be presented and compared to existing model predictions. [Work supported by ONR and the Robert A. Welch Foundation.]

11:00

2aAO10. Sound speed and attenuation in the sea bottom from broadband sound propagation in the Yellow Sea. Lin Wan, Ji-Xun Zhou, and Peter Rogers (School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, lin.wan@gatech.edu)

A joint China-U.S. underwater acoustics experiment was conducted in August, 1996 in 75-m water in the Yellow Sea with a very flat bottom and a strong and sharp thermocline. Broadband explosive sources were detonated both above and below the thermocline along two radial lines and a quarter of circle. Two inversion schemes are used to obtain the sound speed in the bottom. The first scheme is based on extracting normal mode depth functions from the cross-spectral density matrix (CSDM). The CSDM is constructed from explosive signals measured using a 32-element vertical line array at a fixed long range. In the second scheme, the modal arrival times are extracted using an adaptive time-frequency analysis technique. The inverted bottom sound speed is used as a constraint on high-quality TL data to extract the bottom sound attenuation. The inverted bottom attenuation exhibits apparent nonlinear frequency dependence over a frequency range from 80 to 1260 Hz. A possible effect of mode-coupling on the inverted bottom attenuation is discussed. Measured TL as a function of frequency, range, and depth is in good agreement with the predictions based on the inverted bottom acoustic parameters. [Work supported by ONR.]

11:15

2aAO11. Measuring grain roughness for the purpose of high-frequency acoustic modeling. Kevin Briggs, Allen Reed (Seafloor Sci. Branch, Naval Res. Lab., Stennis Space Ctr., MS 39529, kbriggs@nrlssc.navy.mil), Richard Ray and Michael Richardson (Naval Res. Lab., Stennis Space Ctr., MS 39529)

Grain roughness and packing are important sediment properties for newer acoustic models. We present the scanning electron stereomicroscopic imagery of natural sand grains of varying shape and roughness for evaluation of potential grain interactions of individual grains. The grain shape and the presence of microasperities will determine the probability of the contact with surrounding grains being a point or a larger area. Moreover, the use of a micro-roughness power spectrum in characterizing the sand grain roughness may be appropriate. Grain contact information is an essential starting point for developing and evaluating acoustic models that address acoustic losses at high frequencies. This information provides the basis to understand contact mechanics, such as grain slip and frame dilation, during insonification. Because media frame stiffness depends, at the grain scale, on the number and type of grain-to-grain contacts, we characterize sand sediments with imagery identifying such grain contact information. The natural shapes of grains, their variations, and their packing density present a broad distribution of grain contact types and areas. X-ray microfocus computed tomography captures volumetric images of grain interactions, which enable the grain contacts to be classified and quantified. The method and results of this evaluation with computed tomography will be also presented.

11:30

2aAO12. Measurements of sound speed and attenuation for six kinds of water saturated glass beads. K. Lee and W. Seong (Dept. of Ocean Eng., Seoul Natl. Univ., Seoul, Korea, nasalkh2@snu.ac.kr, wseong@snu.ac.kr)

Acoustic measurements of sound speed and attenuation were made in a laboratory water tank containing water saturated granular medium consisting of six kinds of glass beads each with mean grain sizes of 0.090, 0.150, 0.354, 0.507, 0.625, and 0.875 mm for several frequencies extending from 40 kHz to 1 MHz. Cross-correlation technique and time-of-flight technique are used to calculate the sound speed and attenuation of samples. In the presentation, the uncertainties of sediment preparation such as air-removal, grain packing, and depth-varying of sediment will be mainly discussed. Also, the observed patterns of the sound speed and attenuation will be presented for the mean grain size, grain size distribution, and frequency.

Session 2aMU

Musical Acoustics: Dynamical Approaches in the Study of Music Perception and Performance III

Edward W. Large, Chair
Florida Atlantic Univ., Ctr. for Complex Systems, 777 Glades Rd., Boca Raton, FL 33431

Chair's Introduction—8:30

Invited Papers

8:35

2aMU1. Dynamic encoding of pitch, timing, and timbre. Nina Kraus (Commun. Sci., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, nkraus@northwestern.edu)

The perception of music depends on three fundamental elements: pitch, timing, and timbre. It is possible to access the neural transcription of these elements objectively, noninvasively, and with great fidelity with scalp electrodes in humans. The effects of musical experience on subcortical encoding of sound are pervasive and extend beyond music to the domains of language and emotion. The acoustic cues relating to pitch, timing, and timbre have distinct representations, which can be selectively enhanced or degraded in different populations. Musically trained subjects have enhanced subcortical representation of pitch, timing, and timbre, although with increased stimulus complexity, timbre cues become more pronounced while pitch cues are diminished. Musical experience fundamentally shapes auditory processing that occurs early in the sensory processing stream. This top-down influence is likely mediated by the extensive (corticofugal) circuitry of descending efferent fibers that synapse all along the auditory pathway. Sensory shaping occurs not only with lifelong experience but can be effected by short-term training. I will discuss a series of experiments demonstrating the lifelong and short-term malleability of the neural encoding of pitch, timing, and timbre in the auditory brainstem. [Work supported by National Institutes of Health (R01DC001510) and National Science Foundation (NSF 0544846).

9:05

2aMU2. Nonlinear dynamics approach to pitch perception and harmony. Julyan H. E. Cartwright (Instituto Andaluz de Ciencias de la Tierra, CSIC-Universidad de Granada, E-18071 Granada, Spain), Diego L. Gonzalez (FSSG-CNR, I-30124 Venezia, Italy), and Oreste Piro (Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain)

For perception, the brain handles and reacts to an enormous amount of information from the senses in real time. The question of how all this information is represented and processed within the nervous system is addressed by the nonlinear and complex system researches through the proposal that dynamical attractors may form the basis of neural information processing. Sometimes, this rather qualitative approach to modeling provides very accurate quantitative predictions. Here we show an example where this idea can be successfully applied in the human auditory system to describe and, perhaps, explain the nature of some phenomena in the perception of pitch. We will also discuss possible implications of this approach to the theory of harmony.

9:35

2aMU3. Time-frequency transformation by arrays of neural oscillators: Implications for musical structure. Marc Velasco and Edward Large (Ctr. for Complex Systems, Florida Atlantic Univ., Boca Raton, FL 33431)

A growing body of evidence is consistent with the possibility of nonlinear oscillation in both the peripheral and central auditory nervous systems. This talk will introduce a model of nonlinear time-frequency transformation via an array of neural oscillators, each tuned to a distinct frequency, organized along a frequency gradient. Transformation of sound stimuli by neural oscillators is characterized. Predictions about general properties of nonlinear time-frequency transformation, such as frequency detuning and higher-order resonance, are derived. The model is consistent with nonlinear resonance approaches to pitch perception. The perception of tonality is predicted as a global pattern of resonance regions at small integer ratio frequency relationships. Neural oscillation provides a substantive potentially universal principle underlying the basic materials of music, namely, pitch and tonality. [Work supported by AFOSR FA9550-07-C0095.]

10:05—10:30 Break

10:30

2aMU4. Dynamics of tonality. Petr Janata (Ctr. for Mind and Brain, UC Davis, 267 Cousteau Pl., Davis, CA 95618, pjanata@ucdavis.edu)

As listeners of western tonal music, we establish mental representations of tonal space: the distance relationships between major and minor keys (tonal centers). Evidence for the distance relationships underlying tonal space derives from music theory, cognitive psychology, and statistical learning approaches. Tonal space is represented parsimoniously on the surface of a torus. Although the torus itself is a static structure, the movements of music about its surface are shaped by the nuances of melodies and harmonic progressions

that mold the sense of the tonal center through time. Central to an understanding how our brains interact with music's dynamic movements in tonal space, and the main aspect of this presentation, is a consideration of the timescale(s) over which the movements are modeled. At long timescales, movements on the toroidal surface are slow and perhaps tend to reflect symbolic music-theoretic descriptions in which a series of stable tonal centers is established. When biased toward shorter timescales the toroidal dynamics reveal much livelier movement that depends on the tonal content of individual chords. Intermediate timescales reveal tonal dynamics between these two extremes, both stability and rapid transitions, and perhaps reflect the scale that most closely match our sense of tonal movement.

Contributed Papers

11:00

2aMU5. Operator algebra and coherent states for equally tempered scales. Alpar Sevgen (Dept. of Phys., Bogaziçi Univ., Bebek, Istanbul 34342, Turkey, sevgena@boun.edu.tr)

The algebra that the equally tempered scales obey is $[S_z,S_\pm] = \pm c_1 S_\pm$ and $[S_+,S_-] = 0$, where all are modulo N relations, and c_1 is the generator of the scale labels. The coherent states $|\lambda,\mathbf{n}\rangle$ are eigenvectors of S_- with eigenvalues λ and where $\lambda^N = 1$. Similarities and differences between these equations and the quantum mechanical angular momentum algebra are discussed. The mean square variation δS_z^2 in a coherent state depends on M, the number of notes in the scale. δS_z^2 is minimum for $M = 8$, and the next higher δS_z^2 value belongs to the customary $M = 7$. The same conclusions hold true also for the complementary scales with $M' = N - M$ notes. Thus, one may speculate about the possibility of using $M = 8$ and $M = 4$ tones in the $N = 12$ semitone equally tempered scales in musical compositions.

11:15

2aMU6. The speech-to-song illusion. Diana Deutsch, Rachael Lapidis, and Trevor Henthorn (Dept. of Psych., Univ. of California, San Diego, La Jolla, CA 92093)

This paper reports the first formal investigation of a surprising illusion: A particular spoken phrase is made to be heard convincingly as sung rather than spoken, simply by repeating it several times over. Three groups of 18 subjects each listened to the spoken phrase repeated ten times, and they judged after each repetition whether, and how strongly, they heard the phrase as spoken or as sung. For all groups, the first and last repetitions were identical. For the first group, the intervening repetitions were also identical. For the second group, the intervening repetitions were transposed upward and downward slightly (by 2/3 semitone and 1 1/3 semitone), so that the pitches differed but the pitch relationships were preserved. For the third group, the intervening repetitions consisted of the untransposed syllables presented in jumbled orders. Comparing the first with the last repetition, a clear and highly significant transition from perceived speech to perceived song occurred when the intervening repetitions were exact, but not when the intervening repetitions were transformed. This indicates that both the exact pitches and the exact pitch contours need to be repeated for the illusion to occur. This paper is accompanied by sound demonstrations of the illusion.

TUESDAY MORNING, 11 NOVEMBER 2008 LEGENDS 1, 8:45 A.M. TO 12:00 NOON

Session 2aSC

Speech Communication: James J. Jenkins: Teacher, Mentor, Researcher

Winifred Strange, Chair
CUNY Graduate School, Speech, Lang. and Hearing, 365 Fifth Ave., New York, NY 10016

Chair's Introduction—8:45

Invited Papers

8:50

2aSC1. Hearing impairment and speech communication handicap: Deconstructing the Tower of Babel. Sigfrid Soli (Dept. of Human Commun. Sci. and Devices, House Ear Inst., 2100 W. 3rd. St., Los Angeles, CA 90057)

This presentation brings together a body of international research focused on a common problem, the effects of noise on speech communication in normally hearing and hearing impaired individuals. My research on the problem is infuenced by Jim's emphases on interdisciplinary research, measurement, and importance of a good practical problem. The challenges of objectively measuring speech communication handicap are well known and longstanding. The subject research embodies an approach to this problem based on a set of principles that provides a quantitative methodology for equivalent assessment of speech communication handicap in any language. Our approach is based on the earlier work of Stevens and Plomp, as well as Plomp's model of speech communication handicap, which expresses handicap quantitatively as the elevation of the speech reception threshold (SRT) in quiet and/or noise over that of a normally hearing individual (normals). The performance of normals provides a normative reference that is functionally equivalent across languages. Elevation of the SRT above the norm for a language defines a quantitative measure of handicap with the same meaning for each language. Our research has led to the creation of SRT-based measures of handicap in 16 languages. Examples of these measures and their use in clinical outcomes research, device development, and occupational health will be given.

9:15

2aSC2. Revisiting "Acoustic information for objects, places, and events". Valeriy Shafiro (Dept. Commun. Disord., Rush Univ. Medical Ctr., 600 S. Paulina, 1015 AAC, Chicago, IL 60612)

Over two decades ago, Jenkins proposed a broad taxonomy of acoustic information which addresses the perception of objects, places, and events in the world around the listener [J.J. Jenkins "Acoustic information for objects, places, and events," *Persistence and Change: Proceedings of the First International Conference on Event Perception*, edited by W. H. Warren and R. E. Shaw, Erlbaum, (2006)]. Although lesser known than some of his other publications, this chapter has anticipated several future directions in auditory research by asking what is being perceived by the listener and why. Inspired by this taxonomy, our research on environmental sound perception by normal-hearing and cochlear implant listeners has investigated how objects and events are perceived when sensory input is degraded. Experiments with vocoded environmental sounds of varying spectral resolution indicate large differences among individual sounds in the amount of spectral resolution required for identification. Temporally patterned sounds generally need a lower spectral resolution than harmonically rich sounds. However, with training, perception of environmental sounds can substantially improve and generalize to novel sounds, potentially including speech. Findings in cochlear implant listeners further indicate strong correlations between speech and environmental sound perception abilities, suggesting a considerable overlap in the perception of these two ecologically significant sound classes.

9:40

2aSC3. Studies of speech perception: A mentor's influence. Joanne L. Miller (Dept. of Psych., Northeastern Univ., Boston, MA 02115)

A major focus of my research program has been the nature of phonetic categorization. Many of our early studies examined the locations of boundaries between phonetic categories and the kinds of contextual factors that alter boundary locations. Later, as it became increasingly clear that phonetic categories are internally structured, with some category members more representative than others, we extended our efforts to specify the nature of this internal structure and the contextual factors that fine-tune it. Most recently, we have been investigating how these internally structured context-dependent categories might be customized for the systematic acoustic-phonetic variation in the speech of individual talkers. In this presentation, I will provide examples of our findings and drawing from studies of speech perception in adults and infants. Of particular relevance, I will point out how the many important lessons I learned in graduate school from Jim Jenkins, one of my mentors, have influenced my approach to speech research over the years. [Work supported by the NIDCD.]

10:05—10:15 Break

10:15

2aSC4. Adventures of a J. J. Jenkins student in speech technology research. Ann Syrdal (ATT Labs. Res., 180 Park Ave., Rm D159, Florham Park, NJ 07932-0971, syrdal@research.att.com)

Two of the most valuable lessons I learned from Jenkins were (1) to focus on what interests me and explore why it interests me and (2) that an experiment with unexpected results can be more valuable than one that confirms the hypothesis. My graduate student years in Minnesota at the Center for Research in Human Learning, which Jenkins directed, were an exciting and revolutionary time, both on a large societal scale and academically. Experimental psychology at Minnesota had been a bastion of behaviorism and learning theory. Jenkins was our fearless leader in advancing the revolutionary ideas and emerging paradigms of cognitive psychology and psycholinguistics. In this new field, it was now possible to be scientifically rigorous and to study mental as well as behavioral phenomena. Jenkins and co-workers who played a role in my education at Minnesota have influenced both my earlier academic research and my research in speech technology over the past 22 years. After a brief overview of how I as an experimental/cognitive/perceptual/psycholinguist contributed to several speech technology projects, my talk will focus on unit selection text-to-speech synthesis and how studying and modeling the perception of speech is central to the technology.

10:40

2aSC5. What the eyes can tell us about spoken-language comprehension. Andrea Weber (Max Planck Inst. for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, andrea.weber@mpi.nl)

As an academic great-grandchild of Jim, I had the fortune to have Jim's personal advice during the last stages of my graduate studies and during my first steps as a postdoctoral researcher. I was inspired by his great expertise, his respectful and sage comments on my work, and his encouragement for collaboration across disciplines. This paper pays tribute to Jim Jenkins by showing how his mindset influenced my own research in the field of spoken-language comprehension. In particular, I will introduce work that investigates the intersection of lower- and higher-processing levels using eye movements. Eye movements to displayed objects are tightly linked to spoken-language comprehension and are commonly used in psycholinguistic research to gain insights into underlying processes. I will present evidence for how difficulties on the perceptual level of processing can affect lexical processing and how, in turn, sentence processing can be influenced by characteristics of the lexical level, such as lexical frequency.

11:05

2aSC6. Speech perception within a biologically realistic information-theoretic framework. Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706, krkluend@wisc.edu)

Fundamental principles that govern all perception, from hair cell to cortex, are shaping our understanding of the perception of speech and other familiar sounds. Following modest presentation of historical and philosophical background, I will explain how ecological and sensorineural considerations encourage an information theoretical approach to speech perception. Information theory and

biology of sensorineural systems are emphasized in explanations of classic characteristics of speech perception, including perceptual resilience in the face of degradation, signal variability, segmentation, and categorical perception. Experimental findings will be used to illustrate how a series of like processes operate on the acoustic signal with increasing levels of sophistication on the way from wave forms to words. Common to these processes are ways that perceptual systems absorb predictable characteristics of the soundscape, from temporally local to extended periods (learning), and new information (change) is enhanced. [Work supported by NIDCD.]

11:30—12:00 Panel Discussion

TUESDAY MORNING, 11 NOVEMBER 2008 LEGENDS 12, 8:00 TO 11:55 A.M.

Session 2aSP

Signal Processing in Acoustics, Underwater Acoustics, Architectural Acoustics, and Biomedical Ultrasound/ Bioresponse to Vibration: Recent Developments in Coded Signals in Acoustics

David C. Waddington, Chair
Univ. of Salford, Acoustic Research Ctr., Newton Bldg., Salford, M5 4WT, UK

Chair's Introduction—8:00

Invited Papers

8:05

2aSP1. The use of coded signals in the measurement of diffusing surfaces. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com) and Trevor Cox (Univ. of Salford, Salford M5 4WT, UK)

Following the introduction of quantifiable number theoretic, reflection phase grating diffusing surfaces by Schroeder [J. Acoust. Soc. Am. **57**, 149–151 (1975)], several approaches have been used to measure and characterize them. In the early 1980s, time delay spectrometry, developed by Heyser, was used to measure the temporal, spatial, and spectral responses of diffusing surfaces. Once maximum length sequences were introduced as an excitation signal, they were adopted in the late 1980s. These explorations led to the creation of a measurement standard for the diffusion coefficient, which is the circular autocorrelation of the third octave polar responses. The limitations of periodic number theoretic surfaces, namely, bandwidth, grating lobes, and flat plate effects, were addressed by computer optimization, which combined the power of boundary element and multi-dimensional minimization techniques, and modulation. Both one-dimensional (phase variation in one direction) and two-dimensional (phase variation in two orthogonal directions) surfaces were characterized. The data collection and reduction procedures will be reviewed with several examples of diffusing surfaces. A summary of this research can be found in the work of Cox and D'Antonio [*Acoustic Absorbers and Diffusers: Theory, Design and Application* (Spon, 2004)].

8:25

2aSP2. Improving biomedical ultrasonic imaging systems through coded excitation and pulse compression. Michael Oelze and Jose Sanchez (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801)

Coded excitation and pulse compression techniques are being employed in biomedical ultrasonic imaging systems in order to improve image quality. Traditionally, the rationale for using coded excitation techniques in biomedical ultrasonic imaging was to increase the signal-to-noise ratio (SNR) of backscattered signals without increasing the pressure amplitudes above the thresholds predicted to produce bioeffects. Furthermore, coding schemes combined with pulse compression would allow the spatial resolution to be preserved. Recently, unique coded excitation schemes have been developed [i.e., the resolution enhancement compression (REC) technique] that not only produce a significant increase in SNR but also a doubling of the bandwidth of the ultrasonic imaging system. In conventional ultrasound *B*-mode imaging, the increased bandwidth provided by the REC technique could be used to improve the axial resolution of the ultrasonic imaging system or traded off to improve the contrast resolution of the ultrasonic imaging system through frequency compounding. In quantitative imaging techniques using spectral analysis (e.g., scatterer size imaging), the increased bandwidth from REC could be used to increase the trade-off between spatial resolution and estimate variance. An overview of these coded excitation techniques and their application to biomedical ultrasonic imaging will be presented. [This work is supported by NIH EB006741.]

8:45

2aSP3. Coded signals: A challenge for acoustic wind profiling. Sabine von Hünerbein (School of Computing, Sci. and Eng., The Univ. of Salford, Salford M5 4WT, UK, s.vonhunerbein@salford.ac.uk) and Stuart Bradley (Univ. of Auckland, Private Bag 92019, Auckland, New Zealand)

Use of coded signals for acoustic wind profiling presents special challenges because the scattering mechanism is highly bandwidth-sensitive. We outline the physical principles involved and show that the signal coding methods commonly used for underwater acoustics or for transmission in the atmosphere (such as for tomography) do not apply for scattering of sound by turbulence. The degree of detuning and loss of signal strength is examined for FM chirps, in relation to the corresponding increase in signal-to-noise ratio (SNR)

from matched filters. The optimum bandwidth does not give a large improvement in SNR for this physical situation. The alternative, and very conventional, use of stepped frequency chirps is examined in comparison. This too requires special consideration as to the length of each constant-frequency part. Finally, we give some insight into future developments using a combination of these approaches.

9:05

2aSP4. Annular-array 35-MHz chirp imaging applied to biomedical studies. Jonathan Mamou (F. L. Lizzi Ctr. for Biomedical Eng., Riverside Res. Inst., 156 William St., New York, NY 10038, mamou@rrinyc.org), Orlando Aristizábal (New York Univ. School of Medicine, New York, NY 10016), Jeffrey A. Ketterling (Riverside Res. Inst., New York, NY 10038), Ronald H. Silverman (Weill Medical College of Cornell Univ., New York, NY 10021), and Daniel H. Turnbull (New York Univ. School of Medicine, New York, NY 10016)

High-frequency ultrasound (HFU) is an attractive means of obtaining fine-resolution images of biological tissues for ophthalmologic, dermatological, and small-animal applications. Nevertheless, even with current improvements in circuit designs and high-frequency equipment, HFU suffers from two limitations. First, HFU images have a limited depth-of-field (DOF) because of the short wavelength and the low-fixed *F*-number of HFU transducers. Second, HFU is usually limited to shallow imaging because of significant attenuation in tissues. Previously, a 17-MHz five-element annular array was excited using chirp-coded signals and a synthetic-focusing algorithm was used to form images. Results demonstrated significant increase in DOF and penetration depth. In the present study, the same approach was implemented with 35-MHz five-element annular arrays. Vitreous-hemorrhage-mimicking phantom was imaged. Chirp images of a phantom showed increased sensitivity compared to conventional monocycle images and blood droplets into the phantom could be visualized 4 mm deeper. Three-dimensional datasets of 12.5-day-old-mouse-embryo heads were also acquired *in utero* using chirp and conventional excitations. Synthetically focused images were formed and the brain ventricles were segmented and reconstructed in three dimensions. Significant artifacts for the ventricle volumes were observed with the conventional imaging technique but were not apparent on the chirp-based dataset reconstruction. [Work supported by NIH Grant No. EB006509.]

9:25

2aSP5. Orthogonal coding sequences for multiple-source study of outdoor noise propagation. David Waddington and Jamie Angus (Acoust. Res. Ctr., Univ. of Salford, Salford M5 4WT, UK, d.c.waddington@salford.ac.uk)

This paper describes recent developments in the design of a coded acoustic signal for the study of outdoor sound propagation. Low signal-to-noise ratio is a common limitation in outdoor propagation investigations, and time variance of the air limits the amount of averaging. Recent work by the authors presented a coded acoustic signal approach consisting of a frequency carrier biphase-modulated by a specially designed pseudorandom code sequence to overcome these limitations. The "inner and outer" code sequence was specially designed for environmental sound propagation investigations, combining simultaneous fine time resolution and large range ambiguity, together with an ability to average and probe the propagation path. This approach is extended to multiple acoustic sources for the study of outdoor sound propagation, with the design of signals using orthogonal correlation sequences. The results illustrate that combinations of orthogonal coding and multiple carrier frequencies can permit continuous sound speed measurements in numerous vertical and horizontal directions and might therefore be useful for acoustic tomographic applications.

9:45—10:00 Break

10:00

2aSP6. Multimodal coded excitation in medical ultrasonic imaging. Yayun Wan and Emad Ebbini (Dept. of Elec. and Comput. Eng., Univ. of Minnesota, 200 Union St. SE, Minneapolis, MN 55455, wanx0028@umn.edu)

Coded excitation has been used successfully in medical ultrasound to improve the signal to noise ratio and penetration depth. Furthermore, inverse filtering techniques have been shown to improve axial resolution without significant loss in contrast due to the range sidelobe phenomenon. With the next generation ultrasound scanners, the opportunity exist for transmitting multiple codes simultaneously, with or without focusing, into the region of interest. Using appropriate 2D inverse filtering on receive, it is possible to control the axial and lateral resolutions of the imaging system while maintaining an acceptable tradeoff with image contrast. In this paper, we review recent results on 1D and 2D inverse filtering of beamformed echo data resulting from insonification of speckle-generating cyst phantoms with multiple codes. We demonstrate that (regularized) inverse filtering can be derived from the space-time array patterns. Furthermore, we demonstrate the feasibility of reconstructing multiple image lines in parallel to increase the frame rate in 2D and 3D imaging.

10:20

2aSP7. Pseudorandom sequences for binary amplitude diffusers. Konstantinos Dadiotis, Jamie A. S. Angus, and Trevor J. Cox (Acoust. Res. Ctr., Univ. of Salford, Salford M5 4WT, UK, k.dadiotis@pgr.salford.ac.uk)

Binary amplitude diffusers create a mixture of absorption and diffusion and are used to improve room acoustic conditions. They require a two dimensional binary pseudorandom array to state where absorbing and reflecting patches should be. Consider the binary array to have dimensions A×B. Ideally, *A* and *B* should be similar to maximize performance. Such an array can be produced by folding a one dimensional sequence into a two dimensional array using the Chinese Reminder Theorem. This theorem requires *A* and *B* to be coprimes. However, there is a limited set of optimal one dimensional sequences that can be generated from number theory with suitable length *AB*. In the past maximum length sequences have been used because they display desirable autocorrelation characteristics, but their period is 2^{m-1}, where *m* is an integer and so there is a limited number available. Consequently, it is necessary to examine other ways of generating optimal two dimensional arrays. This paper looks into other number theoretic sequences, and other construction techniques for forming binary arrays. The suitability of these for constructing two dimensional binary amplitude diffusers is examined. Boundary element modeling is used to evaluate their performance.

10:40

2aSP8. Coding for underwater acoustic channels. Shengli Zhou (Dept. of Elec. and Comput. Eng., Univ. of Connecticut, 371 Fairfield Rd., Storrs, CT 06269)

This talk will give an overview on channel coding, coded modulation, and their applications in underwater acoustic channels. We will cover classical block codes, convolutional codes, and the traditional coded modulation such as trellis coded modulation and bit-interleaved-coded-modulation. We will then focus on modern Turbo codes and low-density-parity-check codes, space-time coding, and highlight the recent progress on the application of the iterative (or Turbo) principle to underwater acoustic communications.

11:00

2aSP9. Pulse coding for a new scanning bistatic acoustic wind profiler. Stuart Bradley (Dept. of Phys., Univ. of Auckland, Private Bag 92019, Auckland, New Zealand, s.bradley@auckland.ac.nz) and Sabine von Hünerbein (Univ. of Salford, Salford, M5 4WT, UK)

We describe signal coding for a new design of SODAR (atmospheric acoustic RADAR) used for profiling atmospheric turbulence and winds. This bistatic SODAR uses a central vertically pointing transmitter and three synchronized phased-array scanning receivers spatially distributed on the ground and, at any instant, simultaneously pointing at the same region of the central atmospheric column. There are distinct advantages if continuous transmission can be used, since total signal power and fast profile acquisition are enhanced. However, the weak echoes mean that the receivers will record a combination of transmitted and scattered sound, so the two need to be orthogonal. This suggests a signal coding scheme. However, turbulent scattering acts similarly to an acoustic reflective diffraction grating, and this places severe constraints on signal coding. We therefore explore a sawtooth linear FMCW system, designed to optimize separation of echo and direct signal, while not detuning too much the Bragg diffraction grating relationship. An implementation of this new bistatic system and first results are discussed.

11:20

2aSP10. Time-reversed maximal-length sequences for outdoor, underwater sound propagation, and room-acoustic artificial reverberation simulations. Ning Xiang (Graduate Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180) and Dezhang Chu (NOAA/NMFS/NWFSC, Seattle, WA 98112)

Pseudorandom properties of maximal-length sequences (M-sequences) have long been exploited for acoustic measurements. Time-reversed M-sequence pairs also possess unique correlation properties, including a two-valued autocorrelation function and a low-valued cross-correlation function, which make them excellent candidates for acoustic measurements using simultaneous dual-sources [N. Xiang and M. R. Schroeder, J. Acoust. Soc. Am. **113**, 2754–2761 (2003)]. The application of time-reversed M-sequences was originally inspired by the outdoor acoustic tomography. This paper reports on a recent underwater experiment using the time-revered M-sequence pairs in a quasibistatic sonar system. Different from the baseband simultaneous acoustic excitations in outdoor sound propagation, the underwater application in the bistatic sonar requires ultrasonic amplitude modulation technique in order to exploit the desired correlation properties. Their excellent pseudorandom and correlation properties have also been exploited in room-acoustic applications. In addition to simultaneous dual-source room impulse response measurements, this paper also reports on creating artificial enveloping reverberations using time-reversed M-sequences.

Contributed Paper

11:40

2aSP11. Improved interference rejection in code-division multiaccess underwater acoustic communications: The near-far problem. T. C. Yang (Naval Res. Lab., 4555 Overlook Ave., Washington, DC 20375) and Wen-Bin Yang (Natl. Inst. of Standards and Tech., Gaithersburg, MD 20899)

Underwater acoustic communication is known for the signal latency due to the slow sound speed in water. This creates a severe scheduling problem for signal transmission in an underwater acoustic network involving mobile platforms or sensors, particularly when the ranges between the individual nodes are varying with time. Multiple-access communication methods such as code-division multiple-access (CDMA) that are widely used in radio frequency communications (mobile cellular network) have been considered. The problem is that the code orthogonality between the different users is destroyed by the presence of multipath arrivals in an underwater acoustic channel and severe Doppler shift in the signal. As a result, the ability to reject interference that is well known for CDMA signal is severely degraded resulting in a poor bit error rate. One approach is to recover the signal orthogonality by equalizing the multipath arrivals. However, performance of the equalization algorithm is itself degraded in the presence of interference. A different approach using the hyperspace cancellation by coordinate zeroing method is adapted in this paper for interference rejection without requiring channel equalization. Good performance is obtained using at-sea data. [This work is supported by the US Office of Naval Research.)

Meeting of the Standards Committee Plenary Group

to be held jointly with the meetings of the
ANSI-Accredited U.S. Technical Advisory Groups (TAGs) for:
ISO/TC 43, Acoustics,
ISO/TC 43/SC 1, Noise,
ISO/TC 108, Mechanical vibration, shock and condition monitoring,
ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures,
ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring instruments,
ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock,
ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines,
ISO/TC 108/SC 6, Vibration and shock generating systems,
and
IEC/TC 29, Electroacoustics

P. D. Schomer, Chair,
U.S. Technical Advisory Group (TAG) for ISO/TC 43 Acoustics and ISO/TC 43/SC 1 Noise
2117 Robert Drive, Champaign, IL 61821

D. J. Evans, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108 Mechanical vibration shock and condition monitoring, and ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring devices
National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 8220, Gaithersburg, MD 20899

A. F. Kilcullen, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
734 Walden Road, Hedgesville, WV 25427

R. Taddeo, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington, DC 20376

D. D. Reynolds, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock
3939 Briar Crest Court, Las Vegas, NV 89120

D. J. Vendittis, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
701 Northest Harbour Terrace, Boca Raton, FL 33431

R. Taddeo, Vice Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington, DC 20376

C. Peterson, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 6 Vibration and shock generating systems
200 Dixie Ave., Kalamazoo, MI 49001

V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics
National Institute of Standards and Technology (NIST), Sound Building Room A147, 100 Bureau Drive, Stop 8221, Gaithersburg, MD 20899–8221

The reports of the Chairs of these TAGs will not be presented at any other S Committee meeting.

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards of Committees S1, S2, S3, and S12, which are scheduled to take place in the following sequence:

ASC S12 Noise	**11 November 2008**	**11:00 a.m. to 12:45 p.m.**
ASC S1 Acoustics	**11 November 2008**	**2:15 p.m. to 3:30 p.m.**
ASC S2 Mechanical Vibration and Shock	**11 November 2008**	**3:45 p.m. to 5:00 p.m.**
ASC S3 Bioacoustics	**12 November 2008**	**9:00 a.m. to 10:30 a.m.**
ASC S3/SC 1 Animal Bioacoustics	**12 November 2008**	**10:45 a.m. to 12:00 noon**

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees and U.S. TAGs.

The U.S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3, and S12 are as follows:

U.S. TAG Chair/Vice Chair	TC or SC	U.S. Parallel Committee
P. D. Schomer, Chair	**ISO/TC 43** Acoustics	S1 and S3
P. D. Schomer, Chair	**ISO/TC 43/SC1** Noise	S12
D. J. Evans, Chair	**ISO/TC 108** Mechanical vibration, shock and condition monitoring	S2
A. F. Kilcullen, Co-Chair R. Taddeo, Co-Chair	**ISO/TC 108/SC2** Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures	S2
D. J. Evans, Chair	**ISO/TC 108/SC3** Use and calibration of vibration and shock measuring instruments	S2
D. D. Reynolds, Chair	**ISO/TC 108/SC4** Human exposure to mechanical vibration and shock	S3
D. J. Vendittis, Chair R. Taddeo, Vice Chair	**ISO/TC 108/SC5** Condition monitoring and diagnostics of machines	S2
C. Peterson, Chair	**ISO/TC 108/SC6** Vibration and shock generating systems	S2
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	S1 and S3

Meeting of Accredited Standards Committee (ASC) S12 Noise

R. D. Hellweg, Chair, S12
Hellweg Acoustics, 13 Pine Tree Road, Wellesley, MA 02482

W. J. Murphy, Vice Chair, S12
NIOSH, 4676 Columbia Parkway, Mail Stop C27, Cincinnati, OH 45226

Accredited Standards Committee S12 on Noise. Working group chairs will report on the status of noise standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 43/SC 1 Noise, take note - that meeting will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Tuesday, 11 November 2008.

Scope of S12: Standards, specifications and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort, and physical acoustics as related to environmental and occupational noise.

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

P. Battenberg, Chair, S1
Quest Technologies, Inc., 1060 Corporate Center Drive, Oconomowoc, WI 53066-4828

R. J. Peppin, Vice Chair, S1
Scantek, Inc., 7060 #L Oakland Mills Road, Columbia, MD 21046

Accredited Standards Committee S1 on Acoustics. Working group chairs will report on the status of standards currently under development in the areas of physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound, etc. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note - those meetings will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Tuesday, 11 November 2008.

Scope of S1: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics, including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S2 Mechanical Vibration and Shock

R. L. Eshleman, Chair, S2
Vibration Institute, 6262 Kingery Highway, Ste. 212, Willowbrook, IL 60527

A. T. Herfat, Vice Chair, S2
Emerson Climate Technologies, Inc., 1675 West Campbell Road, PO Box 669, Sidney, OH 45365 0669

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working group chairs will report on the status of various shock and vibration standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 108, Mechanical vibration, shock and condition monitoring, and its five subcommittees, take note - that meeting will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Tuesday, 11 November 2008.

Scope of S2: Standards, specification, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

Session 2pAA

Architectural Acoustics: Acoustics of Retrofitted Performance Spaces

Dana S. Hougland, Cochair
Shen Milsom Wilke, Inc., 1822 Blake St., Ste. 2A, Denver,CO 80202

David S. Woolworth, Cochair
Oxford Acoustics Inc., 356 CR 102, Oxford, MS 38655

Chair's Introduction—1:00

Invited Papers

1:05

2pAA1. Adding art to an artless space: case studies in refitting existing buildings as performance facilities. Gregory Miller, Evelyn Way, Byron Harrison, and Richard Talaske (TALASKE, 1033 South Boulevard, Oak Park, IL 60302)

Performing artists will find ways to present their work, often in spaces that they find and adapt for their own use. This paper will provide a number of historic examples of the use of "found spaces" for theatrical performance (specifically referencing the work of Peter Brook and CIRT), as well as contemporary case studies of buildings that have undergone significant renovations. Contemporary projects will include the Lookingglass Theatre (a former municipal pumping station), the Steppenwolf Garage (a parking structure), and the Lincoln Park Music Theatre (a former bank building).

1:25

2pAA2. Barons, barns, bombs: Innovative unlikely homes for the arts. Scott Pfeiffer (Threshold Acoust. LLC, 53 West Jackson Blvd. Ste., 1734, Chicago, IL 60604)

The School of the Art Institute, The Weston Playhouse, and The Peck School of the Arts all have made use of existing structures as augmentations of their facilities. Sustainability is served in these projects regardless of their pursuit of the U.S. Green Building Council Leadership in Energy and Environmental Design Certification. Adaptive reuse preserves the energy spent in the creation of the materials and the construction of the original building, as long as modern systems are introduced that allow the buildings to function efficiently.

In each case, the unique qualities that made the buildings successful in their original use were exploited in their redevelopment. Issues of adapting structure, mechanical systems, and adapting to the local environment all guide the setting of priorities in the acoustic design. Special attention is paid to developing flexible infrastructure since clearly the second (or sometimes the third) use of the building will not be its last. The special problems overcome in retrofitting the existing space for the performing arts are discussed.

1:45

2pAA3. Four buildings converted to performance spaces. David Woolworth (Oxford Acoust., 356 CR 102, Oxford, MS 38655, dave@oxfordacoustics.com)

This paper documents the conversion of two metal buildings and two brick buildings into performance spaces: a metal seafood restaurant into a church sanctuary, a cotton gin into an amplified music venue, a 1928 power house into a black box and gallery, and an office complex and restaurant into a 1200-person capacity amplified venue and movie theater. The challenges and trade offs will be discussed, and the coordination of room acoustics, mechanical noise, and sound systems. Methods of control of amplified low end are addressed for each venue in terms of isolation and absorption.

2:05

2pAA4. Sala São Paulo Concert Hall, 1998 to 2008. Jose Nepomuceno (Acustica Sonica, R. Fradique Coutinho, 955 sala 01, 05416-011 São Paulo, Brasil)

Sala São Paulo is the house of Orquestra Sinonica do Estado de São Paulo (OSESP), and it is an adaptive conversion of a portion of the Julio Prestes railway terminal in São Paulo, Brazil into a concert hall of 1509 seats. Estao Julio Prestes dates from 1938 and the renovation project began in 1997. The concert hall construction started in 1998, and it was opened in July 1999. Ten years after the construction started, Sala São Paulo is a great success. A moveable ceiling with 15 panels weighing 7.5 tons each provides unique sound adjustments. The system calls attention from musicians, from the public, and from researchers. This paper presents a discussion about Sala São Paulo achievements from 1998 to 2008. For example, before Sala São Paulo, acoustics used to start as soon as the architecture design was finished. After Sala São Paulo, this is no longer the rule. Items designed for this project were further developed and installed in other venues and are now commercially available. Moveable ceiling adjustments changed since the project, strongly influenced by listening experiences, showing limitations of judgment based solely on measurements. Construction issues are discussed.

2:25

2pAA5. Rugby Hall. Marshall Long (Marshall Long Acoust., 13636 Riverside Dr., Sherman Oaks, CA 91423)

Rugby Hall began as a flat floored rectangular room with a low ceiling, which had false wooden beams at right angles to the main axis of the room. There was a low-raised platform on the short wall at one end of the room which served as a makeshift stage. The room was used for lectures, plays, and musicals and performed poorly for all uses. After much puzzling it was decided to excavate the floor to yield raked seating and an expanded stage area. The false beams were removed and the ceiling was raised. The new ceiling had to be designed around the large steel beams above and a unique sidewall design was developed. The resulting hall which seats 386 has served the school for over 20 years as a mixed use facility.

2:45

2pAA6. A new performance space for Temple University in the oldest building on campus. David Greenberg (Creative Acoust., LLC, 5 Inwood Ln., Westport, CT 06880) and Felicia Doggett (Metropolitan Acoust., LLC, Philadelphia, PA 19118)

Construction of the Baptist Temple on North Broad Street in Philadelphia was begun in 1899 and completed the following year. Temple University was founded within its walls and consequently where it got its name. The University purchased the building in 1972 where it remained vacant for several decades. In 2007, Temple University announced plans to adapt the structure as a multipurpose venue for orchestra and choral performances, jazz ensembles, drama, lectures, and convocations. The building includes the main worship space located above the Chapel of the Four Chaplains in the basement, both of which will be renovated into performance spaces. How does one renovate a wood-framed, volume-limited, noisy street side building with deep wraparound balconies and less-than-ideal mechanical equipment locations into a performance venue for a high-quality music department on a limited budget with demanding users and an ultra-cost-conscious owner while maintaining the historic fabric of the building? This presentation explores not only the details of the renovations but also the complicated relationships between the owner, the user, the architect, the acoustical consultant, and the peer reviewer.

3:05—3:20 Break

3:20

2pAA7. Retrofitting a performance space: Everyman Theatre. Julie Fischer (Shen Milsom Wilke, 3300 N Fairfax Dr., Ste. 302, Arlington, VA 22201)

The new home of the Everyman Theatre company opened in 1910 as a Vaudeville house and has had many reincarnations throughout its lifetime, including a parking garage and 1550-seat movie theater. In its latest reincarnation, the space under renovation was a large movie theater. A challenge for the project team has been to take this configuration and turn it into something that fits the needs of the Everyman Theater company. The current renovation is underway and will include a large studio theater, with approximately 250 seats and several different configurations including a typical end stage, thrust stage, and arena stage, a Blackbox theater, and support spaces such as classrooms, offices, and rehearsal space. One of the main issues related to the pre-existing space is its current structure. The current design has two theaters stacked vertically and there is concern that noise and vibration from activities from the upper theater

could disturb occupants in the lower theater. Upgrading the structure is complicated by the location of the building which sits directly above a metro tunnel and therefore has strict weight requirements. The approach currently proposed is to include a floating concrete floor and sound barrier ceiling. [Chris Pollock (SMW), Rima Namek and Diane Cho (Cho Benn Holback), Millie Dixon and John Coyne (Theatre Projects), and Vincent Lancisi (Everyman Theatre).]

3:40

2pAA8. 939 Cafe at Berklee College of Music: From an old office to a performing space. Ioana Pieleanu (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

Long due for the world renowned jazz and contemporary music school, and open in October 2007, the 939 Cafe is the first club-type performing space at Berklee College of Music in Boston. Given the urban location of the college and the shortage of available real estate in the area, the Cafe is a retrofit of an old architect's office space. Among the Cafe's neighbors in the building there is the Bostons Tennis and Racket Club, a building tenant since 1902, and an independent restaurant/bar. All aspects of the acoustical design, including the room acoustics, sound isolation, and mechanical system noise control, presented challenges in this limited volume historical setting. This presentation discusses these challenges, limitations, and the proposed acoustical solutions.

4:00

2pAA9. Acoustic challenges of converting a historical hall to multipurpose use. Anthony Bontomase and Kristin Bleedorn (Shen Milsom Wilke, 417 Fifth Ave., New York, NY 10016, abontomase@smwinc.com)

The Great Hall of the historical Cunard Building in New York City was originally used for the Cunard passenger ship lines ticketing beginning in 1921 and later converted to a US Post Office branch. After years of vacancy, a hospitality company proposed to take advantage of the attractive Renaissance Revival inspired hall for events ranging from speeches to weddings and concerts. The Great Hall features a huge vaulted ceiling and is surrounded by occupied office spaces, with original windows between. The acoustic goals were to reduce sound transmission into offices and to tame the reverberation time to accommodate various programs. Several acoustic tests were performed to establish the reverberation time and the noise isolation class between the hall and offices. Further testing with windows covered overyielded the achievable noise isolation without modification of walls within office spaces. Only nonpermanent interior acoustic treatments were permitted as preserving the visual appeal of the interior is critical. Insulating glass additions on the interior of offices is expected to provide a noise isolation class above 50. The use of carpet and suspended drapery is expected to reduce the reverberation time to be more manageable for a variety of programs.

Contributed Papers

4:20

2pAA10. Recording studio rehabilitation. Marshall Long (Marshall Long Acoust., 13636 Riverside Dr., Sherman Oaks, CA 91423)

From time to time projects are designed and built, which do not perform in the expected manner. Sometimes outside events interfere with the design intent. In other cases there is a fundamental misunderstanding of the technical principles involved. This is an example of one of these cases. A recording studio was designed (by a well known studio designer, who shall remain nameless) and built on the first floor of a multistory high rise in Los Angeles, CA. It consisted of a studio approximately $20\times18\times10$ ft^3 high and an adjacent control room. A 3-in. concrete floor was isolated on 1/2-in. rubber and the walls and ceiling were built independently on top of the floor. After completion both footfall and carts being pushed along a corridor approximately 15 ft away were clearly audible inside the studio. This paper describes the steps taken to analyze and correct the problem.

4:35

2pAA11. Multithread implementation for calculating room impulse responses. Wolfgang Ahnert, Stefan Feistel, and Alexandru Miron (Ahnert Feistel Media Group Berlin, Arkonastr. 45-49, 13189 Berlin, Germany, wahnert@ada-acousticdesign.de)

Modern simulation programs are committed to calculate full impulse responses instead of doing simple mapping presentations. The main problem here is how to obtain such IRs quickly but accurately. The usual algorithm simulates sound propagation by tracing a statistical ensemble of acoustic rays. In this contribution, we deal with a new ray-triangle intersection algorithm, based on a three-dimensional grid scheme, that provides significant performance improvement for ray tracing. We report about the introduction of a uniform grid ray-tracing algorithm, which is optimized for acoustic models and can be up to five times faster (standard Pentium 4) than hierarchical space decomposition. Furthermore, the algorithm has been parallelized, taking advantage of the inherently independent propagation of distinct rays. On multicore processors, the computation employs a variable number of thread tracing groups of rays in parallel. The speed increase is almost linear with the number of cores. An overview is given to explain this approach. To support this fast calculation, an additional algorithm derives a suggested particle number and length from the mean free path properties of the room and the desired detection rate at the receiver. So time-consuming guesses about the selecting particle numbers will be avoided.

4:50

2pAA12. Swept sine against maximum length sequences in room acoustics with music signals as background noise. Joel Paulo (DEETC, ISEL -Tecnical Inst. of Lisbon, Rua Conselheiro Emídio Navarro 1, Lisbon, 1959-007, Portugal, jpaulo@deetc.isel.ipl.pt) and José Bento Coelho (CAPS - Instituto Superior Tcnico, Lisbon, Portugal)

The Swept Sine and the MLS techniques are very popular in room acoustic measurement set-ups. Advantages and disadvantages of both methods have been well investigated and can be found in the literature. However, information regarding the performance of these techniques in the presence of high background music levels is scarce. Since the estimation of the room impulse response is based on the correlation between signals, the likelihood between the test signal and the music contents plays an important role on the results' accuracy. This paper explores these issues by taking into account the semantic information of the music signals when used as disturbance. The method used for the assessment of the gain between the two techniques consists of splitting each frame into segments and applying a weighting function depending on a likelihood function. The features used for the likelihood function are the rms value of each segment, spectral energy envelope relation, bandwidth and harmonic structure. Several examples are presented for comparison of the performance of the Swept Sine and the MLS techniques. Advantages and disadvantages of each technique are discussed for music signals as noise.

5:05

2pAA13. Predictions of sound energy flows in coupled spaces using a diffusion equation model. Yun Jing and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

In this work, a diffusion equation model [J. Xiang, J. Acoust. Soc. Am., **123**, 145–153 (2008)] is applied to two coupled spaces to study the time-dependent sound energy flows through the coupling area (aperture). The energy feedback is found when the primary room is less reverberant than the secondary room. The Bayesian framework is applied to quantify double-slope characteristics of sound-energy flow decays obtained from the diffusion equation numerical simulations. This work also reveals that the turning point on a double-sloped energy decay is highly correlated to the time instant when the energy flow direction flips over.

TUESDAY AFTERNOON, 11 NOVEMBER 2008 LEGENDS 2, 1:30 TO 4:55 P.M.

Session 2pAB

Animal Bioacoustics: Marine Mammal Acoustics in Honor of Sam Ridgway II

James J. Finneran, Chair
SPAWARSYSCEN San Diego, 53560 Hull St., San Diego, CA 92152

Invited Papers

1:30

2pAB1. Prey capture by harbor porpoises. Lee Miller (Inst. of Biology, Univ. of Southern Denmark, Campusvej 55, DK-5230 Odense M., Denmark, lee@biology.sdu.dk)

The harbor porpoise (*Phocoena phocoena*) is a small toothed whale living mostly in coastal waters. There are large, but unknown, numbers in the inner Danish waters. Four are in captivity at Fjord Belt Center, Kerteminde, Denmark, one of which was born there in 2006. Harbor porpoises use their ultrasonic clicks as biosonar for orientation and detection of prey (mostly smaller pelagic and bottom dwelling fish), and for communication. For studying wild animals, hydrophone arrays [Villadsgaard *et al.*, J. Exp. Biol. **210** (2007)] and acoustic (time/depth) tags [Akamatsu *et al.*, Deep Sea Res. **2** (2007)] have been used. For studying captive animals, arrays and video techniques [Verfuss *et al.*, J. Exp. Biol. **208** (2005)], as well as miniature acoustic-behavioral tags [Deruiter *et al.*, J. Acoust. Soc. Am **123** (2008)], have been used. While searching for prey, harbor porpoises use clicks at long intervals (50 ms) that progressively decrease when closing an object. After detecting the prey, the click interval stabilizes and then becomes progressively shorter while approaching the prey. The sequence ends in a terminal high-repetition rate buzz (500 clicks/s) just before capturing the prey (a video will be shown). The temporal sequence differs from that of beaked whales but is similar to that of bats.

1:50

2pAB2. Backscatter measurements of three species of salmon using simulated killer whale echolocation signals. Whitlow W. L. Au (Hawaii Inst. of Marine Biology, Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734), John K. Horne and Christopher D. Jones (Univ. of Washington, Seattle, WA 98105-6698)

The resident ecotype of killer whales (*Orcinus orca*) in the waters of British Columbia and Washington State has a strong preference for Chinook salmon even in months when Chinook are 5%–10% of the salmon population. The foraging behavior of killer whales suggests that they rely on echolocation to detect and recognize their prey. In order to determine possible cues in echoes from salmon, a series of backscatter measurements was made at the Applied Physics Laboratory, University of Washington Facility on Lake Union, on three different salmon species using simulated killer whale echolocation signals. The fish were tied to a monofilament net and rotated while echoes were collected, digitized, and stored on a laptop computer. Three transducer depths were used; same depth, 22^{o}, and 45^{o} above the horizontal plane of the fish. Echoes were collected from five Chinook, three Coho, and one Sockeye salmon. Radiograph images of all the specimens were also obtained to examine the swim bladder shape and orientation. Results show that the echo structure from similar sized but different species of salmon were different and probably recognizable by foraging killer whales. The results also suggest that a broadband echo-sounder pointing downward could be used to discriminate salmon species.

2:10

2pAB3. Progress in clinical hearing evaluation of bottlenose dolphins (*Tursiops truncatus*). James J. Finneran (U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Ctr., San Diego, Code 71510, 49620 Beluga Rd., San Diego, CA 92152) and Dorian S. Houser (Biomimetica, Santee, CA 92071)

It has long been recognized that the need to train experimental subjects is a major obstacle to large scale behavioral hearing studies of marine mammals. The time and level of access required for behavioral audiometry have not only prevented hearing assessment in wild individuals but have also limited attempts to assess auditory system fitness in captive populations. Ridgway and co-workers of electrophysiological methods were quick to realize that auditory evoked potential (AEP) measurements could be used to assess marine mammal hearing without lengthy training. Ridgway performed some of the earliest work to characterize dolphin evoked potentials and to develop fast techniques for threshold measurement. In addition to hearing assessment in larger whales not routinely kept in captivity,

Ridgway has also strived to develop techniques that would allow routine clinical hearing assessment of captive animals. This talk reviews the development of AEP techniques for rapid large-scale hearing assessment in dolphins at the US Navy Marine Mammal Program. Topics range from Ridgway's early work on dolphin AEP threshold measurements to recent efforts to acquire longitudinal audiometric data from a large population of dolphins. [Work supported by ONR.]

2:30

2pAB4. Application of the evoked-potential method to the study of the odontocete biosonar. Alexander Supin (Inst. of Ecology and Evolution, Russian Acad. of Sci., 33 Leninsky Prosp., 119071 Moscow, Russia)

In a false killer whale *Pseudorca crassidens*, auditory evoked potentials (AEPs) were recorded in conditions of echolocation using both real targets and electronically synthesized and played-back (phantom) echoes. The electronic echoes were triggered by emitted sonar pulses, their spectra were similar to that of the emitted sonar clicks of the subject, and their intensities were proportional to the level of the emitted sonar pulse. AEPs to both the emitted pulse and real or electronic echoes were well detectable. AEPs related to emitted sonar clicks displayed the regular amplitude dependence on click level: The higher the level, the higher the amplitude. The echo-related AEP amplitudes depended on both echo attenuation and delay with a trade of around 9-dB attenuation per delay doubling. When the echo attenuation was varied in conjunction with delay keeping a rate of 9-dB deeper attenuation per delay doubling, the echo-related AEPs were nearly invariant. This result is well explainable by a hypothesis implying that partial forward masking of echoes by preceding emitted sonar pulses serves as a mechanism of automatic gain control in the auditory system of echolocating odontocetes.

2:50

2pAB5. Perception of echoes with FM1-FM2 delay disparities: Bats have selective direction-of-gaze high-resolution imaging. James A. Simmons (Dept. of Neurosci., Brown Univ., Providence, RI 02912, james-simmons@brown.edu), Mary E. Bates, Sarah A. Stamper and Douglas Benedicto (Brown Univ., Providence, RI 02912)

Echolocating big brown bats (*Eptesicus fuscus*) emit frequency-modulated (FM) biosonar sounds containing two prominent downsweeping harmonics, FM1 from 55 to 23 kHz and FM2 from 100 to 45 kHz. Acoustics of echoes in air ordinarily keep FM1 and FM2 aligned in time; however, if FM1 and FM2 are segregated artificially with low-pass and hig-hpass filters and delivered to bats at different delays, accuracy of delay perception is disrupted. As a direct confirmation of delay-resolution results from echo jitter experiments, harmonic delay offsets of 1–2 us are sufficient for the bat to reject echoes from accurate delay determination. Echoes ordinarily do arrive with differentially greater attenuation of FM2 relative to FM1 for longer ranges (FM2/FM1 is −3 dB for a range of 2 m) and for off-axis targets (FM2/FM1 added −3 dB for angles of 15 deg at a 2-m range). The resulting greater neuronal amplitude-latency trading for FM2 relative to FM1 activates the harmonic delay-offset effect to prevent accurate determination of delay. Noncoherence of harmonics in auditory neuronal spectrograms shunts information about the off-axis and far away targets to a different display of range and crossrange than for targets located in the direction of gaze. [Work supported by the ONR and the NIMH.]

3:10—3:25 Break

Contributed Papers

3:25

2pAB6. Underwater sound measurements of the Hawaii super ferry and the potential impact to humpback whales in the main Hawaiian Islands. Alison K. Stimpert, Alexis Rudd, and Whitlow W. L. Au (Inst. of Marine Biology, Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734)

Underwater measurements of the Hawaii Super Ferry were made with a vertical array of four hydrophones. The depths of the hydrophones were 3, 6.5, 10, and 20 m. The Alakai is a 107 m catamaran passenger and automobile ferry that has a waterjet propulsion system with no exposed propellers. The Alakai made three close proximity passes at speeds of 37, 24, and 12 kn on tracks that allowed for acoustic measurements of sounds radiating from the bow, broadside, and stern aspects. The track of the Alakai was recorded on global positioning system. The measurements were conducted while the ferry proceeded on normal scheduled trips, allowing for three measurements a day. All the sound pressure levels were referenced to 91 m (100 yards), the minimum legal approach distance of boats to humpback whales. Sound intensity increased with speed and was the highest at 37 kn in the broadside aspect (160 dB) and was 14 dB lower for both the bow and stern aspects. The sounds had a broadband noise quality with spectra that had a low frequency peak below 100 Hz and decreased continuously with frequency. The effects of the sounds will be discussed from a humpback whale perspective.

3:40

2pAB7. Patterns of coastal use by Hawaiian spinner dolphins (*Stenella longirostris*) observed using passive acoustic monitoring. Marc O. Lammers (Hawaii Institue of Marine Biology, P.O. Box 1106, Kailua, HI 96734, lammers@hawaii.edu), Simona Bernasconi (CIBRA, Universita Degli Study di Pavia, Pavia 27100, Italy), Whitlow W. L. Au (Hawaii Inst. of Marine Biology, Kaneohe, HI 96744), Kevin Wong, and Russell E. Brainard (Pacific Islands Fisheries Sci. Ctr., Honolulu, HI 96822)

Spinner dolphins are the most commonly sighted cetacean in coastal Hawaiian waters and are an important higher-level trophic component of the near-shore ecosystem. Establishing their long-term patterns of occurrence is important both for their conservation and for a better understanding of the relationships between dolphins, their prey, and other members of the ecosystem. Here we report on an effort to use ecological acoustic recorders (EARs) to monitor the presence of spinner dolphins along the leeward coast of the island of Oahu, Hawaii. The EAR is a bottom-moored recorder with a bandwidth of 30 kHz, which allows the detection of both dolphin whistles and echolocation clicks. Eight EARs were deployed along the coast at depths ranging from 15 to 50 m. Five units were deployed in an array, providing information about the EAR's detection range. The other three units were placed at sites along the coast commonly frequented by spinner dolphins. The results reveal distinct patterns of preference in habitat use,

both for daytime resting behavior and night-time foraging. The recordings also provide a measure of vessel traffic at locations important to spinner dolphins. Combined, these data demonstrate the value of passive acoustic methods for monitoring cetacean populations and their habitat over extended periods.

3:55

2pAB8. Temporal and geographic patterns in the occurrence and distribution of minke whale (*Balaenoptera acutorostrata*) boings in the central and western North Pacific. Julie Oswald (Hawaii Inst. of Marine Biology, Univ. of Hawaii, 46-007 Lilipuna Rd., Kaneohe, HI 96744, jnoswald@hawaii.edu), Tom Norris (Bio-Waves Inc., Encinitas, CA 92024), Whitlow Au (Hawaii Inst. of Marine Biology, Kaneohe, HI 96744), and Fred Duennebier (Univ. of Hawaii, Honolulu, HI 96822)

Minke whales are elusive and difficult to study using visual methods. The source of the "boing" sound was recently linked to North Pacific minke whales, allowing passive acoustics to be used to study this species. The seasonal occurrence of minke whales was examined using data collected at the Station ALOHA Cabled Observatory, an ocean bottom hydrophone 100 km north of Oahu. Preliminary analysis of data collected between February and June 2007 indicates that boings occur during all of these months, peaking in early April. No diurnal variation was evident. Towed hydrophone-array surveys were conducted in the offshore waters of the islands of Oahu, Kauai and Ni'ihau (February 2005) and off Guam and the Northern Mariana Islands (January–April 2007). Although rarely observed visually, the prevalence of boings detected in these areas indicates that minke whales are common. Distribution patterns from both studies suggest that minke whales prefer deep but not the deepest waters. Boings recorded from Guam and the Northern Mariana Islands appear to be more similar to the "central" boing (which includes the Hawaiian Islands) than the "eastern" boing [which includes those recorded east of 138°W, Rankin and Barlow (2005)]. This has important implications for North Pacific minke whale stock structure.

4:10

2pAB9. Contextual sound production by tagged humpback whales (*Megaptera novaeangliae*) on a feeding and breeding ground. Alison K. Stimpert, Whitlow W. L. Au (Hawaii Inst. of Marine Biology Marine Mammal Res. Program, Dept. of Zoology, Univ. of Hawaii at Manoa, P.O. Box 1106, Kailua, HI 96734, stimpert@hawaii.edu), David N. Wiley (Stellwagen Bank Natl. Marine Sanctuary, Scituate, MA 02066), and David Mattila (Hawaiian Islands Humpback Whale Natl. Marine Sanctuary, Kihei, HI 96753)

Humpback whales are amongst the best studied of the baleen whales. The species is also known for its flexibility and variety of behavior on both feeding and breeding grounds. DTAGs were attached to 19 whales on the Hawaiian breeding grounds and to 24 whales on the northwestern Atlantic feeding grounds between the years of 2004 and 2008 to describe the variety of acoustic behavior in each location. Sounds were analyzed using ADOBE AUDITION, XBAT, and custom programs in MATLAB 7, and acoustic records from the tags showed differences in sound production between the two populations engaged in different activities. Recordings from the feeding grounds showed higher social sound production rates and also contained the sounds with the highest signal-to-noise ratios (10–20 dB for 1 μPa higher than on the breeding grounds). Differences in ambient noise may contribute to this: song chorusing is present in Hawaii, but commercial shipping traffic is higher in the northwest Atlantic feeding areas. Some similar sounds were recorded from the two areas, and exemplars of sound types from each location will be described. Overall, the feeding whales appeared to use sound more frequently, and perhaps for longer range communication than sounds produced within competitive groups in Hawaii.

4:25

2pAB10. The presence of low-frequency narrowband phonations in the wild bottlenose dolphins' acoustic repertoire. Natalija Lace and Stan Kuczaj (Dept. of Psych., Univ. of Southern Mississippi, P.O. Box 5025, 118 College Dr., Hattiesburg, MS 39406, kodzaks@yahoo.com)

Bottlenose dolphins' phonations are commonly divided into three major categories: whistles, echolocation clicks, and burst pulses. Other categories are often mentioned and described as yelps, squawks, barks, and low-frequency narrowband sounds. Here, we report the occurrence of low-frequency tonal phonations with fundamental frequency within the 500–2000 Hz range, accompanied by numerous harmonics. Recordings were made using 48 and 192 kHz sampling rates and a 100 Hz high-pass filter in the presence of free-ranging bottlenose dolphins in Mississippi Sound, MS. Auditory sensitivity of the bottlenose dolphins has been studied extensively, and the results indicate that dolphins have their best hearing sensitivity in the higher-frequency range (15–110 kHz). Low-frequency sounds (above 75 Hz) can be detected as well, and it has been suggested that this detection mechanism may be entirely different from that used for higher frequencies and may even include mechanoreception. The occurrence of low-frequency phonations in wild dolphins indicates that low-frequency detection may play an important role in the animals' everyday activities.

4:40

2pAB11. Spatial distribution of right whale "gunshot" sound displays in the Bay of Fundy, Canada. Susan E. Parks, Cara F. Hotchkin (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, sep20@psu.edu), and Christopher W. Clark (Cornell Univ., Ithaca, NY 14850)

North Atlantic right whales (*Eubalaena glacialis*) produce a distinctive loud broadband signal referred to as the "gunshot" sound. Current hypotheses for the function of this signal include an agonistic threat signal between males, a male reproductive advertisement signal to female right whales, or a combination of both. This sound has been documented both in a social context in mixed sex surface active groups and by lone individual males in a stereotyped display. An array of five archival bottom mounted acoustic recorders was deployed in the Center of the North Atlantic Right Whale Conservation Area in the Bay of Fundy, Canada in August 2005. The five recorders were spaced 3–7 nm apart, allowing for localization of whales producing gunshot sounds within or near the array. These recordings were used to describe the regional distribution, spacing, movement patterns, and timing of gunshot sound displays produced by right whales over a two week period to investigate the potential function of this sound. Further investigations included both the diel trends in the sound production and evidence for interactions between individuals simultaneously producing these displays.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration: Microbubble Response and Modeling

Tyrone M. Porter, Chair
Boston Univ., Aerospace and Mechanical Eng., 110 Cummington St., Boston, MA 02215

Contributed Papers

1:30

2pBB1. Coupled two-phase model of a gas microbubble containing suspended light absorbing nonoparticles. Elisabetta Sassaroli, Brian E. O'Neill, and King C.P. Li (The Methodist Hospital Res. Inst., 6565 Fannin St, Houston, TX 77030, esassaroli@comcast.net)

A mathematical model of a micrometer size gas bubble containing nanometer size light absorbing nanoparticles is presented. A description of such a system can be obtained in terms of a coupled two-phase model with the solid particles in suspension in the gas phase. It is assumed that the suspension is diluted so that particle-particle interaction can be ignored. Because the heat exchange between the particles and the gas is of main interest in this calculation, it is assumed in first approximation that the gas and the particles have the same velocity and pressure. The pressure is assumed to be a function of time only. In this case, only the equations of continuity and energy for both the particulate phase and the gas phase are needed. The two-phase model is then solved in a linear approximation, and a system of four differential equations with constant coefficients is obtained. The system is diagonalized and the general solution of the coupled system is obtained. The general solution is a combination of exponential decaying oscillatory functions for the temperature of the two phases, the pressure, and the microbubble radial oscillations. It was found that the oscillatory behavior takes place in the megahertz range.

1:45

2pBB2. The effect of scatterer size distribution on the backscatter coefficient. Michael King (Bioacoustics Res. Lab., Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, mrking2@uiuc.edu), Ernest L. Madsen, Timothy J. Hall (Univ. of Wisconsin-Madison, Madison, WI 53706), Alexander Haak, Michael L. Oelze, and William D. O'Brien, Jr. (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The bias and variance of mean scatterer size estimates made from backscatter coefficients might be reduced by incorporating an estimate of scatterer size distribution into the parameter estimation scheme. Current estimation schemes generally assume a single scatterer size representing the volume-weighted mean diameter of a distribution. When the mean interrogation frequency has a *ka* value near 0.8 for the mean scatterer size, and the size distribution is relatively narrow, the results are unbiased. However, consideration of the distribution of sizes may improve scatterer size estimation, especially at frequencies for which $ka > 1$. A phantom of 2% (dry weight) agar spheres (with diameters between 90 and 125 μm) in a 3.4% agar background was scanned over a broad frequency range using several transducers, and the backscatter coefficients were estimated from these data. When accounting for the size distribution, the mean square error between the measured and theoretical backscatter coefficients over the frequency range of 3–14 MHz was reduced by at least 50% in the size range of interest. These results suggest that accounting for the size distribution in backscatter models will significantly improve the ability to parametrize the backscatter coefficient. [Work supported by NIH CA111289]

2:00

2pBB3. Modeling of the dynamics of a coated microbubble confined in a blood vessel. Sergey Martynov, Eleanor Stride, and Nader Saffari (Dept. of Mech. Eng., Univ. College London, Torrington Pl., London WC1E 7JE, UK, e.stride@ntlworld.com)

Coated microbubbles have been extensively investigated as contrast agents for diagnostic ultrasound imaging and more recently for therapeutic applications such as targeted drug delivery. However, theoretical models for microbubble dynamics have previously been developed either for encapsulated bubbles in an infinite volume of fluid or for uncoated bubbles in a confined volume. In the present study, a numerical model is developed to explore the effects of both encapsulation and confinement in a blood vessel upon the microbubble response in the time domain. Both surfactant-coated microbubbles and polymeric microspheres are examined for a range of vessel:bubble diameter ratios and mechanical properties. The model is validated against other theoretical models of oscillating bubbles, and the theory of buckling of a thin spherical shell. It will be shown that even at low acoustic pressures (20 kPa) the radial oscillations of the bubble and the amplitude and spectrum of the radiated pressure field can be significantly modified as a result of confinement and that these effects are sensitive to the viscoelastic properties of the coating and the vessel. The implications for diagnostic and therapeutic applications of microbubbles will be discussed.

2:15

2pBB4. Definition of a cavitation index for real time monitoring during *in vitro* liposomal drug release. Lucie Somaglino, Guillaume Bouchoux, Jean-Louis Mestas, Adrien Matias, Jean-Yves Chapelon, and Cyril Lafon (Inserm, U556, 151 cours Albert Thomas, 69424 Lyon, cedex 03, France and Universit de Lyon, Lyon F-69003, France, lucie.somaglino@inserm.fr)

Drug release from liposomes can be activated by ultrasound and inertial cavitation is assumed to be the main involved mechanism. Broadband noise is known as a good indicator for such cavitation. The feasibility of assessing the extent of drug release by acoustic noise measurements has been investigated. A 1.17-MHz focused transducer (50-mm diameter and 50-mm focal length) was excited with tone-bursts (duty cycle: 10%–100%, PRF: 100 Hz–10 kHz, and Ispta < 800 W/cm^2). Emitted noise was registered with a broadband needle hydrophone and bandpass filtered between the fundamental and the first harmonic frequency for assessing the instantaneous inertial cavitation activity. A cavitation index (CI) was computed by integrating the filtered signal over time. This index was validated experimentally using terephthalate oxidized to fluorescent hydroxyterephthalate under the action of free hydroxyl radicals generated by inertial cavitation. The hydroxyterephthalate fluorescence and the CI were well correlated ($R^2 \simeq 0.92$). Experiments with liposomes were performed showing very good correlation between CI and drug release extent from liposomes. The definition of this CI allowed applying consistent doses of cavitation and performing comparative tests between different liposome formulations. [This work was funded by a grant from the Norwegian Research Council (NANOMAT programme). Epitarget AS, Norway is acknowledged for the supply of liposome samples.]

2:30

2pBB5. Time reversal acoustic receiver. Laurent Fillinger (Artann Labs., 1459 Lower Ferry Rd., Trenton, NJ 08618 and Stevens Inst. of Technologies, Hoboken, NJ 07030), Yegor Sinelnikov (ProRhythm, Ronkonkoma, NY 11779), Alexander Sutin, and Armen Sarvazyan (Artann Labs., Trenton, NJ 08618)

The time reversal acoustics (TRA) principle has been employed for focusing ultrasonic waves in various biomedical and industrial applications. TRA-based ultrasonic transmitters have numerous advantages over conventional methods of ultrasound focusing such as an ability to effectively focus and steer focal spot with a transmitter comprising even a single piezoelectric transducer. In this study it is shown that such TRA focusing systems with a small number of transducers can also work in the receiving mode and detect both active and passive targets with a spatial resolution close to a half wavelength. The feasibility tests were conducted using two types of TRA receivers. The first TRA receiver comprising a water-filled reverberator with a few 600 kHz piezoceramic elements attached was used for the detection of the signal radiated by a miniature (1 mm) emitter in a water tank. It was shown that the TRA receiver can detect the position of this point emitter with accuracy better than 1 mm. The second test was conducted with an acoustic reverberator made of series of thin randomly shaped membranes and a single disk transducer with a frequency of 3.3 MHz. This TRA receiver allowed the localization of a reflecting target illuminated by a plane acoustic wave with resolution also better than 1 mm.

2:45

2pBB6. Identifying the inertial cavitation threshold in a vessel phantom using focused ultrasound and microbubbles. Yao-Sheng Tung, James Choi, Shougang Wang (Dept. of Biomedical Eng., Columbia Univ., 351 Eng. Terrace, 1210 Amsterdam Ave., New York, NY 10027, YT2235@columbia.edu), Jameel Feshitan, Mark Borden and Elisa Konofagou (Columbia Univ., New York, NY 10027)

Unveiling the mechanism behind the blood brain barrier using focused ultrasound (FUS) and microbubbles is essential for brain molecular delivery. Here, *B*-mode imaging and rf signals were acquired to pinpoint the threshold for inertial cavitation during FUS with microbubbles. A cylindrical hole of 800 μm in diameter was generated inside a polyacrylamide gel to simulate a brain arterial vessel. Definity® (Lantheus Medical Imaging, USA) microbubbles with a 1.1–3.3 μm-diameter and (1,2-distearoyl-sn-glycero-3-phosphocholine) (DSPC) shelled microbubbles with a 1–2-μm-diameter were injected prior to sonication (frequency: 1.525 MHz; pulse length: 100 cycles; PRF: 1 kHz; pulse duration: 40 ms). The cavitation response was passively detected using a 7.5-MHz single-element transducer, confocal with the FUS transducer, and a one-dimensional linear array transducer placed perpendicular to the FUS beam. The broadband spectral response of the acquired rf signals and the *B*-mode images detected the occurrence and location of inertial cavitation, respectively. Findings indicated that the rarefactional pressure threshold was approximately 0.53 MPa for the definity and 0.69 MPa for the DSPC-shelled microbubbles. Further studies will vary the vessel diameter and microbubble size to determine their role in the pressure threshold. [This work was supported by NIH Grant No. R21EY018505 and NSF CAREER 0644713. The authors thank Jennifer Hui for bubble characterization.]

3:00

2pBB7. Single transmitter time reversal focusing characterization using cross-correlation method. Yegor Sinelnikov and Andrey Vedernikov (ProRhythm Inc., 105 Comac St., Ronkonkoma, NY 11770)

Acoustic focusing with reverberating acoustic cavities based on the time reversal acoustic principle has been demonstrated in a range of medical applications from three-dimensional imaging to the generation of high pressure therapeutic pulses for the destruction of kidney and gallbladder stones in the human body. We investigate acoustic focusing in a system comprised of a single transmitter and a layered membrane reverberator. The system is capable of achieving focal peak acoustic pressures up to 0.4 MPa with less than 1 mm focal spot and steering range up to 20 mm. Focusing in the frequency range between 0.5 and 4 MHz is investigated. The parametric investigation of various membrane configurations is performed by a conventional time reversal technique with a hydrophone. In addition, we propose the cross-correlation of hydrophone signals as a method of evaluating the system focusing properties without conducting the actual time reversal focusing experiment. The cross-correlation method predicts the focal spot dimensions for the layered reverberator single transmitter system that compares well with experiment and theory. An important implication is the possibility to use inverse filtering to construct signals of the desired waveform with a low spatial peak pressure amplitude that can enhance the cell's permeability toward macroparticle uptake.

3:15

2pBB8. Information theoretic approaches to ultrasonic system design: Beamforming with interative spatial filters. Nghia Nguyen (Dept. of ECE and Beckman Inst., Univ. of UIUC, 405 N. Mathews, Urbana, IL 61801, nnguyen6@uiuc.edu), Craig Abbey (Univ. of California at Santa Barbara, Santa Barbara, CA 93106), and Michael Insana (Univ. of UIUC, Urbana, IL 61801)

First-principles approaches to the design of medical ultrasonic imaging systems for specific visual tasks are being explored. The study focuses on breast cancer diagnosis and is based on the ideal observer concept for visual discrimination tasks, whereby clinical tasks are expressed mathematically as likelihood functions. Realistic approximations to the ideal strategy for each visual task are proposed to maximize the diagnostic information content of the images. Based on previous studies [Abbey *et al.*, IEEE Trans. Med. Imaging **25**, 198–209 (2006)], it is known that the Wiener filter approach to beamforming, derived as a stationary approximation of the ideal observer, is limited for low-contrast, large area lesion discrimination. In this study an adaptive iterative Wiener filter coupled to a segmentation algorithm is developed to optimally beamform and process radio-frequency (rf) signals. The segmentation is achieved by applying a Markov random field approach. Performance is measured by comparing proportion correct in human-observer studies with the ideal observer and its approximations as required for an estimate of discrimination efficiency. It is found that by adapting to local statistical properties of the rf signal the iterative Wiener filter increases visual discrimination efficiency where the Wiener filter and conventional beamformers fail. [Work supported by the NCI/NIH.]

Session 2pEA

Engineering Acoustics: Acoustic Engineering of Materials and Techniques

Thomas R. Howarth, Cochair
NAVSEA Newport, 1176 Howell St., Newport, RI 02841

Kim C. Benjamin, Cochair
NAVSEA Newport, Newport, RI 02841

Contributed Papers

1:30

2pEA1. Coupling coefficient of segmented stack piezoelectric transducers using high-coupling materials. Stephen C. Thompson, Richard J. Meyer, and Douglas C. Markley (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

High-power piezoelectric transducers normally use a segmented stack of piezoelectric pieces. The pieces are arranged with alternating polarization and are driven by electrodes of alternating voltage polarity. This allows long stacks to be built from easily manufactured pieces and reduces the voltage required for full power operation. The common understanding among transducer designers is that the mechanical properties of the stack of n pieces are equivalent to those of a single piezoelectric piece of equal total length, while the electrical impedance is reduced by the factor n^2. For conventional designs with operating bandwidth less than a half octave, this is quite accurate. However, as new materials with high-electromechanical coupling make it possible to design high-power transducers with bandwidths greater than an octave, the use of segmented stacks must be reexamined. With high coupling materials, there is a significant reduction in effective coupling coefficient caused by stack segmentation. The effect is small for materials such as hard PZT, but is larger in at least some transducers built from materials with coupling coefficient above 80%. This paper will address the effects of stack design based on modeled and measured results. [Work supported by the Office of Naval Research.]

1:45

2pEA2. Galfenol material properties and impact on transducer design. Ryan S. Scott (Knowles Electronics, LLC, 1151 Maplewood Dr., Itasca, IL 60143), Richard J. Meyer, Jr., and Stephen C. Thompson (The Penn State Univ., State College, PA 16804)

A simple measurement technique to accurately characterize magnetostrictive materials such as Terfenol-D and Galfenol is required as they become more widely used. A technique was developed that measures the magnetostrictive d_{33}, μ_{33}^{T}, s_{22}^{H}, and Q_m constants using a pair of Helmholtz coils, a laser vibrometer, and a sense coil. Unlike previous methods, this technique does not require that the magnetostrictive sample be built into a transducerlike apparatus. A matrix of magnetostrictive material properties as a function of magnetic bias and ac drive amplitude can be created using the developed method. To demonstrate the usefulness of this technique, stress annealed laminated Fe–18.4%Ga rods were tested. Property trends observed for these samples will be reported as well as their impact on transducer performance. [This work was sponsored by the Office of Naval Research (Contract No. N00014-06-1-0530) and by Etrema Products, Inc.]

2:00

2pEA3. Verification of a method for measuring magnetostrictive parameters for use in transducer design modeling. Scott Porter, Ryan Scott, Richard Meyer, Jr., and Stephen Thompson (Appl. Res. Lab., The Penn State Univ., State College, PA 16804, scott.porter@psu.edu)

Previous studies have initiated an investigation into a new method of measuring the d_{33}, μ_{33}^{T}, and s_{33}^{H} magnetostrictive constants of a given sample using a pair of Helmholtz coils, a laser vibrometer, and a sense coil. These parameters are calculated from measurements made as a function of dc magnetic bias and ac drive amplitude. This method contrasts with the traditional measurement technique which involves building the sample into a transducerlike device. The new method was demonstrated for laminated stress-annealed Galfenol rods of high aspect ratio. Measurements obtained by this method appear credible and demonstrate good repeatability. For verification, the authors have returned to the established approach and designed a simple transducer using one of the measured Galfenol rods. The transducer was driven in air as a function of dc magnetic bias and ac drive amplitude. Data obtained from this transducer are compared to an analytical model that uses the material properties obtained experimentally with the new method. [This work was sponsored by the Office of Naval Research (Contract N00014-06-1-0530) and by Etrema Products, Inc.]

2:15

2pEA4. Estimating the mechanical properties of acrylic plates via acoustic transmission experiments and modeling. Natasha A. Chang (Dept. of Mech. Eng., Univ. of Michigan, 1231 Beal Ave., 2010 Autolab, Ann Arbor, MI 48109) and David R. Dowling (Univ. of Michigan, Ann Arbor, MI 48109)

Clear acrylic [polymethylmethacrylate (PMMA)] is commonly used to separate instruments from the test-section flow in water tunnel experiments. Thus, for some hydroacoustic studies, knowledge of the sound transmission properties of acrylic plates may be essential. Unfortunately, the actual mechanical properties of PMMA plates cover a relatively large range (e.g., the elastic modulus can vary from 2 to 5.5 GPa) and these mechanical properties may be frequency dependent. In this presentation, numerical and experimental results for acoustic pulse transmission through submerged flat PMMA plates are compared to identify plate properties and to calibrate an approximate plate transmission model. The experiments were conducted in a fish tank at sound incidence angles of 0–35 deg with a spherical wave source and two receivers. The transmitted sounds were short pulses with a nominal frequency range of 40–200 kHz. Several plate thicknesses, d, were tested for kd values (k is the acoustic wave number in the water) from 0.25 to 16. A sound propagation code was used to calculate the transmission characteristics of the plate. By minimizing the difference between the calculated and measured waveforms, the mechanical properties of the PMMA plates were estimated. [Work sponsored by ONR.]

2:30

2pEA5. Two ultrasonic transducer through-wall communication system analysis. Henry A. Scarton (Dept. of Mech., Aerosp., Nuclear Eng., Rensselaer Polytechnic Inst., Troy, NY 12180)

The use of ultrasound to convey data from one side of a metallic wall to the other side is presented. A communication channel is established by attaching two ultrasonic crystals to either side of the wall. The outside transducer injects a continuous ultrasonic wave into the wall. The inside transducer operates as an energy harvester and signal modulator. The outside transducer also receives the modulated signal reflected back from the wall containing the inside transducer. A sensor on the inside provides analog data

(e.g., temperature) that is then digitized. The digitized bits are used to vary the electrical load applied to the electrical terminals of the inside transducer by changing its acoustic impedance in accordance with the data bits. The impedance changes, in turn, modulate the amplitude of the reflected ultrasonic signal. This modulated signal is detected at the outside receiving (as well as transmitting) transducer, where it is then demodulated to recover the data. Additionally, some of the acoustic power received at the inside transducer is harvested to produce the electrical power needed to operate the communication and sensor system on the inside. Digital data communication rates exceeding 50 000 bits/s are achieved.

2:45

2pEA6. Measurement uncertainties caused by tapping machines in floor sound insulation tests. Valentin Buzduga (Scantek Inc., 7060 Oakland Mills Rd., Ste. L, Columbia, MD 21045, buzdugav@scantekinc.com)

This paper illustrates how the parameters of the tapping machine may influence the measurement accuracy when testing the impact sound insulation of floors. The analysis connects the specifications for tapping machines given in ISO 140-6 and ASTM E 492-4 standards with the requirements for repeatability and reproducibility of the sound insulation measurements given in ISO 140-2. The paper also discusses calibration and testability aspects for tapping machines and presents the method developed at Scantek for measuring the impact velocity of the hammers. The paper gives modeling equations, preliminary calculations for measurement uncertainties, and experimental results obtained at Scantek on calibrating the Norsonic tapping machine N-211.

3:00

2pEA7. Ultra-wide-band filter for noise control. Manvir Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, 72570 Puebla, Mexico)

Extensive band structures are performed for two-dimensional periodic arrays of rigid stainless steel cylinders in air, with Bloch vector being perpendicular to the cylinders. We embark on the opening up of a complete acoustic band gap—independent of the polarization of the wave and of the direction of propagation. In addition, we propose the fabrication of a multiperiodic system in tandem that could create a huge hole in sound within the human audible range of frequencies.

3:15

2pEA8. Effect of sound amplitude on the acoustic attenuation characteristics of Helmholtz resonators. Asim Iqbal (Dept. of Mech. Eng. and The Ctr. for Automotive Res., The Ohio State Univ., 201 W. 19th Ave., Columbus, OH 43210, iqbal.27@osu.edu) and Ahmet Selamet (The Ohio State Univ., Columbus, OH 43210)

Influence of sound amplitude on the acoustic attenuation performance of a Helmholtz resonator is investigated computationally. Time-dependent and compressible flow field is determined by solving two-dimensional unsteady, turbulent, and compressible Navier–Stokes equations through an implicit and noniterative pressure-implicit-splitting-of-operators algorithm. The solution of full Navier–Stokes equations remedies the constraints in small-amplitude-wave treatments, thereby allowing a proper modeling of nonlinearities due to large amplitudes. Transmission loss of the Helmholtz resonator is then calculated for discrete inlet amplitudes using the pressure obtained from the flow field. Higher amplitudes are shown in this study to reduce the peak transmission loss significantly.

3:30

2pEA9. Experimental characterization of a biomimetic differential microphone diaphragm. Ronald Miles, Weili Cui (Dept. of Mech. Eng., SUNY at Binghamton, Binghamton, NY 13902-6000, miles@binghamton.edu), and Mihir Shetye (Solteras, City of Industry, CA 91748)

The identification of mechanical properties for a microphone diaphragm based on the coupled ears of the fly *Ormia ochracea* and fabricated out of polycrystalline silicon is described. An acoustic test setup using a laser vibrometer has been developed that facilitates the characterization of the diaphragm on a bare die level. A major problem with using a laser vibrometer for acoustic measurement of microstructures is the close working distance between the sensor head and test device, resulting in measurements corrupted by reflections. Time select windowing procedures are often used to obtain anechoic response estimates from measurements taken in reverberant environments, but are not effective for characterization at lower frequencies or testing of lightly damped structures where the time window length needs to be more than the reverberation time. It is shown that the reflections from the measured response of the biologically inspired diaphragm can be reduced through comparison methods for the calibration of the sound field using a closely placed probe microphone and a commercially available pressure differential microphone. Equivalent mechanical parameters for the diaphragm are estimated with a least squares identification procedure. Characterization results for a diaphragm with two different back volume configurations are compared. [Work supported by NIH.]

3:45

2pEA10. Testing microphones in small acoustical enclosures. Mariana Buzduga and Valentin Buzduga (Scantek, Inc., 7060 Oakland Mills Rd., Ste. L, Columbia, MD 21046, buzdugam@scantekinc.com)

For both instrument calibration and testing, there is a need for faster and more economical test methods, while keeping the test accuracy within acceptable limits. We analyze the use of a small acoustical enclosure that is not strictly "anechoic" but provides a test environment that can be used to make measurements with a predictably high degree of repeatability for testing the frequency characteristics of the microphones or the influence of their accessories (windscreen and grids). First, the characteristics of the acoustical field in the enclosure are discussed. Then the sensitivity of two test methods to specific influencing parameters is analyzed. The first method is the successive comparison using a reference microphone. The second method is a simultaneous comparison based on the constant divergence of the sound pressure in the field [V. Buzduga and M. Buzduga, "Acoustical test methods based on the constant divergence of the sound pressure level," Noise Control Eng. J., **51**(6), 343–347 (2003)]. The accuracy of the tests and the frequency domain of possible use for each method are discussed.

Session 2pMU

Musical Acoustics: Telematic Music Technology

Jonas Braasch, Cochair
Rensselaer Polytechnic Inst., School of Architecture, 110 8th St., Troy, NY 12180

Pauline Oliveros, Cochair
Rensselaer Polytechnic Inst., School of Architecture, 110 8th St., Troy, NY 12180

Invited Papers

1:30

2pMU1. An automated calibration system for telematic music applications. Jonas Braasch (CA^3RL, School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

The complex polyphonic nature of music typically has stricter requirements for sound-pressure level calibration than back-and-forth natured speech dialogs. An onsite sound engineer typically controls the levels for electroacoustically enhanced music performances. In telematic applications, however, this task can easily become difficult because the sound engineer only has restricted access to the remote sites. This makes it difficult to balance the overall sound. For this reason, an automatic calibration system was developed. The system works with closely captured microphone signals, which are needed to avoid echo-feedback between two or more remote telepresence sites. For calibration purposes, a small-aperture pyramidal five-microphone array is mounted in the center of each remote site. The microphone array tracks the positions and sound-pressure levels of the individual instruments, based on an algorithm that can extract this information in the presence of concurrent sound sources [Braasch and Tranby, 19th ICA, Madrid, Spain (2007)]. The extracted data are then transmitted via OpenSound Control to the remote site(s) for accurate spatial reproduction of the original sound-source positions. The signals are calibrated such that they evoke the same sound-pressure levels at the microphone array of each site.

1:55

2pMU2. Synchronization and acoustics in network performance. Juan-Pablo Caceres (CCRMA, Dept. of Music, Stanford Univ., The Knoll, 660 Lomita Dr., Stanford, CA 94305, jcaceres@ccrma.stanford.edu)

The limitations that transmission delays impose on telecommunications are explored and used to the advantage of music making in network performances. This presentation discusses several strategies and practical applications that the SoundWIRE group at Stanford University has implemented in recent years. Performances with Belfast, New York, and Beijing are showcased. Strategies to synchronize remote musicians show how one performance can generate separate instances of a piece in different locations. The Internet's acoustic path (i.e., the delay in the sound transmissions) is used as part of musical effects and physical models that are embedded in the network. These processes make the distance "audible" for performers and support (and sometimes extend) musical ideas that otherwise would be impractical. Finally, a project with an application to attest the quality of service (QoS) in an audio connection is presented. Remote clients can "sound-ping" to a server running the JACKTRIP application and get an idea of its quality. Different physical model instruments can be used. The longer the distance, the lower the pitch. With higher variability in the connection, vibrato is heard. This technique enables musicians to tune a link using the most appropriate and simplest tool, their ears.

2:20

2pMU3. Haptic communication and colocated performance. Curtis Bahn, Jonas Braasch, Dane Kouttron, Kyle McDonald, and Pauline Oliveros (Dept. of the Arts, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY, 12180, crb@rpi.edu)

How can experimental computer-mediated performance attain the very high levels of interpersonal intimacy and temporal intensity attained over millennia of artistic evolution involving traditional analog instruments and affordances? Current practice is lamentably missing key components of a hypothetical possibility space combining enhanced sensory perception with embodied skills inherited from traditional art forms. This presentation examines the development of the "Vibrobyte," a small inexpensive wireless interface for the exploration of haptic communication in computer mediated and colocated performance. The Vibrobyte was developed through a research seminar at Rensselaer Polytechnic Institute and is in its first stages of artistic application. We will present the technical capabilities of the new device and examples of its first applications. Future possibilities and plans for large-scale musical compositions incorporating the Vibrobyte will be introduced.

2:45

2pMU4. Creating systems for collaborative network-based digital music performance. Doug Van Nort (Music Technol., McGill Univ., 555 Sherbrooke St. West, Montreal, QC H3A 1E3, Canada)

The internet has proven to be an important catalyst in bringing together musicians for remote collaboration and performance. Existing technologies for network audio streaming possess varying degrees of technological transparency with regard to allowable bandwidth, latency, and software interface constraints, among other factors. In another realm of digital audio, the performance of "laptop

music" presents a set of challenges with regard to human-computer and interperformer interaction—particularly in the context of improvisation. This paper discusses the limitations as well as newfound freedoms that can arise in the construction of musical performance systems that merge the paradigms of laptop music and network music. Several such systems are presented from personal work created over the past several years that consider the meaning of digital music collaboration, the experience of sound-making in remote physical spaces, and the challenge of improvising across time and space with limited visual feedback. These examples include shared audio processing over high-speed networks, shared control of locally generated sound synthesis, working with artifacts in low-bandwidth audio chat clients, and the use of evolutionary algorithms to guide group improvisations.

3:10—3:30 Break

3:30

2pMU5. Historical perspectives and ongoing developments in telematic performances. Scot Gresham-Lancaster (Media and Technol. Services, California State Univ. Hayward, 25800 Carlos Bee Blvd., Hayward, CA 94542)

This paper presents historical perspective on development 8232, and new technology applications for performances in which the author collaborated with several other dancers, musicians, and media artists to present synchronized colocated concerts at two or more sites. This work grew out of the author's participation in the landmark computer music ensemble, "The HUB." Each of the various performances was made possible by an evolving array of videoconferencing hardware and software. These will be discussed, as well a look back at the early history of "telematic" performance. The problems and interesting side effects presented by latency and dropouts are a unique part of this performance practice. This work leverages the concepts of shared space, and video and audio feedback generate with evolving forms created by the combinations of the space, the sounds, and the movements of the participants. The ubiquity of broadband Internet connections and the integration and constant improvement of videoconferencing software in modern operating systems make this unique mode of performance an essential area of research and development in new media performance. These new techniques will be discussed as well.

3:55

2pMU6. Telematic performance in mixed realities: Virtual and acoustic spaces. Pauline Oliveros (Arts Dept., Rensselaer Polytecnic Inst., 110 Federal St., Troy, NY 12680, olivep@rpi.edu)

Performers are always dealing with latencies. Some of the smaller latencies are long forgotten and integrated into performance practice. Spaces new to performers always present challenges. Halls can be too dry giving the performer insufficient feedback to create a warm instrumental sound or in the other extreme too wet so that the clarity of the sound is blurred in the reverberations. Networked performances between two or more locations present new challenges and latencies. The mixed realities of networked performance spaces will be explored in this paper. How can performers adjust to receiving from and projecting sound into more than one space simultaneously? Where is the focus point when two or more spaces meet in a network? How can the ambiguity of simultaneous spaces provide creative inspiration for new acoustical forms?

Contributed Papers

4:20

2pMU7. Constructing an integrated system for the practice of playing the guitar. Mamoru Ichise (Graduate School of Sci. and Technol., Ryukoku Univ., Japan), Norio Emura (Doshisha Univ., Japan), Masanobu Miura (Ryukoku Univ., Japan), and Masuzo Yanagida (Doshisha Univ., Japan)

Constructed here is an integrated system for the practice of playing the guitar for novice guitar players. Novice guitarists for popular music often give up playing it because they cannot keep practicing for the difficulty of realizing finger positions. An algorithm YG for giving optimal chord-form sequence for given chords was proposed [Emura *et al.*, AST, 104–107 (2006)]. Authors construct an integrated system for guitar practicing by implementing YG. The new system deals with thousands of MIDI files combined with names of chords for them. The proposed system shows chord sequences as the TAB score, with corresponding fingerings generated by YG as the chord diagram. By using the system, users are able to play excerpts using not the difficult but simple pattern of chord-form. The system interacts with the user playing the MIDI guitar by showing the performed notes with an accompanying playback for current excerpt. The user is expected to be supported on playing it under a performance evaluation, where errors of fingerings are simultaneously shown with accompanying playback stopping or slowing down. The performance of the system is confirmed as better than the usual style of practicing in terms of occurrence frequency of errors. [Work supported by HRC, Ryukoku University.]

4:35

2pMU8. New musical interface for playing the guitar solo. Sangjin Cho (School of Elec. Eng., Univ. of Ulsan, Ulsan 680-749, Korea, sjcho75@ulsan.ac.kr), Myeongsu Kang, and Uipil Chong (Univ. of Ulsan, Ulsan 680-749, Korea)

This paper describes the development of a new musical interface to play guitar solo with the nonstringed guitar. The previous work of authors is the implementation of the nonstringed guitar comprised of laser strings and a chord glove [J. Acoust. Soc. Am. **122**, 3055 (2007)]. It is hard to express guitar solo, so voltage dividers are set up on the neck to represent frets. In the case of the chord glove, the same voltage source of the TMS320F2812 is supplied instead of attaching force sensing registers to the fingertip. Consequently, there are three essential components for the proposed interface: voltage dividers, laser strings, and the chord glove. Voltage dividers replace the six strings on the neck and send fret information to DSP, laser strings represent strokes, and the chord glove expresses playing chords and solo. This proposed interface represents many kinds of playing styles. In the case of hammering on, pulling off, and sliding, it needs fret information from voltage divider to express these things. On the other hand, bending that varies the pitch in a fret needs to detect how much flex sensors are bent to play. [Work supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD)(KRF-2006-521-H00002)]

Session 2pPA

Physical Acoustics: Topics in Nonlinear Acoustics

Zhiqu Lu, Cochair
Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

James M. Sabatier, Cochair
Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

1:00

2pPA1. Higher-order statistical analysis of nonlinearly propagated broadband noise. Micah R. Shepherd and Kent L. Gee (Dept of Phys. and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602)

Nonlinearity indicators have recently been used to identify and characterize nonlinear effects present in high-amplitude noise propagation. Although frequency-domain approaches have many benefits, a time-domain approach may be a more natural fit for the time-domain phenomenon of wave form steepening and shock formation. Standard skewness and kurtosis metrics are computed for propagated Gaussian noise at both low and high amplitudes. The same higher-order statistics are also computed for the first time derivative of the propagated noise in order to accentuate non-Gaussian variation in the wave form. The results appear to reveal an asymptotic behavior of the statistical quantities as the time wave form becomes more shock dominated.

1:15

2pPA2. Ultrasonic wave generation by means of highly nonlinear waves' sensor technology. Piervincenzo Rizzo (Dept. of Civil and Environ. Eng., Univ. of Pittsburgh, 942 Benedum Hall, 3700 O'Hara St., Pittsburgh, PA 15261, pir3@pitt.edu), Devvrath Khatri, and Chiara Daraio (California Inst. of Technol., Pasadena, CA 91125)

This paper describes an innovative approach to generate and sense pulse waves in structural materials. The approach is based on the generation of highly nonlinear solitary waves (HNSWs). HNSWs are stress waves that can form and travel in highly nonlinear systems (i.e., granular, layered, fibrous, or porous materials) with a finite spatial dimension independent of the wave amplitude. Compared to conventional linear waves, the generation of HNSWs does not rely on the use of electronic equipment (such as an arbitrary function generator) and on the response of piezoelectric crystals or other transduction mechanism. HNSWs possess unique tunable properties that provide a complete control over tailoring: (1) the choice of the waves' width (spatial size) for defects' investigation, (2) the composition of the excited train of waves (i.e., number and separation of the waves used for testing), and (3) their amplitude and velocity. HNSWs are excited onto concrete samples and steel rebar. The characteristics of the pulses traveling along simple waveguides such as metallic plates and rods are experimentally studied.

1:30

2pPA3. Nonlinear acoustic phenomena in viscous thermally relaxing fluids: Shock bifurcation and the emergence of diffusive solitons. Pedro Jordan (U.S. Naval Res. Lab., Stennis Space Ctr., Code 7181, MS 39529)

In this talk, we will consider the propagation of finite-amplitude acoustic waves in fluids that exhibit both viscosity and thermal relaxation. Under the assumption that the thermal flux vector is given by the Maxwell–Cattaneo law, which is a well known generalization of Fourier's law that includes the effects of thermal inertia, we derive the weakly nonlinear equation of motion in terms of the acoustic potential. We then use singular surface theory to determine how an input signal in the form of a shock wave evolves over time and for different values of the Mach number. Then, numerical methods are used to illustrate our analytical findings. In particular, it is shown that the shock amplitude exhibits a transcritical bifurcation; that a stable nonzero equilibrium solution is possible, and that a Taylor shock (i.e., a diffusive soliton), in the form of a "tanh" profile, can emerge from behind the input shock wave. Finally, an application related to the kinematic-wave theory of traffic flow is noted. [Work supported by ONR/NRL funding (PE 061153N)].

1:45

2pPA4. Intensity and temperature variance in sonoluminescence. Lyric Gillett (1821 Harvard St., Houston, TX 77008, mjcgillett@comcast.net)

Sonoluminescence is the process in which light is produced through ultrasonic pressure waves causing the expansion and subsequent collapse of a gas/vapor bubble, held in an ultrasonic standing wave pattern, in water. This study measured sonoluminescent bubble light intensity output utilizing a photomultiplier tube and a lock-in amplifier to determine the relationship of the sonoluminescent intensity to variance in water temperature. The results could be indicative of the optimal temperatures at which to conduct sonoluminescence research, and assist in facilitating and advancing sonoluminescence research ventures relating to further study of energy production and nuclear reactions. Though not part of the experimental design, the initial experimentation suggested a positive correlation between the drive level of the transducer and the light intensity from the sonoluminescent bubble. Further experimentation was conducted in which the transducer drive signal and frequency were continuously optimized as the water temperature rose. After data were graphed and analyzed, it was determined that there was a decrease in sonoluminescent intensity as the water temperature increased. It is believed that this experimentation resulted in a greater plasma density inside the sonoluminescent bubble. To that end, future experimentation, as well as potential military and medical applications, interests this researcher.

2:00

2pPA5. Nonlinear resonance frequency shifts in acoustical resonators with varying cross sections. Yurii A. Ilinskii, Mark F. Hamilton, and Evgenia A. Zabolotskaya (Appl. Res. Labs., The Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

The frequency response and nonlinear resonance frequency shift for an acoustical resonator with losses and having a varying cross section were investigated previously using Lagrangian mechanics for resonator shapes that are close to cylindrical [J. Acoust. Soc. Am. **110**, 109 (2001)]. The same approach is extended to include resonators having any shape for which a one-dimensional Webster-type equation is a valid model in the linear approximation. Admissible shapes include cones and bulbs proposed for acoustical compressors. The approach is appropriate for approximate but rapid parameter estimations for resonators with complicated shapes, requiring 100 times less computer time than for direct numerical solution of the one-dimensional model equation ordinarily used for these resonators [Ilinskii *et al.*, J. Acoust. Soc. Am. **104**, 2664 (1998)]. Results for cone and bulb shaped resonators with losses are compared with results from the previous one-dimensional theory and with experimental data. It is shown that the di-

rection of the resonance frequency shift is determined by the efficiency of second-harmonic generation in modes having natural frequencies below versus above the frequency of the second harmonic, and how the net effect of this coupling compares with the frequency shifts due to cubic nonlinearity and static deformation.

2:15

2pPA6. Prediction and measurement of particle velocities in ultrasonic standing waves. Bart Lipkens, Jason Dionne, Alex Trask, Brian Szczur (Dept. of Mech. Eng., Western New England College, 1215 Wilbraham Rd., Springfield, MA 01119), and Ed Rietman (Physical Sci. Inc., Andover, MA 01810)

A numerical model has been developed to predict particle trajectories in ultrasonic standing waves. The model includes an electroacoustic model that calculates the characteristics of the one-dimensional standing wave as a function of the input voltage to the piezoelectric transducer driving the cavity. Next, the acoustic radiation force is calculated for particles residing within the water filled cavity. Finally, the particle trajectories are calculated through integration of the equations of motion of the particles. Particle translation is achieved through a periodic sweeping of the excitation frequency. Translational velocities of 6-μm-diameter polystyrene spheres are calculated for a 2-MHz standing wave driven by a PZT-4 transducer. In the experiment a cavity is filled with water and polystyrene particles. A PZT-4 transducer operates near its resonance frequency of 2 MHz. Through a periodic sweeping of the frequency the particles are translated away from the transducer face and ultimately clump together near the rigid reflector at the opposite end. The particle translational velocity is a function of the sweep parameters, i.e., frequency range and sweep period. The calculated particle velocities are then compared to the measured velocities. Needle hydrophone measurements are used to characterize the acoustic standing wave.

2:30

2pPA7. Computation of three-dimensional, pulsed, nonlinear acoustic wavefields from medical phased array transducers in very large domains. Jacob Huijssen, Martin D. Verweij (Lab. of Electromagnetic Res., Fac. of Elec. Eng., Mathematics and Comput. Sci., Delft Univ. of Technol., Mekelweg 4, 2628 CD Delft, The Netherlands, j.huijssen@tudelft.nl), and Nico De Jong (Erasmus Medical Ctr., 3015 GR Rotterdam, The Netherlands)

For the optimization and development of medical ultrasound transducers and imaging modalities, the iterative nonlinear contrast source (INCS) method has been developed. This numerical method predicts the nonlinear acoustic pressure field generated by a pulsed plane source with an arbitrary aperture and propagating in a three-dimensional tissuelike medium that extends over a very large domain of interest. The INCS method obtains the acoustic pressure from the nonlinear acoustic wave equation by treating the nonlinear term as a contrast source. The full nonlinear wave field is then found by iteratively solving the linearized wave problem using a Green's function method. By employing the filtered convolution method discussed in a companion paper, accurate field predictions are obtained at a discretization approaching two points per wavelength or period of the highest frequency of interest. In this paper, very large-scale nonlinear field profiles are presented for transducers with cylindrical as well as phased array geometries excited with a pulsed waveform having a center frequency of 1–2 MHz. Comparison with the results obtained from models of reduced complexity shows that in all cases the INCS method accurately predicts the nonlinear field. [Work supported by the STW and the NCF.]

2:45

2pPA8. A filtered convolution method for the numerical evaluation of very large-scale ultrasound problems. Martin Verweij, Jacob Huijssen (Lab. of Electromagnetic Res., Fac. of Elec. Eng., Mathematics and Comput. Sci., Delft Univ. of Technol., Mekelweg 4, 2628 CD Delft, The Netherlands, m.d.verweij@tudelft.nl), and Nico de Jong (Erasmus Medical Ctr., 3015 GR Rotterdam, The Netherlands)

The pulsed ultrasound field in nonlinear biomedical tissue may be computed by the recurrent evaluation of the field of a distributed contrast source in a linear homogeneous background medium. For human organs, the computational domain may easily measure hundreds of wavelengths or periods in each spatiotemporal dimension. Even with today's supercomputers, the evaluation of the transient field from the contrast sources in these very large inhomogeneous domains is a big challenge, and is only feasible if the applied numerical method is extremely efficient in terms of memory usage and computational speed. Using a Green's function method, the critical operation is a spatiotemporal convolution over the entire computational domain. It is shown how a systematic filtering and windowing of the involved functions yield accurate numerical convolutions with a discretization approaching the Nyquist limit of two points per wavelength or period of the highest frequency. This minimizes the storage requirement. Application of fast Fourier transforms for evaluation of the discrete convolution sum renders a method that is also computationally fast. The performance of a Neumann method equipped with these numerical convolutions is shown by presenting the computed transient wave field that traverses a very large inhomogeneous domain. [Work supported by the STW and the NCF.]

Session 2pPP

Psychological and Physiological Acoustics: Complex Sound Perception

Lawrence L. Feth, Chair
Ohio State Univ., Speech and Hearing Sci., 110 Pressey Hall, 1070 Carmack, Columbus, OH 43210-1372

Contributed Papers

3:00

2pPP1. Spectral weights and the transition between two auditory processes. Bruce Berg (Dept. of Cognitve Sci., Univ. of California, Irvine, 3151 Social Sci. Plaza, Irvine, CA 92697-5100, bgberg@uci.edu)

Evidence of broad auditory filters for processing temporal information was obtained with a molecular psychophysical technique. The standard in the 2IFC discrimination task consisted of n equal intensity sinusoids, with frequency separations ranging from 10 to 40 Hz. The signal was an intensity increment of the central sinusoid, which on average was 1000 Hz. Pitch cues were degraded by randomly selecting the center frequency over a 100 Hz range on each stimulus presentation. Single channel intensity cues were degraded with a roving-level procedure. As tones were added symmetrically to the ends of the complex, thresholds increased until a certain bandwidth was reached, beyond which thresholds decreased. Previous results showed that transition bandwidths ranged from 100 Hz to more than 1000 Hz across individuals [B. G. Berg, J. Acoust. Soc. Am. **121**, 3639–3645 (2007)]. This discontinuity presumably represents a change in the underlying auditory processes, with envelope cues at bandwidths narrower than the breakpoint and across channel level comparisons for wider bandwidths. Findings show that the pattern of spectral weight estimates before the breakpoint is consistent with an envelope processing model, whereas the weights for bandwidths greater than the transition point were consistent with an across-channel comparison. [Work supported by ONR].

3:15

2pPP2. Matching the temporal window and the fine-structure. William M. Hartmann (Dept. of Phys. and Astronomy, Michigan State Univ., East Lansing, MI, hartman2@msu.edu)

Research in psychological and physiological acoustics requires careful stimulus generation. This is especially true for research in binaural hearing because of the importance of interaural phase relationships. Knowing the stimulus precisely becomes even more important when the experiment employs reproducible noise because there is less averaging over stimulus variations. The discrete Fourier transform can be used to generate noise bursts with precise timing. Although the bursts are not band limited, the amplitude and phase spectra are well characterized within any chosen band if the fine-structure is matched to the temporal window. For the example, with a 100-ms rectangular window, the components are taken to be integer multiples of a fundamental frequency of 10 Hz. With this ideal fundamental frequency, the long-term spectrum is predictable. If the window is not rectangular, the ideal fundamental frequency depends on the temporal parameters of the window. The ideal frequency can be found from the Fourier transform of the window itself. For trapezoidal and Hann (raised-cosine) symmetrical windows, the ideal frequency turns out to be the reciprocal of the total (nonsilent) duration minus the rise time (only). Windows that are not symmetrical about their center lead to phase shifts and should be avoided.

3:30

2pPP3. Rapid efficient encoding of correlated complex acoustic properties. Christian E. Stilp, Timothy T. Rogers, Stephanie L. Jacobs, and Keith Kluender (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706, cestilp@wisc.edu)

Early sensory processing is long thought to extract and exploit redundancy and regularity in the input, but perceptual evidence supporting this hypothesis has been limited. Experiments are reported demonstrating efficient encoding of multiple acoustic attributes resulting in the reduction of perceptual dimensionality. Stimuli varied across two complex independent dimensions: attack/decay (AD) and spectral shape (SS). A subset of 24 stimuli captured a near-perfect correlation between these dimensions ($r^2 = 0.95$). Participants passively listened to 25 presentations of these correlated stimuli over 7.5 min before discriminating sound pairs in AXB trials without feedback. For trials where both SS and AD varied, listeners successfully discriminated sounds consistent with the correlation of SS and AD during exposure, but they were at chance discriminating pairs varying orthogonal to the correlation presented in exposure. After exposure, participants were also significantly poorer at detecting differences among stimuli varying in only SS or AD. Thus, passive listening to stimuli with correlated attributes resulted in the collapse of two physically and perceptually distinct stimulus dimensions onto a single dimension that efficiently encoded exposure covariance. Data are consistent with Hebbian models of cortical adaptation but contradict models derived from anti-Hebbian and competitive-learning principles. [Work supported by NIDCD.]

3:45

2pPP4. Timbre perception. Lydia Hennig (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany, lydia_muwi@web.de)

An electroencephalographic study dealing with timbre perception has been performed. Timbre is next to tone and chroma is the basic element of sound and is perceived as a multidimensional space in which perceptual coordinates correlate with acoustic parameters. To show these correlations and their associations with different brain areas, sounds all constant with respect to their fundamental pitch of 311 Hz were varied in timbre. Due to these variations of acoustical parameters clear differences could be found concerning the perceptional pathway and corresponding brain areas. The modification applied to the sounds included variations of attack time, spectral center of gravity, and even harmonic attenuation. Next to other findings a hierarchy of perception was found where variations of the attack time was overwritten by the even harmonic attenuation and the spectral centroit variations, but only if all three were varied simultaneously. In all other cases the attack time variations caused clear perceptional delays in the prefrontal lobe.

4:00—4:15 Break

4:15

2pPP5. Assessment of a computational model for auditory perception for rotary-wing aircraft using human responses. Evelyn Hoglund, Nandini Iyer, Douglas Brungart, Frank Mobley, and John Hall (Air Force Res. Labs., 2610 Seventh St., Bldg 441, Area B, Wright Patterson Air Force Base, OH 45433, evelyn.hoglund@wpafb.af.mil)

Traditional auditory perceptual models for detection of complex signals against complex ambient soundscapes are based on the human audibility threshold imposed upon computed representations of auditory critical band filters. Such models attempt to locate a positive signal-to-noise ratio in any critical band and then apply classic signal detection theory to derive detectability measures (d') and probability of detection values for the event. One limitation to these models is the limited experimental validation against human sound jury performance, especially using very low-frequency target

signals such as helicopters. This study compares computational auditory detection model predictions against a corresponding large sample of human sound jury data points obtained in the laboratory. Helicopter and ambient soundscape signals were obtained from high-sensitivity recordings in the field. Playback in the laboratory was achieved under high-fidelity headphones calibrated to accommodate helicopter primary rotor frequencies with minimal distortion above human sensation level. All listeners completed at least 12 000 trials detecting helicopters against rural and urban soundscapes to represent the spectrum of potential environments involved in a real world scenario. Analysis compares the human sound jury performance against a contemporary computational auditory detection model, called "AUDIB," developed by the U.S. Army and NASA.

4:30

2pPP6. Simulating the precedence effect by means of autocorrelation. Jonas Braasch (CA^3RL, School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

A number of algorithms have been developed to simulate the robust localization performance of humans in reverberant conditions, the most successful of which are based on (contralateral) inhibition. While these models can demonstrate the precedence effect in general, their nonlinear behavior makes it difficult to optimize their settings. A linear algorithm has now been developed to overcome this limitation. An autocorrelation algorithm determines the delay between lead and lag and their amplitude ratio for both channels. An inverse filter is then used to eliminate the lag signal before it is localized with a standard localization algorithm. Interestingly, the filter contains both inhibitory and excitatory elements, and the filter's impulse response looks somewhat similar to the response of a chopper cell. The algorithm operates robustly on top of a model of the auditory periphery (gammatone filterbank and halfwave rectification). Due to its linear nature, the model performs better if the full waveform is reconstructed by subtracting a delayed version of the halfwave-rectified signal, with a delay time that corresponds to half the period of each frequency band's center frequency. The model is able to simulate a number of experiments with ongoing stimuli and performs robustly with onset-truncated and interaural-level-difference based stimuli.

4:45

2pPP7. Binaural loudness summation by cochlear implant users. Monika Kordus and Richard S. Tyler (Dept. Otolaryngol.-Head Neck Surgery, Univ. of Iowa Hospital Clinics, 200 Hawkins Dr., Iowa City, IA 52246, monika-kordus@uiowa.edu)

The loudness summation in people with bilateral and unilateral cochlear implants (CIs) was investigated. The stimuli were 1/3-octave bands of noise centered at frequencies of 0.25, 1, and 4 kHz presented frontally from a loudspeaker in an ascending or random sequence of level. Subjects rated different intensity levels between the threshold of hearing and level of discomfort on scales from 0 to 100. Eight intensity levels were used for normally hearing subjects who served as a control group. For the CI users, results showed difference in loudness between monaural and binaural conditions with substantial amount of binaural loudness summation. The shape of the intensity-loudness function varied between monaural and binaural conditions, depending on the listener. The power function of control group exhibited an exponent of near 0.5. In binaural condition, stimuli of the same sound pressure level were about 1.8 louder than in monaural condition. For 0.25- and 4-kHz bands of noise, no consistent differences between monaural and binaural thresholds were observed. For a 1-kHz band, the binaural loudness ratings were about 3% and 9% lower for mid- or high-intensity levels as compared to the monaural condition. The results suggest future application for fitting bilateral devices. [Work supported by NIH 5 P50 DC00242.]

TUESDAY AFTERNOON, 11 NOVEMBER 2008 LEGENDS 7, 1:30 TO 4:30 P.M.

Session 2pSC

Speech Communication: Prosody (Poster Session)

Rahul Shrivastav, Chair
Univ. of Florida, Comm. Sci. and Disorders, Dauer Hall, Gainesville, FL 32611

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbers papers will be at their posters from 3:00 p.m. to 4:30 p.m.

2pSC1. Passive exposure to speech stimuli facilitates within-category discrimination. Laurent Bonnasse-Gahot (Ctr. d'Analyse et de Mathématique Sociales (CAMS, UMR 8557 CNRS-EHESS), Ecole des Hautes Etudes en Sci. Sociales, 54 Bd. Raspail, Cedex 06, F-75270 Paris, France) and Jessica Maye (Northwestern Univ., Evanston, IL 60208)

Phonetic categories are known to induce good discrimination of stimuli from different categories and poor discrimination within category [Liberman *et al.*, J. Exp. Psych. **54**, 358–368 (1957)]. In addition, passive exposure to a phonetic continuum improves discrimination if stimuli are presented according to a bimodal distribution of occurrence frequency, whereas a unimodal frequency distribution results in decreased discriminability [Maye *et al.*, Dev. Sci. **11**, 122–134 (2008)]. The present study examined whether passive exposure to bimodal versus unimodal frequency distributions results in a categorical pattern of perception. Stimuli were tokens of the syllable [ma] with rising pitch contour, where pitch slope varied across the continuum. Participants completed an AX discrimination task both before and after passive familiarization to either a unimodal or bimodal distribution of the continuum stimuli. At post-test participants in the bimodal condition showed a significant improvement in discrimination and outperformed participants in the unimodal condition. However, contrary to the predictions of categorical perception, the bimodal group's discrimination improved most for stimuli that did not cross the category's boundary. These data suggest that short-term changes in speech perception after passive exposure may not reflect phonetic category learning but rather enhanced encoding of frequently occurring sounds.

2pSC2. Temporal structure and syntactic disambiguation across lexical biases. Yoon-Shil Jeon, Amy J. Schafer, and Victoria B. Anderson (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Honolulu, HI 96822, yoonshil@hawaii.edu)

The present study investigates prosodic cues across three lexical biases in the production of syntactically ambiguous sentences such as "Walkers and runners with dogs use the outer portion of the park," in which "with dogs" can modify either the second conjunct alone (low association) or both conjuncts (high association). A phrase-combination task induced speakers to

produce high- and low-association tokens for each of 48 test sentences, split across three pretested lexical biases (high bias, equibias, and low-association bias) that were controlled for the number of syllables in the critical region. Results showed a significant lengthening difference between association conditions in the second noun, while the lengthening effect was minor in the first noun. In the high-association condition, the second noun region was significantly longer than the same region in the low-association condition. This is consistent with a stronger prosodic boundary at the end of the second conjunct for high-association productions than for low-association ones. The interaction of lexical bias and association was not statistically significant, suggesting that speakers' tendency to disambiguate through durational differences was not significantly affected by lexical bias.

2pSC3. On the reliability of rhythm metrics. Amalia Arvaniti, Tristie Ross, and Naja Ferjan (Dept. of Linguist., UCSD, 9500 Gilman Dr., La Jolla, CA 92093-D108, amalia@ling.ucsd.edu)

In the past decade or so, various metrics of vocalic and consonantal variability have been used to quantify linguistic rhythm, often yielding disparate results. The reliability of several such metrics (percentage of vocalic intervals and consonantal standard deviation, pairwise variability indices, and variation coefficient) was tested using materials from stress-timed English and German, syllable-timed Spanish and Italian, Korean, an unclassified language, and Greek, which has resisted classification. The materials for each language were divided into three subsets: an uncontrolled subset of sentences excerpted from a representative author of each language, a subset exhibiting as much as possible "stress-timing" properties (complex syllable structure and extensive vocalic variability), and a subset exhibiting as much as possible "syllable-timing" properties (simple syllable structure and limited vocalic variability). The results suggest that rhythmic scores can be severely affected by the choice of materials, especially in languages such as Italian, in which it is easy to avoid or accentuate variability (e.g., by excluding or including geminates). Variation coefficient scores were the most resistant to the manipulation of materials but failed to show statistical differences across most of the languages examined. The overall results cast doubt on the reliability of metric scores as indicators of timing and linguistic rhythm.

2pSC4. Impact of segmentation rules on the rhythm metrics. Diana Stojanovic (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Moore Hall 569, Honnolulu, HI 96822)

Rhythm of speech in the literature of the past decade is quantified using measures of durations: first order, such as mean duration and percentage of the total sample, or second order, such as variance and pairwise variability index. Results of different studies, however, vary even when the same materials and speech mode are used and imply uncertainty of cross-study comparisons. Allen (1978) discussed the issue of validity (the criteria used for segmenting) and reliability (how accurately these criteria can be applied) on duration studies. The current paper examines how different segmentation criteria affect the five most commonly used rhythm metrics. All metrics are computed on the segmented sample of the read story (The North Wind and the Sun) for three speakers each of American English, Indonesian, and Serbo-Croatian. Different segmentation criteria were applied to (1) borders between voiceless plosives and vowels and (2) borders between vowels and trills. As a result, durations of some segments were changed in the range of 10–50 ms. The results show how different metrics are affected by the choice of two segmentation rules at two different speech rates. Implications for cross-linguistic studies are discussed.

2pSC5. A rough task: Defining a measure for rough voice quality. Rahul Shrivastav, David A. Eddins, Sona Patel, and Stacie Cummings (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611)

Perceptual judgments of voice quality on a rating scale or a magnitude estimation task are highly context dependent, making it difficult to compare data from one experiment to another. In previous work, a matching task using a sawtooth wave mixed with speech-shaped noise was found useful for quantifying breathiness in vowels [Patel *et al.*, 2006, JASA, **119**, 3340]. In the present experiment, we attempted to adapt this matching task for the estimation of roughness in vowels. Ten listeners participated in a matching task where listeners compared the roughness of a voice standard and the reference signal. A low-pass filtered sawtooth wave mixed with speech-shaped noise was used as the reference signal. The roughness of this signal was manipulated by amplitude modulating the waveform with a 40-Hz square wave. The modulation depth at which the signal and standard were perceived to be equally rough was used as an index of vowel roughness. Preliminary results show that listeners are able to use the matching task to estimate roughness in vowels, particularly for moderately to severely roughness voices. However, voices with little or no roughness are difficult to match to the signal because of the inherent timbre of the sawtooth.

2pSC6. Phonetic realization of prominence among lexical tones in Mandarin Chinese. Mingzhen Bao and Ratree Wayland (Program in Linguist., Univ. of Florida, Gainesville, FL 32601, joanneb@ufl.edu)

Linguistic prominence is defined as words or syllables perceived auditorily as standing out from their environment Terken (1994). It is explored through changes in pitch, duration, and loudness Ladd (1996). In this study, phonetic realization of prominence was investigated among lexical tones in Mandarin Chinese. The primary concern was to compare how accent and focus are acoustically realized under four conditions: (a) unaccented and unfocused, (b) accented but unfocused, (c) unaccented but focused, and (d) accented and focused, among four tones. Ten native speakers of Chinese were recorded reading materials in a natural fashion. The recorded data were segmented and acoustically measured for acoustic parameters: vowel duration; mean and maximum of intensity; and mean, maximum, minimum, and slope of *F*0. The results showed that vowel duration lengthening was the main acoustic parameter associated with accent, while an increase in vowel duration, mean and maximum of intensity and *F*0, and slope of *F*0 was associated with focus realization. It was also found that acoustic parameters used to realize focus were differentially ranked and varied from tone to tone. These results suggested that phonetic realization of prominence in Mandarin Chinese was affected by category of prominence (i.e., focus or accent) and tonal contexts.

2pSC7. The relationship between open quotient and $H1^*$-$H2^*$. Jody Kreiman (Div. of Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794), Markus Iseli, Juergen Neubauer, Yen-Liang Shue, Bruce R. Gerratt (UCLA School of Medicine, Los Angeles, CA 90095-1794), and Abeer Alwan (UCLA, Los Angeles, CA 90095)

It is widely assumed that changes in open quotient (OQ) produce corresponding changes in $H1^*$-$H2^*$, but empirical data supporting this relationship are scant. To provide such data, high-speed video images and audio signals were simultaneously recorded from six speakers producing the vowel /i/ while varying *F*0 from high to low and voice quality from pressed to breathy. Across speakers, the observed relationship between OQ and $H1^*$-$H2^*$ was much weaker than generally assumed. Patterns of covariation also differed substantially from speaker to speaker. Estimation of harmonic amplitudes was complicated by difficulties in determining the frequency of *F*1 when *F*0 was high and by uncertainties regarding the *F*1 bandwidth in the presence of a persistent glottal chink. Use of analysis-by-synthesis allowed correction of formant values, but bandwidth estimation remains problematic and will be discussed further at the conference. [Work supported by NIH Grant DC01797 and NSF Grant BCS-0720304.]

2pSC8. Music melody perception in tone-language- and nontone-language speakers. Jennifer A. Alexander, Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208), Richard D. Ashley, and Patrick C. M. Wong (Northwestern Univ., Evanston, IL 60208)

Speech and music both utilize pitch variation to convey contrasts in meaning: music-pitch expresses composition (e.g., the key of a piece) and effect; speech-pitch conveys pragmatic meaning and, in tone languages, lexical information. This study investigated whether experience with processing lexical pitch affects music pitch processing. 28 nonmusicians (14 native English speakers and 14 native Mandarin speakers) discriminated (experiment 1) and identified (experiment 2) short melodies. The Mandarin listeners more accurately discriminated the melodies than the English listeners (Mann-Whitney $U=140.5$, $p<0.05$; two-tailed $t(21.86)=2.45$, $p<0.05$, $d=0.93$), but the English listeners more accurately matched the melodies with graphical representations of the pitch movements that the mandarin listeners (Mann-Whitney $U=26.5$, $p<0.005$; two-tailed $t(25.44)=-3.94$, $p<0.001$, $d=1.15$). This indicates that experience with lexical-pitch processing

may enhance attention to pitch and thereby facilitate pitch-pattern discrimination. However, learned linguistic pitch-pattern categories may interfere with novel, nonlinguistic, pitch-patterns, thereby impairing identification of musical pitch-patterns. Our study contributes to a literature suggesting that processing of music-pitch and speech-pitch utilizes shared cognitive mechanisms (e.g., Alexander *et al.*, Interspeech 2005). Results are discussed with regard to a cognitive-processing framework that involves the influence of experientially acquired top-down pitch category information upon novel bottom-up pitch input during certain tasks. [Work supported by NU Cognitive Science Graduate Fellowship to J.A. and NIH Grants DC005794 (to A.B.) and HD051827 DC007468 (to P.W.)]

2pSC9. Similarities in the acoustic expression of emotions in English, German, Hindi, and Arabic. Marc Pell, Silke Paulmann, Chinar Dara, Areej Alasseri (School of Commun. Sci. and Disord., McGill Univ., Montreal, QC H3G 1A8, Canada), and Sonja Kotz (Max Planck Inst. for Human Cognit. and Brain Sci., Leipzig, Germany)

Based on the hypothesis that emotion expression is in large part biologically determined ("universal"), this study examined whether spoken utterances conveying seven emotions (anger, disgust, fear, sadness, happiness, surprise, and neutral) demonstrate similar acoustic patterns in four distinct languages (English, German, Hindi, and Arabic). Emotional pseudoutterances (the dirms are in the cindabal) were recorded by four native speakers of each language using an elicitation paradigm. Across languages, approximately 2500 utterances, which were perceptually identified as communicating the intended target emotion, were analyzed for three acoustic parameters: *f*0Mean, *f*0Range, and speaking rate. Combined variance in the three acoustic measures contributed significantly to differences among the seven emotions in each language, although *f*0Mean played the largest role for each language. Disgust, sadness, and neutral were always produced with a low *f*0Mean, whereas surprise (and usually fear and anger) exhibited an elevated *f*0Mean. Surprise displayed an extremely wide *f*0Range and disgust exhibited a much slower speaking rate than the other emotions in each language. Overall, the acoustic measures demonstrated many similarities among languages consistent with the notion of universal patterns of vocal emotion expression, although certain emotions were poorly predicted by the three acoustic measures and probably rely on additional acoustic parameters for perceptual recognition.

2pSC10. Acoustic correlates of vocal effort. Stephen Tasko (Dept. of Speech Pathol. and Audiol., Western Michigan Univ., 1903 W. Michigan Ave., Kalamazoo, MI 49008-5355, stephen.tasko@wmich.edu), Madalyn Parker (Kalamazoo Area Math and Sci. Ctr., Kalamazoo, MI 49008), and James Hillenbrand (Western Michigan Univ., Kalamazoo, MI 49008-5355)

This study evaluated the influence of open quotient (OQ), fundamental frequency (*F*0), and intensity on listener ratings of vocal effort for isolated synthetic vowels. Glottal waveforms (duration: 1000 ms) were synthesized using a model developed by Rosenberg [J. Acoust. Soc. Am. **49**, 583–590 (1971)]. Signals were generated using six different OQs ranging from 0.2 to 0.7 at each of eight distinct *F*0 values. Four *F*0 values were in a typically male range (100–142 Hz) and four in a typically female range (178–253 Hz). The glottal waveforms were passed through formant resonators with formant frequency settings characteristic of the vowel /a/. Signals were randomly presented at 80, 83, and 86 dBA to 20 listeners who used visual analog scales to rate each sample on dimensions of vocal effort, loudness, pitch, breathiness, and naturalness. Mean vocal effort ratings showed a weak negative association with OQ ($r=-0.34$) and weak-moderate positive associations with F0 ($r=0.45$) and intensity ($r=0.55$). Multiple regression analysis revealed that a linear combination of the three variables accounted for 68% of the variance in vocal effort ratings. These results suggest that the perception of vocal effort relies on multiple acoustic cues.

2pSC11. Perceived age in normal and disordered voices. Dorothy Bourgeois, W.S. Brown, Jr., Rahul Shrivastav, Howard Rothman, and James D. Harnsberger (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611)

An experiment was conducted to identify the contribution of voice quality to perceived age. Voice quality effects were assessed in age estimation experiments using (1) natural pathological stimuli that incorporated voice qualities of interest and (2) young normal voices in which two voice cues, hoarseness and tremor, were modified through resynthesis. The disordered samples included single sentences from 227 talkers included in the Kay Elemetrics database of disordered samples. The resynthesized samples were sentences from ten young males in which (1) the *f*0 contour was multiplied by a random number within a fixed range (hoarseness) or (2) a 5–9-Hz wave with an amplitude of 7.5% of the original *f*0 value was incorporated (tremor). Sixty navive listeners estimated all speakers ages in years. The results demonstrated that disordered voices were overestimated in age by 11 years, with the greatest mismatch in chronologic and perceived age reaching 47 years older. The addition of synthetic tremor and hoarseness shifted perceived age older by 8 and 5 years, respectively. In summary, voice quality appears to play a significant role in the perception of perceived age, comparable to or exceeding other cues that have been identified to date.

2pSC12. Perception of falling and rising pitch contour contrast: Effects of first language background. Ratree Wayland, Mingzhen Bao, and Edith Kaan (Program in Linguist., Univ. of Florida, Gainesville, FL 32607, ratree@ufl.edu)

The main objective of this study is to examine the ability to discriminate falling and rising pitch contour contrasts among native speakers of a tonal language (Chinese) and native speakers of a nontonal language (English). Linearly falling and rising pitch contours on [ba:] syllables are presented to participants in (1) the "same different categorial" discrimination task and (2) an "oddball" detection task. Preliminary results obtained from 10 Chinese and 12 English speakers suggested that (a) native Chinese speakers found the falling contour to be easier to discriminate than the rising contour, while the ability to discriminate between the two pitch contours was comparable among the English speakers, (b) reaction time for the falling contour false alarms was longer than for the rising contour among the English speakers, (c) no difference in either "detection" rate or "reaction" time between the two contours was found among both groups of speakers. More data will be collected from both groups of speakers and from speakers of another tonal language. Language general as well as language specific factors will be considered to account for the patterns of results obtained.

2pSC13. The production and perception of prenuclear second occurrence focus. Melissa Wake, Jason Bishop, and Cathryn Panganiban (Dept. of Linguist., Univ. of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095, mwake@ucla.edu)

Semantic focus in English is typically marked by intonational prominence, most canonically by a pitch accent. One case that has been presented as an exception to this generalization, however, is that of second occurrence (SO) focus. An SO focus is a repeated but focused item, usually associated with a focus sensitive operator such as "only" or "even." Previous studies have suggested that SO foci lack pitch accents for phonological rather than information structural reasons, in most cases examining such foci in the postnuclear domain. We present acoustic and phonological data that demonstrate that SO foci also lack considerably in intonational prominence (particularly in terms of F0) when prenuclear, although they show increased duration. Additionally, the perceptibility of prenuclear SO focus is tested. Perceptually weak prosodic marking would suggest an important role for pragmatics in a focal interpretation, as has been suggested previously for SO focus.

2pSC14. Perceived prosody: Phonetic bases of prominence and boundaries. Jennifer Cole (Dept. of Linguist., Univ. of Illinois, 707 South Mathews, Urbana, IL 61801, jscole@uiuc.edu), Louis Goldstein, Argyro Katsika (Haskins Labs., New Haven, CT 06511), Yoonsook Mo (Dept. of Linguist., Univ. of Illinois, Urbana, IL 61801), Emily Nava, and Mark Tiede (Haskins Labs., New Haven, CT 06511)

In comprehending speech, listeners are sensitive to the prosodic features that signal the phrasing and the discourse salience of words (prominence). Findings from two experiments on prosody perception show that acoustic and articulatory kinematic properties of speech correlate with native listeners' perception of phrasing and prominence. Subjects in this study were 114 university-age adults (74 UIUC + 40 Haskins), monolingual speakers of American English who were untrained in prosody transcription. Subjects listened to short recorded excerpts (about 20 s) from two corpora of spontaneous and read speech (Buckeye Corpus and Wisconsin Microbeam Database) and marked prominent words and the location of phrase bound-

aries on a transcript. Intertranscriber agreement rates across subsets of 17–40 subjects are significantly above chance based on Fleiss' statistic, indicating that listeners' perception of prosody is reliable, with higher agreement rates for boundary perception than for prominence. Prosody perception varies across listeners (both corpora) and across speakers (WMD, where perceived prosody varies for the same utterance produced by different speakers). Acoustic measures from stressed vowels (Buckeye: duration, intensity, *F*1, *F*2) and articulatory kinematic measures (WMD) are correlated with the perceived prosodic features of the word. [Work supported by NSF.]

2pSC15. Perception of contrastive meaning through the L+H*L−H% contour. Heeyeon Y. Dennison, Amy J. Schafer, and Victoria B. Anderson (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Honolulu, HI 96822, linguist@hawaii.edu)

This study establishes empirical evidence regarding listeners's perceptions of the contrastive tune [L+H*L−H%; e.g., Lee *et al.* (2007)]. Eighteen native English speakers heard three types of test sentences: (1) contrastive, "The mailbox was(L+H*) full(L−H%)," (2) positive neutral, "The mailbox(H*) was full(H*L−L%);" and (3) negated neutral, "The mailbox(H*) was not(H*) full(H*L−L%)." The participants first scored them by naturalness, and then typed continuation sentences based on the perceived meaning. Three other native English speakers independently coded the continuations to evaluate participants' interpretations of the test sentences. The results clearly demonstrated that the L+H*L−H% tune generated contrastive meanings (e.g., "…but the mailman took the mail and now it is empty" significantly more often than both the positive and negative neutral counterparts. Moreover, sentences presented in the contrastive tune were perceived as natural utterances. High coder agreement indicated a reliable function of the contrastive tune, conforming to the existing literature based on intuitive examples [e.g., Lee (1999)]. Interestingly, however, the contrastive tune produced the expected contrastive meaning in only about 60% of trials (versus less than 10% contrastive continuations for the other contours). This finding shows that the interpretation of the L+H*L−H% contour is more complex than previously suggested.

2pSC16. Order of presentation asymmetry in intonational contour discrimination in English. Hyekyung Hwang (Dept. of Linguist., McGill Univ., 1085 Dr. Penfield Ave., Montreal PQ H3A 1A7, Canada, hye.hwang@mail.mcgill.ca), Amy J. Schafer, and Victoria B. Anderson (Univ. of Hawaii, Honolulu, HI 96822)

In the work of Hwang *et al.* (2007), native English speakers showed overall poor accuracy in distinguishing initially rising versus level (e.g., L*L*H- H*L-L% vs L*L*L- H*L-L%) or initially falling versus level (e.g., H*H*L- H*L-L% vs H*H*H- H*L-L%) contour contrasts on English phrases in an AX discrimination task. Results not reported in that paper found that it was easier to discriminate when a more complex *F*0 contour occurred second than when it occurred first. Several orders of presentation effects in the perception of intonation have been reported (e.g., L. Morton (1997); S. Lintfert (2003); Cummins *et al.* (2006)] but no satisfying account has been provided. This study investigated these asymmetries more systematically. The order effect was significant for falling-level contrast pairs: pairs with a more complex *F*0 contour last were discriminated more easily than the reverse order. Rising versus level contrasts showed a similar tendency. The results thus extend intonational discrimination asymmetries to these additional contours. They suggest that the cause of the asymmetries may depend more on *F*0 complexity than on *F*0 peak.

2pSC17. Alternatives to *f*0 turning points in American English intonation. Jonathan Barnes (Dept. of Romance Studies, Boston Univ., 621 Commonwealth Ave., Boston, MA 02215, jabarnes@bu.edu), Nanette Veilleux (Dept of Comput. Sci., Simmons College, Boston, MA 02115, veilleux@simmons.edu), Alejna Brugos (Boston Univ., Boston, MA 02215, abrugos@bu.edu), and Stefanie Shattuck-Hufnagel (Res. Lab of Electrons, MIT, Cambridge, MA 02139, stef@speech.mit.edu)

Since the inception of the autosegmental-metrical approach to intonation (Bruce 1977, Pierrehumbert 1980, Ladd 1996), the location and scaling of *f*0 turning points have been used to characterize phonologically distinct *f*0 contours in various languages, including American English. This approach is undermined, however, by the difficulty listeners experience in perceiving differences in turning point location. Numerous studies have demonstrated either listener insensitivity to changes in turning point location or the capacity for other aspects of contour "shape" to override turning-point alignment for contour identification (Chen 2003, D'Imperio 2000, Niebuhr 2008). Even labelers with access to visual representations of the *f*0 encounter similar challenges. By contrast, a family of related measurements using area under the *f*0 curve to quantify differences in contour shape appear more robust. For example, a measure of the synchronization of the center of gravity of the accentual rise with the boundaries of the accented vowel yields 93.9% correct classification in a logistic regression model on a data set of 115 labeled utterances differing in pitch accent type. (L*+H L+H* in ToBI terminology). This classification proceeds entirely without explicit reference to the turning points (i.e., beginning of rise, peak) traditionally used to characterize this distinction.

2pSC18. Comparison of a child's fundamental frequencies during structured and unstructured activities: A case study. Eric Hunter (Natl. Ctr. for Voice and Speech, 1101 13th St., Denver, CO 80126, eric.hunter@ncvs2.org)

This case study investigates the difference between children's fundamental frequency (F_0) during structured and unstructured activities, building on the concept that task type influences F_0 values. A healthy male child (67 months) was evaluated (31 h, 4 days). During all activities, a National Center for Voice and Speech voice dosimeter was worn to measure long-term unstructured vocal usage. Four structured tasks from previous F_0 studies were also completed: (1) sustaining the vowel /ɑ/, (2) sustaining the vowel /ɑ/ embedded in a phrase-end word, (3) repeating a sentence, and (4) counting from one to ten. Mean F_0 during vocal tasks (≈257 Hz), as measured by the dosimeter and acoustic analysis of microphone data, matched the literature's average results for the child's age. However, the child's mean F_0 during unstructured activities was significantly higher (376 Hz). The mode and median of the vocal tasks were respectively 260 and 259 Hz, while the dosimeter's mode and median were 290 and 355 Hz. Results suggest that children produce significantly different voice patterns during clinical observations than in routine activities. Further, long-term F_0 distribution is not normal, making statistical mean an invalid measure for such. F_0 mode and median are suggested as two replacement parameters to convey basic information about F_0 usage.

2pSC19. Effects of acoustic cue manipulations on emotional prosody recognition. Chinar Dara and Marc Pell (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine West, Montreal, QC H3G 1A8, Canada, chinar.dara@mail.mcgill.ca)

Studies on emotion recognition from prosody have largely focused on the role and effectiveness of isolated acoustic parameters and less is known about how information from these cues is perceived and combined to infer emotional meaning. To better understand how acoustic cues influence recognition of discrete emotions from voice, this study investigated how listeners perceptually combine information from two critical acoustic cues, pitch and speech rate, to identify emotions. For all the utterances, pitch and speech rate measures of the whole utterance were independently manipulated by factors of 1.25 (+25%) and 0.75 (−25%). To examine the influence of one cue with reference to the other cue the three manipulations of pitch (+25%, 0%, and −25%) were crossed with the three manipulations of speech rate (+25%, 0%, and −25%). Pseudoutterances spoken in five emotional tones (happy, sad, angry, fear, and disgust) and neutral that have undergone acoustic cue manipulations were presented to 15 male and 15 female participants for an emotion identification task. Results indicated that both pitch and speech rate are important acoustic parameters to identify emotions and more critically, it is the relative weight of each cue which seems to contribute significantly for categorizing happy, sad, fear, and neutral.

2pSC20. Perception of emphasis in urban Jordanian Arabic. Allard Jongman, Sonja Combest, Wendy Herd, and Mohammad Al-Masri (Linguist. Dept., Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044, jongman@ku.edu)

Previous acoustic analyses of minimal pairs of emphatic versus plain CVC stimuli showed that (1) emphatic consonants have a lower spectral mean than their plain counterparts and (2) vowels surrounding emphatic consonants are characterized by a higher *F*1, lower *F*2, and higher *F*3 than

vowels surrounding plain consonants [Jongman *et al.*, J. Acoust. Soc. Am. **121**, 3169 (2007)]. The present perception study explores whether Arabic listeners' recognition of emphasis is based on information in the consonant or vowel. Monosyllabic words were used with emphatic and plain consonants in either initial or final position. By means of cross-splicing, the emphatic consonant or its adjacent vowel replaced the plain consonant or its adjacent vowel. Thirty Jordanian listeners participated in the experiment. On each trial, they indicated which word (emphatic or plain) they heard. Results show that the contribution of consonantal and vocalic information to the perception of emphasis depends on vowel quality: In the context of |a|, listeners seem to make their decision primarily based on the vowel while in the context of |i| and |u|, properties of the consonant carry more weight in this decision. The perceptual data will be compared to the acoustic measurements. [Research supported by the NSF.]

2pSC21. The weighting of vowel quality in perception of English lexical stress. Yanhong Zhang (411 Windsor Court, Ewing, NJ 0868, zhang66@purdue.edu) and Alexander Francis (Purdue Univ., West Lafayette, IN 470907-2038)

Acoustically, English lexical stress is multidimensional, involving *F*0, duration, intensity, and vowel quality. Previous research found that Mandarin speakers had problems using vowel reduction in English lexical stress production. Assuming nativelike perception is a prerequisite to nativelike production for non-native speech, the weight of vowel quality with comparison to that of *F*0, duration, and intensity in Mandarin listeners' stress perception was examined. Mandarin and English listeners judged lexical stress placement in synthesized tokens of desert, in which the first syllable /de/ was varied along vowel quality and each of the other cues depending on the pair of cues in focus. Results showed that both Mandarin and English listeners consistently weighted vowel quality more than the other cues. Vowel quality and duration were treated as combinational cues by both groups. English listeners used both intensity and vowel quality (separately), while Mandarin listeners did not use intensity at all. Findings suggest that Mandarin listeners had a nativelike use of vowel quality for perceiving English stress. However, Mandarin listeners treated *F*0 in a different way from English listeners, possibly owing to the influence of their native tonal background. Implications for the interaction between production and perception in second-language learning will be also discussed. [Work supported by Purdue Linguistics.]

2pSC22. Duration of tone sandhi in Mandarin Chinese. Bei Yang (FLARE, the Univ. of Iowa, PH658, Univ. of Iowa, Iowa City, IA 52240, bei-yang@uiowa.edu)

Tone sandhi is the tonal alternation when they are connected in speech flow. Six types of tone sandhi of disyllabic words in Mandarin Chinese are investigated in this research, including the neutral tone, the dipping tone alternation, /bu/, /yi1/, /yi2/, and double change (both of tone changes in a disyllabic word). The paper explores the duration of these six types and whether the neutral tone can change into the citation tones (level, rising, dipping, and falling tones) if we change the duration of the neutral tone. Five Chinese native speakers participate in the study. Two tasks are used to elicit data. First, speakers are required to read 40 disyllabic words. The second one is an identification task. The duration of eight neutral tones are elongated by acoustic technology. The participants hear the eight processed neutral tones and eight unprocessed neutral tones, and are asked to judge the tone types. The results indicate that the neutral tone has the shortest duration, and the duration of other sandhi tone is shorter than that of the normal citation tones. The data also show that the neutral tone and citation tones can be altered based on duration and pitch.

2pSC23. The development of tonal duration in Mandarin-speaking children. Jie Yang (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, 1 Univ. Station A1100, Austin, TX 78712, thyjessie@mail.utexas.edu), Randy Diehl and Barbara Davis (Univ. of Texas at Austin, Austin, TX 78712)

Previous research found that the duration of segments decreases as children grow older. The development of suprasegmental duration, however, has not been explored. The present study investigated developmental changes in duration of the four Mandarin tones. 5-, 8-, and 12-year-old monolingual Mandarin-speaking children and young adults participated in the study. Tone durations were measured in participants' production of monosyllabic target words elicited by picture identification tasks. The results were as follows (1) For each tone category, tone duration and variability decreased with age: 5- and 8-year-old children showed significantly longer durations than adults. Tone durations in 12-year-old children approximated adult values. (2) Despite longer durations, adultlike duration patterns across tone categories existed in all children: dipping tones were the longest, followed by rising and level tones, with falling tones being the shortest. (3) Duration differences between the rising and dipping tones became larger as children grew older. The results may be indicative of the general maturation of laryngeal control over age. Although 5- and 8-year-old children have already established lexical contrasts of tone, adultlike phonetic norms are still in the process of development. The developmental data also provide support for a hybrid account of speech production from a suprasegmental perspective.

2pSC24. Effects of language background on tonal perception for young school-aged children. Chang Liu and Jie Yang (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, Austin, TX 78712)

Given that Chinese is a tonal language while English is not, the present study investigates how language background affects tonal perception for young children. Tonal identification and discrimination are measured on for three groups of school-aged children (6–7 years old): Chinese-monolingual (CM), English-monolingual (EM), and English-Chinese bilingual (ECB) children. Children's task is to identify and discriminate tone 1 (level), tone 2 (rising), and tone 4 (falling) of a Chinese syllable /ma/, for which the fundamental frequency contour is systematically manipulated from tone 1 to tone 2 and from tone 1 to tone 4. CM and ECB children show typical categorical perception in both identification and discrimination, while EM children cannot identify and discriminate the three tones. These results suggest that learning and exposure experience of the tonal language is critical for children to perceive tonal changes in speech sounds.

2pSC25. Hemispheric processing of pitch accent in Japanese by native and non-native listeners. Jung-Yueh Tu, Xianghua Wu, and Yue Wang (Dept. of Linguist., Simon Fraser Univ., 8888 Univ. Dr., Burnaby, BC V5A 1S6, Canada, jta31@sfu.ca)

It is well established that language processing is left hemisphere dominant. Previous findings, however, indicate that lateralization of different levels of linguistic prosody varies with their functional load as well as listeners' linguistic experience. This study explored the hemispheric processing of Japanese pitch accent by native and non-native listeners differing in experience with pitch, including 16 native Japanese participants, 16 Mandarin Chinese participants whose native language has linguistic tonal contrasts, and 16 English participants with no tone or pitch accent background. Pitch accent pairs were dichotically presented and the listeners were asked to identify which pitch accent pattern they heard in each ear. Preliminary results showed that for all the three groups, the percentage of errors for the left ear and that for the right ear were comparable, indicating no hemispheric dominance. The Japanese group did not reveal left hemisphere dominance, as previously found for linguistic tone processing by native listeners. The performance of Mandarin group infers that tone language background did not significantly affect the lateralization of pitch accent. These findings are discussed in terms of how linguistic function differentially influences the hemispheric specialization of different domains of prosodic processing by native and non-native listeners. [Work supported by the NSERC.]

2pSC26. The effect of weak tone on the *f*0 peak alignment. Seung-Eun Chang (Dept. of Linguist., Univ. of Texas at Austin, 1 University Station B5100, Austin, TX 78712, sechang71@gmail.com)

The *f*0 peak sometimes occurs after the syllable with which it is associated, and the peak alignment varies, depending on several factors such as lexical tone target, neighboring tones, focus, and so forth. This study investigates the effect of weak tones on the alignment of *f*0 peaks with three tone types (i.e., H, M, and R) of South Kyungsang Korean, spoken in the southeastern part of Korea. When three tone types are followed by one or two unstressed suffixes, R was found to have the maximum amount of peak delay and M was found to have the minimum amount, i.e., the peak came in the second syllable, following the R-toned syllable, but the peak came in the syllable following the H-toned syllable. This peak delay was not found for M. Thus, it is argued that the tone alternation patterns in suffixed words are

not random; rather, they systematically reflect the phonetic implementation of each tonal target. For example, the peak is in the final portion of a syllable in R, and it takes more time for the peak to be fully realized. This effect can be clearly observed when the following tone is weak.

2pSC27. Investigating the influence of context frequency on lexical tone perception. Jingyuan Huang and Lori Holt (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15232, jingyuan@andrew.cmu.edu)

Tone languages such as Mandarin use pitch variations to contrast meaning. Within tone languages, large variability exists in the pitch of tones produced by different speakers. However, previous studies of speaker normalization for contour tones have produced inconsistent results; whether speakers rely on context information in tone perception is unclear. The present study intended to provide an unambiguous test of the effect of context on contour lexical tone perception and to explore its underlying mechanisms and sources of information. In four experiments, Mandarin listeners' perceptions of Mandarin first and second (level and rising) tones were investigated with preceding speech and nonspeech contexts. Results indicate that (1) the mean fundamental frequency ($f0$) of a preceding sentence affects the perception of contour lexical tones and the effect is spectrally contrastive: Following a sentence with a higher-frequency mean $f0$, a following word is more likely to be perceived as a low-frequency tone and vice versa; (2) nonspeech precursors also elicit this effect, suggesting general perceptual rather than articulatory-based mechanisms; (3) listeners can use information from both fundamental frequency and periodicity to normalize tone perception. [Work supported by NIH NIDCD 2 R01DC004674-04A2].

TUESDAY AFTERNOON, 11 NOVEMBER 2008 LEGENDS 12, 1:55 TO 4:30 P.M.

Session 2pSP

Signal Processing in Acoustics and Underwater Acoustics: Signal Processing for High Clutter Environments

Ronald A. Wagstaff, Cochair
Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Joal Newcomb, Cochair
NAVOCEANO, 1002 Balch Blvd., Stennis Space Center, MS 39522-5001

Chair's Introduction—1:55

Invited Papers

2:00

2pSP1. Signal processor for detection of signals in cluttered environments. Ronald Wagstaff and Heath Rice (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, rwagstaf@olemiss.edu)

The well-publicized experience of the USS Cole in a foreign port demonstrates the potential danger our military ships are frequently exposed to. When docked, they are the most vulnerable to attack by small fast boats that can hit and run before their threat is recognized, and defenses can be activated. This is a challenge that has many facets. Attack by a small high-powered fast-boat is just one type of threat. However, it is important. One way of quickly identifying a fast-boat is with an underwater acoustic sensing system. Such a system is not without challenges. Harbors are typically busy and contaminated by many forms of acoustic clutter. Detecting and separating the signals of fast-boats from among the clutter are difficult tasks. Fortunately, a signal processor has been developed with highly coherent fast-boat signals, and harbor clutter in mind. This fast-boat processor invokes temporal coherence constraints, by degree, to strip away incoherent noise and less coherent shipping signals, and leaves the fast-boat signals exposed. This signal processor will be discussed, and the results will be presented to illustrate the substantial signal-to-noise ratio and automatic signal detection gains that can be achieved.

2:20

2pSP2. Human detection algorithm for seismic and ultrasonic detectors. Alexander E. Ekimov and James M. Sabatier (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Range detection methodologies for human acoustic signals are discussed. Seismic, passive, and active Doppler ultrasonic sensors are used in the presented methods. The algorithm developed for recognition of human's presence in the measured signals is based on detection of the specific acoustic signatures resulting from specific human motion characteristics. These signatures have two characteristic times, $T1$ (the footstep repetition time, which is equal to the time of the whole body periodic motion) and $T2$ (the footstep duration time). The footstep duration time is equal to the time interval for a single footstep from "heel strike" to "toe slap and weight transfer." Taking advantage of these times in signal processing for optimization of the signal-to-noise ratio and applying a procedure of cadence frequency detection allow recognition of human presence in the analyzed signals. This algorithm was tested and it was experimentally demonstrated that cadence frequencies for walking human and their harmonics for seismic and ultrasonic signatures (passive and active Doppler signatures) were the same. Cadence frequencies were stable and had no detectable variation in time for a regular walking style. This stability feature potentially may be used for human tracking. [Work supported by the Department of the Army, Army Research Office, Contract No. W911NF-04-1-0190.]

2:40

2pSP3. Hardware and software solutions to noise in resonance measurements. Joseph Gladden, III (Dept. of Phys., Univ. of Mississippi, University, MS 38677, jgladden@phy.olemiss.edu)

Resonance measurements have the benefit that the system itself is a natural signal amplifier with a gain proportional to the quality factor (Q) of the resonator. However, even this small respite from Murphy's law is not always enough to coax weak resonance signals out of a noisy background. In this talk I will discuss several aspects of improving signal to noise ratios in resonant ultrasound spectroscopy measurements using both hardware and software based solutions such as custom low-noise preamplifiers, lock-in amplifiers, dynamic averaging during data acquisition, and nonlinear fitting of noisy and overlapping (low-Q) peaks to an analytical model. Some particular issues of signal transmission across high-thermal gradients will also be discussed.

3:00—3:15 Break

3:15

2pSP4. Identifying individual clicking whales acoustically amid the clutter of other clicking whales. George E. Ioup, Juliette W. Ioup, Lisa A. Pflug, Arslan M. Tashmukhambetov (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, geioup@uno.edu), Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA), Charles H. Thompson (NOAA/NMFS/SEFSC, Stennis Space Ctr., MS), and Christopher O. Tiemann (Univ. of Texas at Austin, Austin, TX)

Exceptionally clear recordings of sperm whale (*Physeter macrocephalus*) codas reveal time and frequency properties, which show that clicks within a coda are remarkably similar to each other but that they can differ from clicks in other codas. This is consistent with the hypothesis that individual whales can be identified by the characteristics within their codas. Research has centered on the cluster analysis of these codas to help establish whether acoustic identification of individuals is possible. Recently the cluster analysis has been made more robust. This increases the confidence in the applicability of the approach. Data are now available to couple visual sightings and acoustic tracking with recordings for acoustic identification to give an independent verification of the analysis. Acoustic identification of individuals has also been attempted using isolated echolocation clicks of sperm and beaked whales (*Mesoplodon densirostris* and *Ziphius cavirostris*). Although this is a more difficult and problematic undertaking, there has been some promising cluster analysis. Again data for acoustic tracking and visual observations are now available with the recordings for acoustic identification using echolocation clicks to test and perhaps validate the method. [Research supported in part by SPAWAR.]

3:35

2pSP5. Clutter prediction using artificial neural networks. Juliette Ioup (Naval Res. Lab. at Stennis Space Ctr., MS 39576 and Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, jioup@uno.edu), Maura Lohrenz, and Marlin Gendron (Naval Res. Lab., Stennis Space Ctr., MS 39576)

Clutter is a known problem in electronic geospatial (map) displays, on which many different types of data can be combined and presented as a single image. In this context, clutter may be thought of as an overabundance of information, which reduces display usability by the viewer. To declutter a geospatial display, it is necessary to first predict the amount of clutter a human observer might perceive. Computer-aided classification of maps according to the amount of clutter likely to be perceived by a human viewer is the goal of the research described here. Artificial neural networks are among the possible choices of computing techniques that can be used for this task. The network is trained using prior clutter classifications of training maps made by humans, so that automated prediction of clutter for a new map is possible. The objective is to have the network classify maps according to clutter as perceived by human judges. Several neural network algorithms and trials with data consisting of cluttered map classifications and response times by human judges are described. The choice of input features, including the use of principal components, is discussed. Preliminary results with test maps show good prediction of clutter.

3:55

2pSP6. Mitigation of speckle noise due to laser Doppler vibrometer motion across a vibrating target. Richard Burgett (Planning Systems, Inc., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, burgett@psislidell.com), Vyacheslav Aranchuk, and James Sabatier (Univ. of Mississippi, University, MS 38677)

Laser Doppler vibrometers (LDVs) are widely used for remote vibration measurements. In many applications, such as measurements with an LDV on a moving platform or measurements with a continuously scanning beam, an LDV beam moves across a target. The motion of a laser beam across the target generates noise at the LDV output due to dynamic speckles—speckle noise. Speckle noise is caused by phase fluctuations of laser speckles and by Doppler signal dropouts and increases with the speed of the beam. It has been observed that the noise floor increase due to speckles follows a logarithmic curve as the speed increases. Of course as the speed increases, the dwell time over the target area decreases, which limits the signal of interest. This smaller sampling of the vibration signal, combined with the increase in noise due to speckles, places a severe limit on the scanning speed. A way is being investigated to take advantage of this nonlinear increase in noise with respect to speed by using time multiplexing of multiple LDV beams to increase the length of time the LDV samples the vibrations of the target area, thereby allowing the target signal to be detected above the noise floor.

Contributed Paper

4:15

2pSP7. Acoustic Dopplergram for intruder defense. T. C. Yang (Naval Res. Lab., 4555 Overlook Ave., Washington, DC 20375)

This paper discusses the concept and presents preliminary experimental results using the Dopplergram to detect and localize an underwater vehicle, intended for harbor defense and/or protection of high value assets. The acoustic Dopplergram displays the Doppler frequency of the target echo from an active source as a function of time similar to the Lofargram (or spectral gram), which is widely used in passive sonar for detection of tonal and/or wideband transient signals. *m*-sequence signals, which are sensitive to Doppler shift, are transmitted with a rapid repetition rate from a source and received on a colocated receiver. Target detection is improved using the processing gain of the *m*-sequence and using Doppler discrimination (signal association) by eye-ball integration of the (Doppler) gram data. The target is localized using a two-way travel time and bearing estimation. [This work is supported by the US Office of Naval Research.]

TUESDAY AFTERNOON, 11 NOVEMBER 2008 LEGENDS 8, 1:30 TO 4:00 P.M.

Session 2pUW

Underwater Acoustics: Inversion and Uncertainty

Steven Finette, Chair
Naval Research Laboratory, Washington, DC 20375-5320

Contributed Papers

1:30

2pUW1. A sparse-grid, nonintrusive formulation of acoustic field uncertainty in ocean waveguides. Steven Finette (Naval Res. Lab., Washington, DC 20375-5320) and John Burkhardt (U.S. Naval Acad., Annapolis, MD 21402)

The inclusion of environmental uncertainty in simulations of acoustic wave propagation in ocean waveguides is important for the development of simulation-based prediction methods that quantify the influence of multiple sources of incomplete environmental knowledge on the simulation results. Polynomial chaos expansions have been suggested as a natural mathematical framework for describing both environmental and acoustic field uncertainties, their interaction, and propagation through the waveguide [S. Finette, J. Acoust. Soc. Am. **120** (2006)]. Previous research has described the inclusion of these expansions directly into the propagation equation (the intrusive approach), yielding coupled differential equations for the expansion coefficients. The solution for the coefficients contains the statistical properties of the uncertain field. Here we describe an alternative nonintrusive formulation, where existing acoustic propagation codes can be used to estimate the chaos coefficients rather than solve for them via a complex set of coupled differential equations. The nonintrusive formulation involves multiple solutions of an existing deterministic code, e.g., a wide-angle parabolic equation solver, in conjunction with the Smolyak sparse-grid interpolation and multidimensional quadrature to obtain uncertainty statistics on the acoustic field. [Research supported by the Office of Naval Research.]

1:45

2pUW2. A chaos-based wide-angle parabolic equation model for sound propagation in random ocean. Li Ma and Hao Xing (Inst. of Acoust., Chinese Acad. of Sci., Beijing 100190, China, mary1968@tom.com)

Recently, the chaos-based methods have attracted much attention in ocean acoustics committee. In this paper, a set of partial differential equations (PDEs), which are based on the wide-angle parabolic equation and Wiener-chaos decomposition, was established to investigate the propagation of sound propagation in the ocean environment with an arbitrary sound speed profile. Simultaneously, a numerical treatment of the PDEs, which employed an implicit difference scheme and a spectral method, was developed. By using this method, three cases with different kinds of randomness under a distinct ocean environment are studied numerically.

2:00

2pUW3. Uncertainty and resolution in continuum inversion of ocean bottom geoacoustic properties. Andrew A. Ganse and Robert I. Odom (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, aganse@apl.washington.edu)

Inversion of ocean bottom geoacoustic properties from acoustic receptions in the water column is a nonlinear inverse problem that is inherently unstable and nonunique. One common approach to stabilizing this problem is to assume that the ocean bottom is made up of a small number of layers. The solution from this approach does allow one to reproduce the scattered sound field if all the other experiment parameters such as frequency and geometry are also reproduced. However without extensive prior information about that ocean bottom, this approach yields only one of many equivalent nonunique solutions and may not accurately describe the actual ocean bottom itself. An alternate approach, which may allow one to reuse the results later with a different frequency or geometry, is to use the tools of geophysical continuum inversion to specify the degree of nonuniqueness by quantifying both the uncertainty and limited resolution of the continuum bottom solution. This work compares inversion uncertainty and resolution results for different formulations of the data (e.g., matched field versus matched modes versus waveform), different geometries, and different formulations of the uncertainty (e.g., normally distributed versus including some higher-order moments). [Work partially supported by ONR.]

2:15

2pUW4. Resolution matrix perturbation series applied to a nonlinear ocean acoustic inverse problem. Robert I. Odom and Andrew A. Ganse (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St, Seattle, WA 98105, odom@apl.washington.edu)

The resolution operator for a nonlinear inverse problem is the product of the estimated inverse of the forward model operator and the forward model operator itself. If the inverse of the forward model operator were exact, then the resolution operator would be the identity, but in general the resolution operator describes the (noninvertible) transfer function between the unknown true environmental model and the limited-resolution version, which can be estimated from measurements. The nonlinear model resolution operator can be computed iteratively from the Neumann series representation of an estimate of the inverse of the forward model operator with the assumption that both the data functional and the model perturbation functional possess regular perturbation expansions. An example of a problem that fits these

criteria is normal mode acoustic propagation with "slow enough" perturbations such that the modes adjust adiabatically to the perturbations, and the mode eigenvalues are "far" from cutoff. We examine the effects of the nonlinear components of the model and the higher order components of the model resolution on the reconstruction of the model estimate for a simple ocean acoustic propagation problem. [Work supported by ONR.]

2:30—2:45 Break

2:45

2pUW5. Open ocean seamount geoacoustic inversion. Kevin D. Heaney (OASIS Inc., Fairfax Station, VA 22039)

During the BASSEX-2004 (Basin Acoustic Seamount Scattering Experiment), sound from three sources [(250 Hz LFM, 100 Hz bandwidth) and (75 Hz, m-seq, 50 Hz bandwidth)] was transmitted in a range of 200–250 km to a moving towed array receiver (200 m aperture, cut for 250 Hz) in the vicinity of two seamounts in the central North Pacific. The seamounts rise from a depth of 6000 m to approximately 1000 m below the sea surface. A simple sediment model (soft surficial unconsolidated sediment overlying a hard basement) is used to try and model the energetics of the receptions. Two sets of receptions are used to perform this geoacoustic inversion. The first is a received energy versus range (equivalent to TL versus range) in the shadow behind the seamount. The depth of the received level minimum behind the seamount is sensitive to the sediment cover of the seamount. This approach is limited by the available signal-to-noise ratio at the receiver. The second approach is to use data taken when the receiver is just above the seamount. For this geometry, the beamformer is capable of resolving direct propagating paths and paths that bounce off the seamount up to the array from below.

3:00

2pUW6. Geoacoustic inversion over a basalt seafloor. Kevin D. Heaney and Michael Vera (Univ. of Southern Missisippi, 118 College Dr., Hattiesburg, MS 39406)

Geoacoustic inversion from data just off a volcanic island (Kauai) poses significant challenges. One of the significant challenges is to separate the effects of seafloor roughness and shear wave conversion in the basement. Lava flow fields are known to have significant roughness on several scales relevant to low-frequency acoustic propagation (75 Hz). Basalt is known to have a very high-compressional speed of 3000 m/s and a variable shear speed ranging from 500 to 2000 m/s. The combination of shear and roughness can lead to significant ambiguity while performing a geoacoustic inversion because both mechanisms can lead to attenuation of the coherent field. During the BASSEX 2004 experiment up to 20 transmissions from the fixed NPAL source (75 Hz, m-seq with a 50-Hz bandwidth) were received on a towed horizontal line array at many orientations and ranges from the source. In this paper we look at several short-range transmissions and attempt to sort out the scattering effects of roughness with the attenuation effects of shear wave conversion in the sediment. The two mechanisms will be sorted out using simultaneous narrowband (transmission loss) and broadband inversions.

3:15

2pUW7. Inverting for the properties of an elastic seafloor using complex-density equivalent fluids. Michael Vera (Univ. of Southern Mississippi, 118 College Dr., #5046, Hattiesburg, MS 39406) and Kevin Heaney (OASIS Inc., Fairfax Station, VA 22039)

Receptions were recorded from a bottom-mounted broadband source (with a center frequency of 75 Hz) located near Kauai as part of the basin acoustic seamount scattering experiment (BASSEX). Travel times for arrivals from this source have been modeled at basin-scale ranges using complex-density equivalent fluids for the elastic seafloor material. The collection of acoustic data at shorter ranges as part of BASSEX allows for a more detailed examination of the accuracy attainable by an equivalent-fluid representation of the seafloor. The use of equivalent fluids is intended to depict the conversion of acoustic energy into shear waves. The performance and stability of propagation models can be substantially improved if equivalent fluids can be used to accurately characterize bottom interaction. Simulations have been performed in order to search for the equivalent fluid that best explains received data at a range of a few kilometers. The reflection coefficient of this equivalent fluid can then be matched to parameters of an elastic solid (density, sound speed, and shear speed) for the relevant grazing angles. The ability of these models to reproduce features of the data will serve to indicate the importance of elastic effects in the experimental results. [Work supported by ONR.]

3:30

2pUW8. Time domain integration to improve geoacoustic parameter estimation. Donald DelBalzo (Planning Systems, 40201 Hwy 190 East, Slidell, LA 70461, ddelbalzo@plansys.com), James Leclere, and George Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Shallow-water acoustic predictions are affected by uncertainty in sediment property characteristics. Inverse methods using controlled active source signals have been developed to estimate seabed properties. However, some applications prefer a more covert approach. Our work is focused exclusively on passive inversion techniques using signals from surface ships of opportunity. This study addresses the accuracy of low-frequency (100 Hz) matched-field correlations and subsequent geoacoustic estimates using signals from moving surface ships with unknown source levels at unknown ranges. A time-staggered technique is employed to reduce ambiguities in thick-sediment descriptions, with ever increasing confidence as time evolves. Matched-field techniques are applied in a simulated shallow-water environment with a single vertical line array and high signal-to-noise ratios. The simulations indicate significant potential for accurate estimates of thick-sediment characterizations of grain size and layer thickness out to ranges of tens of water depths in shallow water, despite moderate mismatch conditions in the environmental model. [Work sponsored by SPAWAR.]

3:45

2pUW9. Attenuation inversions using broadband acoustic sources. Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882), Preston Wilson (The Univ. of Texas at Austin, Austin, TX 78712-0292), James F. Lynch, Arthur Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882)

Low-frequency attenuation data are widely scattered. The spread most likely reflects differences in sediment type, degree of consolidation, layering, as well as other causes of effective attenuation encountered during low-frequency field measurements. Moreover, attenuation varies with depth, and longer wavelengths, i.e., low frequencies, encounter deeper sediments with different properties. The assumption that the attenuation in the sediment is constant with depth often results in interpretations which give sediment attenuations with a frequency dependence different when depth variations are included in the sediment description. Low-frequency attenuation inversions using broadband acoustic data from different field experiments will be presented and compared. Compressional wave attenuation will be estimated using modal amplitude ratios. Variation of modal attenuation coefficients with frequency and depth will be discussed. [Work supported by the Office of Naval Research.]

Session 3aAA

Architectural Acoustics: Special Session Celebrating the Work of Russell Johnson

Damian J. Doria, Cochair
Artec Consultants, 114 W. 26th St., New York, NY 10001

William J. Cavanaugh, Cochair
Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776-3027

Chair's Introduction—8:00

Invited Papers

8:05

3aAA1. Russell Johnson: The Bolt Beranek and Newman years. William J. Cavanaugh (Cavanaugh Tocci Assoc. Inc., 327F Boston Post Rd., Sudbury, MA 01776, wcavanaugh@cavtocci.com)

In 1954, the author and Russell Johnson both joined the architectural acoustics staff of the pioneering research and consulting firm, Bolt Beranek and Newman (BBN), in Cambridge, MA. Just over 16 years later, in 1970, after exposure to some of the brightest talents in the relatively new discipline of applied acoustics, both left BBN to form their own independent consulting firms: Cavanaugh, a general consulting practice, and Johnson, which was to become an internationally renowned firm, providing comprehensive services on all phases of performance facility design including acoustics. In the mid-1990s, the author began the process of nominating his long time friend and colleague for the Acoustical Society's prestigious Wallace Clement Sabine Medal which was awarded in 1997 with the citation "for contributions to the understanding of the acoustics of performance spaces and the design of concert halls, theaters and opera houses throughout the world." This paper traces Russell Johnson's developing performance facility design interests from his early pre-WWII days, his US Army wartime service, his postwar studies at Carnegie Tech (now Carnegie Mellon University), and at Yale as well as his "extended" initial committment of two years at BBN to learn all he could about performance space acoustics.

8:25

3aAA2. Tribute to Russell Johnson: W. C. Sabine's personal paper. Leo L. Beranek (975 Memorial Dr. 804, Cambridge, MA 02138, beranekleo@ieee.org)

Russell Johnson was one of the most active acoustical consultants involved with concert hall design. He was a Senior Consultant at Bolt Beranek and Newman in Cambridge before he founded Artec. Sabine's interactions with Mr. Johnson will be featured. In 1975, the notebooks of Wallace Clement Sabine were discovered and in 1979 his consulting files were discovered. Both findings were reported [J. Acoust. Soc. Am. **61**, 629–639 (1977); J. Acoust. Soc. Am. **69**, 1–6 (1981)]. By chance, Sabine's personal papers were discovered recently and highlights from them will be presented with emphasis on his European activities from 1906 to 1917.

8:45

3aAA3. My colleague and competitor: Russell Johnson. Christopher Jaffe (Jaffe Holden Acoust., 114 A. Washington St., Norwalk, CT 06854)

This presentation documents the relationship between two architectural acoustic consultants that were active throughout the second half of the 20th century and into the 21st century. It illustrates the fact that practioners in a competive setting, devoted to the search for truth, can work together to arrive at solutions that enable a discipline to move forward.

9:05

3aAA4. Russell Johnson: Four decades of collaboration. Robert Wolff (239 Highland St., Milton, MA 02186, robertwwolff@gmail.com)

Collaboration of Russell Johnson at Bolt Beranek and Newman, Russell Johnson Associates, and Artec Consultants Inc., from 1966 to 2001, with focus on fundamental approaches to acoustics design and consulting gleaned from experiences with Russell Johnson. Insights into Johnson's beliefs about the relation of acoustics design and consulting to theater design and to buildings for the performing arts will be discussed.

9:25

3aAA5. Russell Johnson's "University of Artec". Robert Essert (Sound Space Design, 2 St. George's Court, 131 Putney Bridge Rd., London SW15 2PA, UK, bob.essert@soundspacedesign.co.uk)

Johnson's legacy was one of passion for holistic design excellence in performance buildings. Architecture and theater were early passions. His interest and skill in acoustical design was a means to influence as much as he could of the auditorium architecture because he cared about quality for the performers and audiences and also because he wanted to express his design aesthetic. I had the privilege of being at Artec from 1980 through 1996, a period when the firm outgrew the home office and emerged into the big time. Although the complexion of the firm changed over those years, Johnson's focus on quality and learning carried through. I will explore these two threads—quality and learning—as he motivated clients, architects, and the Artec staff to do excellent work, with examples of both process and creativity.

9:45

3aAA6. The Ballet Opera House competition and the New Jersey Performing Arts Center: My collaboration with Russell Johnson. Barton Myers (Barton Myers Assoc., Inc., 1025 Westwood Blvd., Los Angeles, CA 90024, mail@bartonmyers.com)

I am honored and pleased to accept Damian Doria's (ARTEC) invitation to speak about my relationship to the late Russ Johnson at the November meeting of the ASA in Miami honoring Russ. I propose to speak about my relationship as an architect working with Russ, the acoustician on two major projects: the Toronto Ballet Opera House competition (unbuilt) and the New Jersey Performing Arts Center (built). One would of course have to add to the numerous anecdotes that helped create his mythical and eccentric reputation.

10:05—10:20 Break

10:20

3aAA7. Auditorium acoustics newsletters, edited by Russell Johnson. Carl J. Rosenberg (Acentech Inc., 33 Moulton St., Cambridge, MA 02138, crosenberg@acentech.com)

From January 1968 to May 1970, Russell Johnson was the principal editor of an internal newsletter at BBN called Auditorium Acoustics. Review of the 43 extant issues presents an interesting perspective on topics of concern and interest to Johnson at that time and reflects on the public attitudes toward acoustics and concert hall design. This paper shares some of those perspectives.

10:40

3aAA8. Acoustical modeling of Russell Johnson's Singapore projects. Gary Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702, gsiebein@siebeinacoustic.com), Robert Lilkendey, Adam Bettcher, Threcia Robinson, and Joshua Fisher (Siebein Assoc., Inc., Gainesville, FL 32607)

The University of Florida Acoustics Lab had the pleasure of working with Russell Johnson and others at Artec in acoustical modeling of several halls, most notably, the Singapore concert and opera rooms. It was truly a pleasure to meet with the talented multidisciplinary staff that Russell put together at Artec to address the development of what can arguably be called the Concert Hall of the 21st Century: a traditional concert room that is amazingly adaptable through reverberation chambers, moving canopies, and adjustable absorption to a wide range of venues and a wide range of architectural styles. The inherently architectural and aesthetic basis Russell used to develop his style inspired a generation of architectural and acoustics students who worked on these projects with us.

11:00

3aAA9. A video retrospective of the life of Russell Johnson. Tateo Nakajima (Artec Consultants Inc., 114 West 26th St., New York, NY 10001, tn@artecconsultants.com)

A video collection of slides set to music from the life of Russell Johnson will be shown. Included are photos from early childhood provided by Johnson's sister, Barbara Mansfield. These depict background and early childhood in Pennsylvania and Johnson's military service during World War II. Later photos from Artec's archives show Johnson with colleagues, friends, and artists throughout his long career in acoustics.

11:30—12:00
Open Mic for Russell Johnson Stories

Session 3aAB

Animal Bioacoustics: Marine Mammal Acoustics in Honor of Sam Ridgway III

Dorian S. Houser, Chair
Biomimetica, 5750 Amaya Dr., #24, La Mesa, CA 91941

Invited Papers

8:00

3aAB1. Building on the shoulders of a giant: Using technology to extend biosonar research pathways first blazed by Sam H. Ridgway. Ted Cranford (Dept. of Biology, San Diego State Univ., 2674 Russmar, San Diego, CA 92123, tcranfor@mail.sdsu.edu)

Underwater biosonar research began in the middle of the last century by a handful of pioneers. Among the debates that raged in the scientific literature for the first 30 years were (1) the origin of biosonar signals and their formation into a transmitted beam and (2) the pathways and mechanisms for sound reception. Ridgway and his colleagues (1980) provided definitive proof for the nasal origin of sonar signals using electromyography and small pressure catheters inserted into various airways. Ridgway also collaborated on a series of experiments that attempted to discern the structure and function of the tympanoperiotic complex. My colleagues and I have combined remote imaging technology, high-speed video endoscopy, and finite element analysis to pinpoint the site of sonar signal generation, posit novel sound propagation pathways into and out of the heads of various odontocetes, and uncover new elements of the structure and function of the tympanoperiotic complex. These results will be presented in light of Ridgway's pioneering work.

8:20

3aAB2. Non anthropogenic deafness in marine mammals: Hearing that is going, going, gone. D. R. Ketten (Biology, WHOI, Woods Hole, MA and Harvard Med. School, Boston, MA 02114, dketten@whoi.edu), J. Arruda (WHOI, Woods Hole, MA and MEEI, Boston, MA), J. O'Malley (MEEI, Boston, MA), S. Cramer (WHOI, Woods Hole, MA), and J. Hammock (Smithsonian Inst., Washington, DC)

In humans, hearing is absent or diminished as a result of congenital defects, aging, noise exposures, traumatic events, and disease. Until concern arose about anthropogenic noise impacts on marine mammals, little was known about the mechanisms or incidence of marine mammal hearing losses. Over the past decade, we have gained substantial information from behavioral and electrophysiologic audiometres, *in vivo* imaging, and postmortem examinations. Dr. Sam Ridgway has been a pivotal element in these investigations, pioneering many of the techniques and facilitating broad collaborative studies. In this paper, the results of computerized tomographic and histologic studies of pinniped and cetacean ears, the majority of which Dr. Ridgway supplied, will be presented. The data show that marine mammals sustain precipitous and progressive hearing loss from multiple etiologies, including labyrinthitis, infestations, trauma, chronic multistage otitis, and presbycusis. In particular, older dolphins and seals develop degenerative pathologies (neural, hair cell, support cell loss, and demineralization) paralleling presbycusic changes in older humans. For captive cases, the distribution of inner ear pathologies will be compared with their premortem hearing curves to demonstrate the feasibility of back-projection hearing analyses on ears from noncaptive animals to better understand hearing status of wild populations. [Work supported by the ONR and the NIH/NIDCD]

8:40

3aAB3. Biomedical imaging for the study of auditory processes in the bottlenose dolphin. Dorian Houser (Biomimetica, 7951 Shantung Dr., Santee, CA 92071, biomimetica@cox.net), Sam Ridgway (Univ. of California, San Diego, CA 92093, ridgway@spawar.navy.mil), Patrick Moore, and James Finneran (US Navy Marine Mammal Program, San Diego, CA 92152)

The use of medical imaging techniques in marine mammalogy has traditionally consisted of anatomical investigations within *postmortem* specimens. Within the last decade, Ridgway pioneered the use of *in vivo* scanning techniques to study physiological processes in the bottlenose dolphin. His work utilizing PET and SPECT imagings to study processes that occur during half-brain sleep and long-term persistent vigilance have opened avenues for marine mammal research not previously realized. Leveraging off of the successes of Ridgway, new work on the processing of auditory signals in the dolphin brain is underway. A dolphin is trained to perform an echolocation task while out of water. Echolocation clicks are captured at the melon surface with a suction-cup embedded hydrophone and synthetic target echoes are returned to the dolphin via a jawphone attached to the pan of the lower jaw. Under this paradigm, the dolphin performs a target discrimination task while positioned in a PET scanner. Injection of radiotracers at the time of the discrimination task will be used to map regions of blood flow and/or metabolic activity during the task. Results will provide the first visualization of regional physiological processes associated with the processing of echoic information in the dolphin brain.

9:00

3aAB4. A Forrest Gump effect? Being "there" for surprises and delights in cetacean and pinniped neurobiology. Sam Ridgway (SPAWAR Systems Ctr. San Diego, Div. 71500, 53560 Hull St., San Diego, CA 92152-5001 and School of Medicine, Univ. of California, San Diego, CA 92093)

Anaesthesia research enabled collaborations with pioneers in animal bioacoustics and neurobiology. Being "there" with Evans, Schusterman, and Scott Johnson stoked interest in hearing. Surgery on the ear with McCormick and Wever helped to elucidate sound conduction. In Wever's laboratory were also Simmons, Fay, and Saunders. (Simmons' bats were there. Bringing giant turtles and dolphins to Princeton was challenging.) With Bullock, recordings were made from auditory areas of the dolphin brain demonstrating separate centers for pulse and whistle responses. Suga helped with similar sea lion research. Sleeping seals transmitted their EEGs from pools atop Harrisons' department in Cambridge. With Woods and Galambos, far field auditory responses were characterized and correlated with the direct responses. Event-related brain potentials were also described. Discovering a deaf/mute dolphin with Don Carder and finding others with presbycusis were a surprise (ears eventually went to Ketten). Having Finneran, Schlundt, and Houser take over marine mammal TTS and sonar issues, as well as sound production work with Moore and Cranford, was delightful. There on Oahu with Mohl, Au, Nachtigall, and Supin when a baby sperm whale came onto Maui for testing again illustrated the Gump effect: "Life was like a box of chocolates. You never know what you're gonna get."

Contributed Papers

9:20

3aAB5. Estimation of the temporal structure of the echo from the fish by using the broadband sonar signal of the dolphin. Ikuo Matsuo (Dept. of Information Sci., Tohoku Gakuin Univ., Sendai 9813193, Japan, matsuo@cs.tohoku-gakuin.ac.jp), Tomohito Imaizumi, Tomonari Akamatsu (Fisheries Res. Agency, Hasaki 7620-7, Kamisu 3140408, Japan), Masahiko Furusawa (Tokyo Univ. of Marine Sci. and Technol., Minato-ku, Tokyo 1088477, Japan), Yasushi Nishimori, and Shinji Ogawa (Furuno Electric Co., Ltd., Nishinomiya 6628580, Japan)

Dolphins can identify prey by using the broadband sonar signals. Behavioral experiments indicate that dolphins can detect and discriminate targets in highly cluttered and noisy environments (Au, (1993)]. The sonar system can be improved by clarifying the ability of the dolphin and this improved sonar system is useful for fish species identification. The echo from the fish contains components resulting from multiple reflections, for example, the swim-bladder and body surface of the fish. Therefore, it is necessary to determine the distance and intensity, corresponding to the reflectivity, of these reflectors. We analyzed the echoes from fishes in seas close to Japan by using the Hilbert transform. First, we analyzed echoes which were measured from three species of the anaesthetized fish in a water tank. It was clarified that both the numbers and intensity of reflectors and the duration of the echo were varied dependent on the species, the individual, and the tilt angle. Second, we analyzed the echo which was measured from the fish suspended by nylon monofilament line in the open sea. It was clarified that the numbers and intensity of reflectors were varied dependent on the species and movement of the fish.

9:35

3aAB6. The dynamics of the bottlenose dolphin sonar beam. Josefin Starkhammar (Dept. of Elec. Measurements and Industrial Elec. Eng. and Automation, Lund Inst. of Technol. LTH, Lund Univ., Lund, Sweden, josefin.starkhammar@elmat.lth.se), Stan A. Kuczaj (Univ. of Southern Mississippi, Hattiesburg, MS), Mats Amundin (Linkoping Univ., Linkoping, Sweden), Johan Nilsson, Monica Almqvist, and Hans W. Persson (Lund Inst. of Technol. LTH, Lund Univ., Lund, Sweden)

Much research on dolphin echolocation has focused on animals that have been trained to remain stationary or to carry a device that allows the animals to move but restricts the location of the sonar beam. In such cases, a small number of hydrophones measures sonar characteristics while dolphins solve echolocation tasks. As a result, much is known about the beam axis but relatively little is known about other parts of the beam. One reason for this disparity is that it has been difficult to interpret the results from off axis measurements using a small number of hydrophones (that may or may not sample simultaneously). In this paper, we report results from a system of 47 hydrophones in a 0.75×0.75 m^2 matrix that allowed measurements to be made at multiple locations in the beam simultaneously, with a sample rate of 1 Msample/s. The system both visualizes and records echolocation clicks in real time across the whole cross section of the beam, hence allowing the full dynamics of the sonar beam to be revealed. As a demonstration of the system's utility, we present results obtained with the system to assess the acoustic properties of clicks produced spontaneously by free-swimming dolphins.

9:50—10:05 Break

10:05

3aAB7. Automated classification of frequency-modulated bowhead whale calls using contour tracing and image segmentation methods. Delphine Mathias, Aaron Thode (Scripps Inst. of Oceanogr., Marine Physical Lab., 9500 Gliman Dr. La Jolla, CA 92037-0238), Susanna B. Blackwell, Charles R. Greene, Jr., and Katherine H. Kim (Greeneridge Sci., Inc., Goleta, CA 93117)

In 2007, directional autonomous seafloor acoustic recorders (DASARs) were deployed at 35 locations over a 280 km swath in the Beaufort Sea in order to monitor potential changes in bowhead whale (*Balaena mysticetus*) locations and/or acoustic activity during seismic exploration activities. The large amount of acoustic data generated motivated the development of computer-aided methods to assist in detecting and classifying bowhead whale calls. In the classification stage, bowhead whale calls were divided into six categories: (1) upsweeps, (2) downsweeps, (3) constant calls, (4) u-shaped and (5) n-shaped undulated calls, and (6) complex calls. The frequency range, duration, and fine structure of individual calls vary considerably, creating difficulties when using match-filtering or spectrogram correlation methods. A manually reviewed test data set was assembled, containing examples from each call category, arranged by signal-to-noise ratio. The data set was then used to test several methods (based on image segmentation techniques) for extracting relevant parameters from the signal for subsequent classification. An optimization procedure was then used to generate receiver operating characteristic curves and thus determine appropriate decision boundaries for optimum statistical classifiers. [Work supported by Shell Exploration and Production Company.]

10:20

3aAB8. Automated detection and localization of nonstereotyped bowhead whale calls in the presence of seismic airgun signals, incorporating multiple directional autonomous recording packages (DASARs). Aaron Thode, Delphine Mathias (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238, athode@ucsd.edu), Susanna B. Blackwell, Katherine H. Kim, and Charles R. Greene, Jr. (Greeneridge Sci. Inc., Goleta, CA 93117)

In 2007 and 2008, 35 directional autonomous recording packages (DASARs) [Greene *et al.*, J. Acoust. Soc. Am. **116**, 799–813 (2004)] were deployed at five sites over a 280 km swath in the Beaufort Sea to monitor the annual migration of the bowhead whale (*Balaena mysticetus*) population during seismic exploration activities. To expedite analysis, automated detection and localization methods are applied to each DASAR data set, employing a four-stage detection and classification algorithm. The first three stages consist of the detection phase, while the final classification stage uses image

thresholding, morphological processing, edge detection, and image segmentation to extract key features of signals from equalized spectrograms. These features are then compared to results from other DASARS at a given site in order to identify a single call across all DASARs. Bearings from each DASAR are then crossed to provide a location. Manually reviewed data sets are used to optimize the algorithm parameters for a given site and a given set of environmental conditions. The total detection and classification performance of the combined DASAR set will be compared with the performance of a single DASAR. [Work supported by Shell Exploration and Production Company.]

10:35

3aAB9. Geographic variation in northwest Atlantic fin whale (*Balaenoptera physalus*) song: Implications for stock structure assessment. Julien Delarue (College of the Atlantic, 48 N. Marriner St., South Portland, ME 04106), Sean Todd (College of the Atlantic, Bar Harbor, ME 04609), Sofie Van Parijs (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543-1026), and Lucia Di Iorio (Universitt Zurich, Rmistrasse 71, 8006 Zurich, Switzerland)

Passive acoustic data are increasingly being used as a tool for helping to define marine mammal populations and stocks. Fin whale (*Balaenoptera physalus*) songs present a unique opportunity to determine interpopulation differences. Their highly stereotyped interpulse interval has been shown to vary between geographic areas and to remain stable over time in some areas. In this study, the structure of songs recorded at two geographically close feeding aggregations in the Gulf of St. Lawrence (GSL) and Gulf of Maine (GoM) was compared. Recordings were made from September 2005 to June 2006 in the GSL and intermittently between January 2006 and September 2007 at two locations in the GoM. 6257 pulse intervals corresponding to 19 GSL and 29 GoM songs were measured to characterize songs from the three locations. Classification trees showed that GSL songs differ significantly from those in the GoM. The results are consistent with those derived from the other stock structure assessment methodologies, such as chemical signature and photoidentification analysis, suggesting that fin whales in these areas may form separate management stocks. Song structure analysis could therefore provide a useful and cost-efficient tool for defining conservation units over the temporal and geographical scales relevant to management objectives.

10:50

3aAB10. Killer whales at the Pribilof Islands: Who are they and what are they doing. Kelly Newman, Alan Springer (UAF School of Fisheries and Ocean Sci., 905 N. Koyukuk Dr., 245 O'Neill Bldg., Fairbanks, AK 99775-7220), and Craig Matkin (North Gulf Oceanic Society, Homer, AK 99603)

The Pribilof Islands, located in the eastern Bering Sea, are the site of the largest concentration of northern fur seals in the world. Despite the occurrence of transient killer whales on a near daily basis during the fur seal breeding season, these whales are unknown in terms of their social structure and overall range. Orcas were recorded with autonomous acoustic recorders at the Pribilof Islands in the summers of 2006 and 2008. This is the first study of call types and whistles of these killer whales. Whistles have been described for other odontocete species but not yet for transient killer whales. This study also reports on temporal and spatial dynamics of killer whale occurrence at the two major Pribilof Islands, St. Paul and St. George Islands. Call matches between the years and locations were analyzed to ascertain visitation patterns, predation sounds, and diurnal activity near the fur seal rookeries. Whale vocalizations from the Pribilofs were compared to other regions in Alaska. Thus far, there is no match between the calls of the Pribilof whales and those in regions of Southeast Alaska and Prince William Sound, but some calls bear a structural resemblance to calls produced by whales recorded near False Pass, AK.

11:05

3aAB11. Localization of killer whale mother, juvenile, and calf vocalizations. Dawn M. Grebner, David L. Bradley, Dean E. Capone, Susan E. Parks, Jennifer L. Miksis-Olds (Grad. Prog. in Acs., Penn State Univ., State College, PA 16804, dmg302@psu.edu), and John K. B. Ford (Pac. Biol. Station, Nanaimo, BC V9T 6N7, Canada)

A dataset of killer whale female (mother), juvenile, and calf vocalizations was collected with a hydrophone array in Johnstone Strait in the summers of 2006 and 2007. The vocalizations were spatially and temporally localized together with behavioral observations. The analysis is focused on pulsed call production variation when a mother is with her juvenile(s) and/or calf and when she is separated from one or more of them. Analysis is centered on the frequency-time structure of the vocalizations. This circumstance provides a unique opportunity to gain insight into the communication between individuals within a common social subset of a matriline or pod. This subset is characteristically tighter in association than many other intramatrilineal groups and can offer a glimpse into early vocal exchange and development.

11:20

3aAB12. Consistency of frequency dependent echolocation beam focus in an echolocating false killer whale (*Pseudorca crassidens*) performing a discrimination task. Laura Kloepper (Hawaii Inst. of Marine Biology Marine Mammal Res. Program, Dept. of Zoology, Univ. of Hawaii at Manoa, P.O. Box 1106, Kailua, HI 96734, kloepper@hawaii.edu), Stuart Ibsen (Univ. of California, San Diego, La Jolla, CA 92093), Paul Nachtigall, and Marlee Breese (Univ. of Hawaii at Manoa, Kailua, HI 96734)

Tremendous variation in individual echolocation click parameters has been shown for odontocetes, making trend analysis of individual clicks within a click train necessary to understand their generalized echolocation behavior. The frequency dependent spatial echolocation beam profile of clicks for a single false killer whale (*Pseudorca crassidens*) performing a discrimination task was collected using a star shaped 16 hydrophone array. These spatial profiles were analyzed to compare the relative spatial location and focusing of various frequencies from click to click under different target exposures. Using a go/no go paradigm, the whale actively discriminated between objects of varying transmission reflectance properties. Very tight focusing and consistent spatial location was observed for frequencies near peak frequency but deteriorated for frequencies above peak frequency. This trend did not vary between targets, suggesting that the animal maintained a constant beam focus pattern regardless of the reflection characteristics of the target. Additionally, these patterns were consistent for all clicks within a train, suggesting that the animal employs a similar beam focus pattern throughout the entire click train.

Session 3aEA

Engineering Acoustics: Acoustics for Battlefield Operations and Homeland Security I

Michael V. Scanlon, Chair
U. S. Army Research Lab., 2800 Powder Mill Rd., Adelphi, MD 20783-1197

Chair's Introduction—9:00

Invited Papers

9:05

3aEA1. Environmental effects on short-range acoustic propagation and sensor performance. Donald G. Albert (ERDC Cold Regions Res. and Eng. Lab., 72 Lyme Rd., Hanover, NH 03755, donald.g.albert@us.army.mil)

Autonomous passive acoustic sensor systems can potentially detect, identify, and locate sound emissions from gunfire, explosions, vehicles, and animals. However, since sound propagation through the atmosphere is influenced by many environmental conditions including meteorology, ground impedance, and barriers or obstacles, these effects need to be understood and included in sensor system design to achieve accurate performance. Low-frequency short-range measurements using impulsive sources will be presented to quantify these effects on acoustic wave form shape and amplitude attenuation. In urban areas, multiple reflections and diffractions can add enough complexity to explosion wave forms that it becomes difficult for a human listener to identify them as explosions. In these areas, classical methods to determine source locations such as beam-forming or direction-of-arrival bearing estimates from time delays can fail completely. Novel signal processing methods that incorporate the complexity of propagation in these environments can produce accurate localizations. [Work funded by U.S. Army.]

9:30

3aEA2. Orthogonal acoustic sensor package for human detection in quiet and noisy environments. James M. Sabatier and Alexander E. Ekimov (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Seismic and low-frequency acoustic sensors are severely limited by background vibrations and sound in noisy (urban) areas and a human's ability to mask the impulsive signal from his or her footsteps. Typically, noise levels are highest at low frequencies and roll off at higher frequencies. Recent research has shown that the ultrasonic signals from footsteps can be passively detected in urban areas and in buildings and are relatively independent of attempts at masking. Coupling this sensor with ultrasonic Doppler sonar and processing algorithms that tie footsteps to corresponding human motion yields an orthogonal sensor package with high probability of detection and minimal false alarms. Coupling the ultrasonic sensor package with the low-frequency traditional sensors (such as seismic sensors) forms the universal acoustic sensor package that potentially increases the probability of human detection in high-noise urban and low-noise rural environments. Successful signal processing algorithms for the sensor package are developed and tested. [Work supported by the Department of the Army, Army Research Office, Contract No. W911NF-04-1-0190.]

Contributed Papers

9:55

3aEA3. Remote detection of vehicle obscured by forest canopy using laser Doppler vibrometer. James Sabatier, Alexander Ekimov, Vyacheslav Aranchuk, and Richard Mack (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, sabatier@olemiss.edu)

Detection and identification of vehicles obscured by forest canopy is a particularly challenging military problem. Imaging techniques (laser radar imaging a target through gaps in foliage, for example) require extensive data, making this approach processing-intensive and time consuming. A new method for detection of a vehicle obscured by forest canopy by remotely sensing the vibration of foliage with a laser Doppler vibrometer has been proposed. The method uses the effect of coupling of sound waves from the vehicle's engine through the air to tree leaves, causing them to vibrate. The presence of a vehicle can be determined by the spectrum of the leaves' vibrations. The feasibility of the proposed method has been studied experimentally. The acoustical impedance of various types of leaves was measured in the laboratory and outdoor settings. The results show that the vibration response of a leaf to the sound from an idling vehicle is high enough to be reliably sensed with an LDV in a wide frequency band. The vibrations of leaves excited with simulated vehicle acoustic stimuli were successfully measured with an LDV in the laboratory and outdoor environment. A real vehicle masked by foliage was successfully detected by the proposed method.

10:10

3aEA4. Acoustic classification of battlefield vehicles based on their seismic detection. Peter William (Dept. of Elec. Eng., Univ. of Nebraska Lincoln, 209N WSEC, Lincoln, NE 68588, peter.ezzat@ieee.org) and Michael Hoffman (Univ. of Nebraska Lincoln, Lincoln, NE 68588)

Identification of battlefield vehicles using a wireless sensor network is an event driven application that requires simple detection and energy efficient classification. Seismic and acoustic sensors have the capability to detect and track targets in a passive non-line-of-sight manner. The proposed approach uses seismic emissions for event detection and acoustic emissions for classification of detected events. Seismic detection using unpowered geophones achieves minimal energy for wakeup and event detection in addition to elimination of wind and background noise events. Time domain extracted features from the acoustic emission of vehicles are very efficient in battlefield vehicle discrimination preserving the coupled harmonic signature for both tracked and wheeled vehicles. We propose a simple energy efficient time domain detection and classification scheme that enables the sensor node to perform detection, feature extraction, and classification. Classifica-

tion of acoustic features is performed using a multilayer feedforward neural network. The proposed approach shows that selective time domain acoustic features surpass equivalent spectral features. Evaluation of the proposed approach using recorded data shows an average classification rate (single node and single event) that exceeds 83% and an average false alarm rate of 16% (over both tracked and wheeled vehicles).

10:25

3aEA5. Localization of sources on the ground from an elevated acoustic sensor array. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305 and Phys. Dept., New Mexico State Univ., Las Cruces, NM 88003), Michael V. Scanlon, Christian Reiff (U.S. Army Res. Lab., AMSRD-ARL-SE-SP, Adelphi, MD 20783), D. Keith Wilson, and Sergey N. Vecherin (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755)

Acoustic sensor arrays suspended below tethered aerostats are employed for localization of sources on the ground. The aerostats can be elevated up to 1–2 km above the ground. Since the values of temperature and wind velocity vary significantly from the ground up to such heights, sound refraction should be taken into account when determining the source coordinates. Previously [Ostashev *et al.*, J. Acoust. Soc. Am. **122**, 3084 (2007)], a theory was developed that enables determination of the source coordinates from the direction of sound signal propagation as measured by the sensor array, its coordinates, and the vertical profiles of temperature and wind velocity. However, these profiles can be known only with some errors and uncertainties. In the present paper, the effects of uncertainties in the vertical profiles of temperature and wind velocity on determining the source coordinates are studied. To this end, the vertical profiles are approximated with the Monin–Obukhov theory of similarity. The uncertainties in the profiles are modeled by allowing the surface roughness length, the friction velocity, and the surface sensible heat flux vary from their mean values. Some of the results obtained are compared with experimental data.

10:40

3aEA6. Impulsive sound source localization using distributed elevated acoustic sensors. Xiao Di, Ronald Wagstaff, John Anderson, and Kenneth Gilbert (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, xiaodi@olemiss.edu)

For detecting and localizing impulsive sources in the daytime, distributed elevated acoustic sensors with large base line separations have distinct advantages over small ground-based arrays. First, during the day, the signal levels are generally higher at altitude than on the ground. Second, the large base lines provide improved localization accuracy. Results are reported from a distributed array of acoustic sensors deployed during an experiment in Bourges, France during June of 2008. The distributed array consisted of microphones and global positioning system (GPS) receivers attached to the tether lines of several widely separated balloons. The sound sources were various impulsive devices. Data from the experiment are presented and discussed. In particular, the localization error is analyzed as a function of the GPS accuracy and the motion of the balloons. Possible ways to improve the localization accuracy are suggested. [Research supported by the U. S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

WEDNESDAY MORNING, 12 NOVEMBER 2008 PLUM A & B, 8:00 TO 11:45 A.M.

Session 3aMU

Musical Acoustics and Structural Acoustics and Vibration: Structural Vibrations in Musical Instruments

Uwe J. Hansen, Chair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Invited Papers

8:00

3aMU1. Modal analysis of musical instruments. Thomas Rossing (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305 rossing@ccrma.stanford.edu)

The complex vibrations of musical instruments can be described in terms of normal modes of vibration. Modal analysis can be done either experimentally or mathematically. We review some of the ways in which modal analysis has been successfully applied to musical instruments in the string, wind, and percussion families.

8:20

3aMU2. Applying electronic speckle pattern interferometry to the study of musical instruments. Thomas Moore (Dept. of Phys., Rollins College, Winter Park, FL 32789, tmoore@rollins.edu)

An overview of the electronic speckle pattern interferometer as a tool for the study of structural vibrations will be presented, highlighting both the theory and construction of the device. Two specific applications to the study of musical instruments will then be discussed: time-averaged interferometry of harmonic motion and time resolved interferometry of transient motion.

8:40

3aMU3. Constructing and using a low cost speckle pattern interferometer for analysis of musical instruments. Andrew Morrison (Phys. Dept., Illinois Wesleyan Univ., P.O. Box 2900, Bloomington, IL 61702-2900, achmorrison@gmail.com)

Following the design of Moore [Am. J. Phys. **72**, 1380 (2004)], a speckle-pattern interferometer for use in exploring the mode shapes of vibrating structures, especially musical instruments, has been constructed. The interferometer was built primarily by undergraduates with mostly common optical elements found in any undergraduate optics laboratory. The interferometer uses a FireWire (IEEE 1394) camera, which has a built-in frame-grabbing capability, eliminating the need for expensive frame-grabbing hardware. The interferometer has been used to study vibrating structures such as snare drum heads, mandolins, and a tennis racket. The interferometer has also been used to study the the vibrational properties of a cured carbon fiber disk. The design and implementation of an affordable speckle-pattern interferometer along with a brief overview of some of the results of research projects conducted with undergraduate students will be presented.

Contributed Papers

9:00

3aMU4. Modes of the kalimba resonator box. Daniel Ludwigsen (Dept. of Phys., Kettering Univ., 1700 W Third Ave., Flint, MI 48504)

The kalimba is an African instrument with plucked metal tines mounted on a resonator box made of hardwood. The box features both acoustic (related to air in the cavity and holes) and structural resonances. An experimental modal analysis of the box indicates some agreement between the frequencies of these structural resonances and quickly decaying spectral components of the attack. Furthermore, direct measurements of mechanical impedance at the bridge show support for the effective radiation of sound at the fundamental frequencies of the tines. A finite element model of the resonator box corroborates these results, which show several ways the design of the box affects a musical and distinctive instrument.

9:15

3aMU5. Drum tuning: An experimental analysis of membrane modes under slightly nonuniform tension. Randy Worland (Dept. of Phys., Univ. of Puget Sound, 1500 N. Warner, Tacoma, WA 98416, worland@ups.edu)

Results of an experimental study of normal mode vibrations in single-headed musical drums under nonuniform tension will be presented. Although uniform tension is often assumed in theoretical treatments, in practice the musical drum only approximates this condition, even after careful tuning by the drummer. This study investigates the behavior of normal mode shapes and frequencies under (slightly) nonuniform tension, as they relate to the tuning process. In particular, the role of the (1,1) mode will be described. Experimental results include electronic speckle pattern interferometry images of modal shapes along with the associated frequencies. A finite element model will be used for comparison with the experimental results.

Invited Papers

9:30

3aMU6. Generalized structural acoustics model of violin radiativity profiles. George Bissinger (Phys. Dept., East Carolina Univ., Greenville, NC 27858, bissingerg@ecu.edu)

Violin radiativity profiles are dominated by contributions from the Helmholtz-like *A*0 cavity mode near 280 Hz, the first corpus bending (*B*1) modes near 500 Hz, followed by a slow rise, peaking near 2.4 kHz, and a slow falloff above this peak. A blend of total damping trends, *f*-hole air motions, bridge-filter effects, bridge versus bridge-island impedances, effective critical frequencies f_{crit}, and radiation efficiency-damping trends was used to create a structural acoustics model for the entire profile. The lowest top and back plate bending modes 2 and 5 subsumed into the *B*1 modes provide nominal values for the violin f_{crit} as well as significantly influencing the *B*1 mode frequencies. *A*0 excitation via coupling of *f*-hole volume flows with those of the *B*1 modes is proposed. This coupling predicts *A*0 radiativity increasing as the *A*0-*B*1 frequency difference decreases, in accord with recent VIOCADEAS data (including violin octet experimental results) and a 1937 experiment involving too-thick/too-thin plate thickness and soundpost removal. Simplified simulations show the fraction-of-vibrational-energy-radiated F_{rad} peaking at f_{crit} (nominally coinciding with the bridge-filter peak). Dropping f_{crit} drops this peak frequency, reduces the maximum F_{rad}, and changes the relative spectral balance in vibration-acoustic energy conversion.

9:50—10:05 Break

10:05

3aMU7. Acoustic holography of musical instruments using radiation directivity method. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstr. 13, 20354 Hamburg, r_bader@t-online.de)

Static and time-dependent acoustic holography is applied to the radiation of musical instruments using the acoustic camera developed at the Institute of Musicology consisting of 128 microphones simultaneously recording at a 48-kHz sample rate. Here a radiation directivity method was developed suiting the great amount of microphones as well as the high sampling frequency. It uses not only the radiation strength but also the radiation directivity pattern of 128 points on the geometry as a dependent variable. This does not only allow for a precise reconstruction of the vibration of the instruments but also results in a detailed description of the radiation directivity for the overall geometry or for each radiating point individually. Furthermore the time-dependent reconstruction avoids any temporal integration and reconstructs the vibration of the geometry with the samplerate precision. With this method the different role and strength instrument parts have in the overall body radiation, scattering effects caused by geometrical inhomogeneities, and time-dependent initial transients reconstructed as radiation patterns were investigated. The material as well as the air parts of the instrument are considered simultaneously, i.e., with guitars or violins, the enclosed air radiating through the sound holes is investigated in relation to the surrounding plates.

10:25

3aMU8. Finite element studies of hand bell modes. Uwe J. Hansen (Dept. of Chemistry Phys., Indiana State Univ., Terre Haute, IN 47809)

Hand bell modes are identified by indices m and n (m, n), where m refers to the number of nodal lines crossing the crown and n refers to the number of circumferential nodal lines. For high m values, depending on the bell size, no $n = 0$ modes exist. For such large m modes, the $n = 1$ nodal line is located very close to the mouth of the bell, rather than near the middle of the bell as is the case for (m,1) modes, where for the same value of m, an $n = 0$ mode exists. These modes are generally labeled as (m, 1^*) modes. Finite element studies of mode sections of increasing complexity indicate that only when two curvatures are introduced, conditions are adequate to give rise to (m, 1^*) modes. Computer generated images and animations of mode patterns corresponding to increasingly complex geometries will be illustrated.

Contributed Papers

10:45

3aMU9. Radiation of guitar loudspeakers. Heiko Timm (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany, heiko_t@web.de)

Loudspeakers of guitar amplifiers used by electric guitarists are analyzed in terms of their radiation characteristics. This was performed using an acoustic camera consisting of 128 microphones. The data recorded were digitalized with a sample rate of 48 kHz and analyzed with methods of near field acoustic holography (NAH) and statistical optimized NAH. By these algorithms an acoustic hologram was obtained. The radiation is visualized to get information about the different parts of the complete electric guitar sound box. Differences in sound between different loudspeaker constructions can be explained by taking this close look at the sound coming out of the speakers. The radiation behavior, the impulse behavior, and the diffraction of sound coming from the backside of the speaker are investigated in this way. So the amount of diffraction of sound influencing the overall character of the speaker is also addressed.

11:00

3aMU10. Measurement and modeling of the open end correction for driven pipes. Daniel Ludwigsen and Linda Hunt (Dept. of Phys., Kettering Univ., 1700 W Third Ave., Flint, MI 48504, dludwigs@kettering.edu)

The end correction that accounts for the inertial impedance at the open end of a pipe is a widely used refinement to lumped-element models, and many variations have been investigated. This study measures the end correction for 1 m closed-open pipes of varying diameters, exposed to a single-frequency external sound field created by a subwoofer in an anechoic facility. The frequency response of each pipe was measured as the ratio of the pressure at the closed end of the pipe to a reference pressure at the open end. Using an expression for the predicted mode frequencies based on the assumption that the driving pressure is uniform across the open end, the driven end correction was derived from the measured frequency response. The interaction of the pipe and the external sound field can be visualized in a finite element model, with alternative estimates of the driven end correction. Preliminary results indicate that the driven end correction can be as large as the diameter, more than three times the expected value for a passive open end.

11:15

3aMU11. Parameter estimation and inversion for wind instrument physical models. Mark Sterling, Xiaoxiao Dong, and Mark Bocko (Dept. of Elec. and Comp. Eng., Univ. of Rochester, Comput. Studies Bldg. 526, Rochester, NY 14623, sterling@ece.rochester.edu)

An outstanding goal in the field of musical instrument physical modeling has been the definition of a robust parametric and expressive representation or coding for musical audio. For this purpose, accurate dynamical models of musical instrument systems must be supported by algorithms that can provide suitable sets of control parameters for producing some desired output. We describe an estimation algorithm for clarinet physical modeling. However, the results could be applied to a broader class of convolutional wind instrument models, which fit the McIntyre–Schumacher–Woodhouse framework. Three fundamental signals in clarinet physical modeling are the blowing pressure p_m, bore pressure p_b, and flow u inside the mouthpiece. These are related via the nonlinear reed-valve characteristic and the air column impulse response. Assuming p_b is given or approximated from a digital recording using known radiation characteristics, p_m and u are iteratively estimated against one another, fitting the model to p_b. This is accomplished with an informed inversion of the reed valve and an adaptive (LMS) inverse of the bore impulse response. This method supplants prior work we have done where blowing pressure waveforms were computed by envelope detection. Results are verified with numerous audio examples. [Work partially supported by NSF Grant IIS-555457.]

11:30

3aMU12. Experimental modal study of the Kurdish long necked lute, Tanbour. Hedayat Alghassi (Dept. of Elec. and Comput. Eng., Univ. of British Columbia, Vancouver, BC V6T 1Z4, Canada), Sohrab Ferdowsi, Roozbeh Alghassi (Tehran Univ., Tehran, Iran), and Babak Khademi (Stanford Univ., Stanford, CA 94305)

This research presents the results of an experimental study of the vibration properties of the Kurdish long necked lute called Tanbour (aka Tambour, Tanbur, or Tamurah). A system transfer function of this musical instrument has been derived. [Hedayat Alghassi *et al.*, "Time series analysis of the Kurdish long necked lute Tanbour,", J. Acoust. Soc. Am. **122** (2007)]. To investigate the complex vibrational dynamics of Tanbour, the deflection patterns of Tanbour's pear shaped soundboard was acquired using a low-end time-averaged electronic speckle pattern interferometery (ESPI) imaging system. Later, the acquired image patterns were categorized as three resonance modes: near certain, dipolar, and quadrupolar. The three noticeable modes are assigned to the specific plectrum angles and positions. Moreover the short term Fourier transform (STFT) of the sound produced is derived for those plectrums. The strings were excited by a controlled mechanical plucker to maintain a high precision. Finally the sound spectra were related to the vibrational modes of the instrument, so one can approximately speculate the vibrational mode of Tanbour's soundboard from the STFT spectra for a basic plectrum.The criteria of observing three obvious vibration modes from the ESPI acquired images can also be used to achieve the optimum location of the bridge on the sound board.

Session 3aNS

Noise and Architectural Acoustics: Fire Codes and Acoustics

Matthew V. Golden, Cochair
Kinetics Noise Control, 6300 Irelan Pl., Dublin, OH 43017-0655

Ralph T. Muehleisen, Cochair
Illinois Inst. of Tech., Civil and Architectural Eng., 3201 S. Dearborn, Rm. 228, Chicago, IL 60616

Chair's Introduction—8:00

Invited Papers

8:05

3aNS1. Fire resistance and noise control in multifamily buildings. J. D. Quirt and T. R. T. Nightingale (Natl. Res. Council, Inst. for Res. in Construction, Ottawa, ON, K1A 0R6, Canada)

The walls and floors that separate dwellings in apartment buildings and condominiums must serve many functions other than structural support. To select wall or floor assemblies, designers need ready access to collections of test ratings, so this talk begins with a brief overview of some key factors affecting both the fire resistance and sound transmission ratings of wall assemblies and floor/ceiling assemblies with gypsum board surfaces. For successful buildings, however, designers must also appreciate how the overall fire resistance and sound isolation may be compromised by details at junctions and penetrations, and use a system approach that respects the requirements for both sound and fire controls. This presentation will focus on fire stops and fire blocks in the context of Canadian codes and standards, illustrate design options for fire stops at some common junctions and penetrations (mainly in the context of lightweight framed constructions), and provide guidance on corresponding acoustical issues.

8:25

3aNS2. Impact of suspended acoustical ceilings on the fire performance of open spaces. Tom Fritz and Kenneth Roy (Armstrong World Industries, ABP Technol., 2500 Columbia Ave., Lancaster, PA 17604, twfritz@armstrong.com)

An architectural design trend referred to as the "exposed structure" look is becoming increasingly popular due to a number of emerging design factors. In some cases, there is a desire to expose the overhead mechanical systems and roof deck, thus providing a feeling of spaciousness and economy. Or, in other cases, the desire for openness is driven by the installation of underfloor air distribution, in which case a traditional overhead ceiling is not needed to cover plenum mechanicals. The architects and building owners must, however, be reminded that the occupant safety, satisfaction, and performance will often be negatively impacted by these architectural design decisions. In particular, the response time of fire safety devices may be adversely affected, as shown by simulation tests.

8:45

3aNS3. Impact of material flammability regulations on acoustical treatment selection for heavy-duty vehicles. Charles Moritz, Kendall Bush, and Jennifer Shaw (Blachford Acoust. Lab., 1445 Powis Rd., West Chicago, IL 60185, cmoritz@blachfordinc.com)

Typical acoustical materials used in cars and trucks include polyurethane foam, polyester fiber blends, vinyl barriers, and various plastic or fabric facings. These materials are required to pass Federal Motor Vehicle Safety Standard 302, which regulates the speed at which a horizontal sample of the material can burn. Buses and specialty vehicles can be subject to more stringent flammability or smoke development requirements. These additional requirements may preclude the use of some common acoustical materials or require the addition of fire retardant chemicals or special facings. This paper reviews the commonly specified flammability regulations used with industrial vehicle acoustical materials and discusses their impact on material selection, performance, and cost.

9:05

3aNS4. Understanding the acoustical implications of Underwriters Laboratory listings. Matthew Golden (6300 Irelan Pl., Dublin, OH 43017, mgolden@kineticsnoise.com)

With today's ever more complicated building design, acoustical consultants should be aware of how their recommendations affect other aspects of floor/ceiling design. The acoustical ratings must be balanced with the need for fire rated assemblies as well as structural requirements. As the industry leader in the listing of fire rated designs, the Underwriters Laboratory (UL) maintains a list of assemblies that will meet various levels of fire resistance. These listings are very detailed and even slight changes to the design can have a profound impact on the assembly's fire resistance rating. This paper will compare various listed wood-frame floor/ceiling assemblies with respect to acoustical design. These comparisons will include fiberglass thickness, number of drywall layers, resilient channel spacing, and types of resilient sound clips. The general affect of each of those variables on acoustical performance will be shown. Several insights will also be shared from the acoustical supplier's perspective on the UL listing experience.

9:25

3aNS5. Gypsum board fire resistant assembly codes, standards, and certification. Rick Curkeet (Intertek Testing Services, 8431 Murphy Dr., Middleton, WI 53562, rick.curkeet@intertek.com) and John Nicholas (Intertek Testing Services, Elmendorf, TX 78112)

Fire resistant (FR) assemblies are key components of building a life safety design. These assemblies are typically required to divide buildings into smaller compartments so that, should a fire start, it will remain confined for a sufficient period to allow evacuation and arrival of fire fighting personnel. Properly installed and maintained FR construction has a long history of performance in both saving lives and limiting fire related property damage. This presentation provides an overview of the process used to develop FR assembly ratings and listings where a gypsum wallboard is a key assembly component. The discussion includes the test methods and standards required by model building codes, the process used by designers and manufacturers to obtain design certifications, and how these designs are intended to be applied in the field. There are often occasions where a basic FR assembly may need to be altered to improve performance not directly related to fire resistance. This frequently includes addition or changes in materials or other aspects of a design to improve acoustical performance. However, all changes to FR assembly designs have the potential to change the fire resistance of the assembly. The process required to obtain acceptance of design modifications is discussed.

Contributed Paper

9:45

3aNS6. The fire code and speech intelligibility. Richard J. Peppin (Scantek, Inc., 7060 Oakland Mills Rd. L, Columbia, MD 21046)

The National Fire Code has provisions for speech intelligibility for annunciators. These requirements and their measurements will be discussed.

WEDNESDAY MORNING, 12 NOVEMBER 2008 LEGENDS 10, 10:15 A.M. TO 12:00 NOON

Session 3aPA

Physical Acoustics: Material Properties and Elasticity

James P. Chambers, Chair
National Ctr. for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Contributed Papers

10:15

3aPA1. Elastic constant measurements at elevated temperatures. Rasheed Adebisi and Josh Gladden (Dept. of Phys., Univ. of Mississippi, University, MS 38677)

Resonant ultrasound spectroscopy (RUS) is a technique in obtaining the full elastic tensor of single crystal materials by measuring the mechanical resonances of a polished sample. Elastic constants of materials are a sensitive probe into the atomic environment; therefore, they are useful tools for studying phase transitions at high temperatures. A high-temperature system in which test samples are in direct contact with two transducers was designed to conduct RUS measurements up to 650 °C. Extremely high-temperature (~650 °C) measurements are made possible by separating the sample, placed in a tube furnace, and the transducers with buffer rods made of low-acoustic attenuation materials with good thermal stability such as ceramic alumina or fused quartz. Results from the high-temperature measurements on a standard thermoelectric material, silicon-germanium, demonstrated that the system has the ability of acquiring resonance signals at temperatures up to 1000 °C. Experimental issues such as additional resonance peaks introduced by the buffer rods and sample loading will be addressed. The apparatus has been used to study phase transitions in transition metal oxides and novel Zintl phase thermoelectric materials. These results demonstrate the great potential of RUS methods in high-temperature physics.

10:30

3aPA2. Phononic band gaps in periodically corrugated lithium niobate plate. Igor Ostrovskii (Phys. and Astronomy, Univ. of Mississippi, University, MS 38677)

A *ZX*-cut lithium niobate crystalline plate with a periodic corrugation is shown to have the band gaps in its spectrum of the plate acoustic waves. The corrugated plate is modeled numerically using finite element method. The depth of corrugations is 20% of the plate thickness along the *Z*-axis, and the corrugation period along the *X*-axis is taken from 0.3 to 1 mm for different model samples. The phase and group velocities of propagating acoustic modes will be affected significantly by the periodic change of plate thickness. The cutoff frequencies of higher-order modes will increase, and some new plate acoustic waves will appear. The periodic corrugations will cause the reflections and band gaps in the spectra of propagating modes. The strongest effects occur when the period of corrugations is equal to an integer number of half wavelengths of a mode. However, the distortions of the spectra do not appear in the spectrum if only 10% of the plate length is corrugated. One can achieve a complete band gap with certain corrugation periods and plate thicknesses. A periodically poled corrugated $LiNbO_3$ is also considered. These acoustic properties of a phononic $LiNbO_3$ crystal may be employed in a new generation of acoustic filters.

10:45

3aPA3. Dispersion of plate acoustic waves in periodically poled lithium niobate. Andriy Nadtochiy and Igor Ostrovskii (Phys. and Astronomy, Univ. of Mississippi, University, MS 38677)

The phase and group velocity dispersion of the plate acoustic waves (PAWs) in *ZX*-cut periodically poled lithium niobate (PPLN) wafer is calculated theoretically and verified experimentally. The dispersion curves of eight lowest PAW modes are calculated with the help of the finite element method. The PAWs in PPLN have some peculiarities in comparison with the case of a single crystal waveguide. First, some PAW modes may reflect out of the multidomain structure, and so a band gap may occur in the PAW spectrum. For example, a zero antisymmetric mode $A0$ has 0.2-MHz-wide band gap at frequency where its wavelength is equal to the period of a domain pattern. Second effect includes a variation of a cutoff frequency of PAW. The experimental dispersion curves are in agreement with calculations. The phononic band gap of $A0$ mode and the cutoff frequency shift of first antisymmetric mode $A1$ are confirmed by our experiments. The

group velocity is calculated from the experimental phase-velocity dispersion curves. The calculated and measured group velocities are consistent. It is shown that the group velocity of $A0$ mode approaches zero within the band gap.

11:00

3aPA4. Large elastic softening in highly magnetostrictive Fe-based alloys. Jared LeBlanc, Gabriela Petculescu (Dept. of Phys., Univ. of Louisiana, Lafayette, LA 70504), Marilyn Wun-Fogle, James Restorff (Naval Surface Warfare Ctr., W. Bethesda, MD 20817), Thomas Lograsso (Ames Lab., Ames, IA 50011), Arthur Clark (Clark Assoc., Adelphi, MD 20783), and Kristl Hathaway (Univ. of Maryland, College Park, MD 20742)

In order to practically exploit the large magnetostriction effects observed in Fe-based alloys, the magnetoelastic interactions in these materials have to be understood. Until now, Fe–Ga alloys have exhibited the largest magnetostriction in non-rare-earth alloys, but they also present a complex phase diagram, which does not allow a simple interpretation of magnetoelastic coupling at large Ga concentrations (above 20%). Fe–Ge alloys, which show an increase in the tetragonal magnetostriction but not as pronounced as in Fe–Ga, bring, however, a cleaner and more reproducible crystalline structure for the entire range of practical compositions. Measurements of elastic constants using resonant ultrasound spectroscopy in a saturating magnetic field and as a function of temperature for different compositions of the Fe–Ge alloys will be presented and contrasted to previously published Fe–Ga results. The addition of Ge to the structure of alpha-Fe softens the $c'=(c11-c12)/2$ shear modulus significantly. For example, only with 7 at. % Ge addition, c' falls by 21% from the 48 GPa value of the pure Fe crystal.

11:15

3aPA5. Generalized acoustic cloaking theory. Andrew Norris (Dept. of Mech. and Aerosp. Eng., Rutgers Univ., Piscataway, NJ 08854, norris@rutgers.edu)

An acoustic cloak is a region enclosing an object such that acoustic waves from all directions pass through and around the cloak as though the object was not present in the fluid. The object has zero scattering strength and is therefore acoustically invisible. Several researchers have proposed using fluids with anisotropic inertia for the cloaking material, but this has been shown by the author to require a cloak with infinite mass. An alternative framework is to use pentamodal materials, which guarantees finite mass and, under certain circumstances, isotropic inertia. The acoustic cloaking theory recently proposed by the author for acoustics is but one element in a class of cloaking theories for pentamodal materials. The main result is that pentamodal materials with anisotropic inertia form an invariant set of materials under arbitrary finite deformation and under the condition that the deformed material is cloaked. The general theory is explained using the language of finite deformations, common in continuum mechanics. The group properties follow by considering finite deformation in combination with gauge transformations. The talk will attempt to explain the meaning of pentamodal materials and show that the mathematics of transformation is related to simple physical requirements of transparency.

11:30

3aPA6. Using sound radiation to study collisions in granular media. Joshua Riner, Chola Regmi, and Andi Petculescu (Dept. of Phys., Univ. of Louisiana, P.O. Box 44210, Lafayette, LA 70504)

An acoustic method to check the validity of Hertz's law and assess the actual collision force law in granular media is proposed. The technique uses a microphone to detect the sound produced by collisions between solid polypropylene balls. The impact is observed to produce an underpressure followed by an overpressure. It can be conjectured that the initial compression of the balls lowers the pressure, followed by an expansion which leads to a sharp pressure increase. The structure following the overpressure peak could be due to the bulk and surface waves produced in the spheres. There is a visible difference between the underpressure and overpressure peaks (of about 9%), which can indicate energy losses. The power spectrum has a dominant peak at $\approx$7 kHz, accompanied by several smaller peaks, some of which may be intermodulation products. [Work funded by the Louisiana Board of Regents.]

11:45

3aPA7. Green's function model for time-reversal focusing of elastic surface waves. Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Time-reversal focusing applied to an object buried in soil improves the ability to focus energy on the object, suggesting a promising method for mine and IED detection. As a rule, these devices are buried close to the ground surface at a depth smaller than an elastic surface wave penetration depth. A Green's function approach is used to evaluate the field scattered by the object. Both the Green's function and the source function representing the scattering site are expanded in eigenfunctions of the homogeneous wave equation. When the object is acoustically compact, the scattered wave amplitude on the surface is described in the frequency domain by a Hankel function of order zero. Below the surface, the scattered wave amplitude is described by a Bessel function of order zero multiplied by exponentials that decay with depth, plus its Hilbert transform. The Green's function in the time domain is zero for negative time, and for positive time it is the Fourier transform of the Bessel function of order zero multiplied by the exponential depth dependence. The Green's function is used twice to calculate scattering from the object, first by the probe wave and next by the time-reversed wave. [Work supported by ONR.]

Session 3aSAa

Structural Acoustics and Vibration: Building Structural Acoustics and Vibrations

Kai Ming Li, Chair
Purdue Univ., School of Mechanical Eng., Ray W. Herrick Lab., 140 Martin Jischke Dr., West Lafayette, IN 47907

Invited Papers

8:30

3aSAa1. Transmission of noise through a double-pane window. Zhaohui Yu and Kai Ming Li (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., 140 Martin Jischke Dr., West Lafayette, IN 47907)

In the present study, the sound transmission of low-frequency noise through a double-pane window is analyzed by a coupled finite element/boundary element method (FEM/BEM). Particularly, the response of solid structure due to the acoustic excitations is treated by FEM while the behavior of fluid medium, which includes the fluid loading above and below the windows and the fluid cavity inside the window (between the double panes), is simulated by BEM. The stiffness matrix in the FEM formulation is adjusted by the technique of the component mode synthesis in order to handle the mounting conditions of windows effectively. Using the proposed model, the transmission of sound pressures through double-pane windows is calculated and validated with published experimental data. A set of comprehensive parametric studies is then conducted to explore for characterizing the acoustical properties of the double-pane windows, which may be used in the design stage.

8:50

3aSAa2. Modeling of transmission of shaped sonic booms into residential structures. Natalia Sizov, Kenneth Plotkin, and Christopher Hobbs (Wyle, 241 18th St., Ste. 701, Arlington, VA 22202, natalia.sizov@wyle.com)

The indoor perception of shaped minimized sonic booms is a concern for the acceptability of overland supersonic flight. An investigation of the transmission of shaped booms into residential structures has been conducted. Two classical methods were used. The first is single degree of freedom analysis. The second is modal analysis, modeling a wall as a simply supported rectangular panel and the interior as a rectangular cavity. Time domain responses were computed for both methods. Both methods allowed arbitrary waveforms, and the modal model allows arbitrary wave incidence angle. Sound pressures inside a room and acceleration of the wall were calculated and compared to measured data collected during the NASA sonic boom tests conducted at the Edwards Air Force Base in 2007. The effects of wave signature parameters on transmission were studied to evaluate the advantages of various kinds of low boom shapes, and the influence of thickened shock structures. Future extension of this research will be to employ finite element modeling. [This work was funded by NASA Langley Research Center, Contract No. NNL07AD96T.]

9:10

3aSAa3. Vibration and sound radiation from a rectangular panel with any boundary condition. Xuefeng Zhang and Wen L. Li (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This presentation is concerned with the vibroacoustic analysis of rectangular panels such as building walls, which may be isotropic or anisotropic in nature. A general analytical method is first described for modeling the vibration of a panel arbitrarily supported along each of its edges. The displacement field is analytically expressed as a two-dimensional Fourier cosine series supplemented with several one-dimensional series. Mathematically, such a series expansion is capable of representing and uniformly converging to any function (including the classical solution, if exists), which is adequately smooth over the solution domain. This series solution is exact in the sense that both the governing differential equation and the boundary conditions can be satisfied, on a pointwise basis, to any specified accuracy. Once the vibration field is determined, the resulting acoustic pressure and power radiation can be calculated in a simple and systematic manner. Numerical examples are given to demonstrate the vibroacoustic effects of the panel supporting conditions.

9:30

3aSAa4. Create an audio system to sonic boom an entire house? Victor Sparrow and Steven Garrett (Grad. Prog. Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, vws1@psu.edu)

To establish thresholds of public acceptability, the Federal Aviation Administration (FAA) would like to determine if it is possible to design and build a sonic boom simulation device that can accurately reproduce a sonic boom over an entire building, or portion of a building, such as a residential home. Having such a sonic boom reproduction device would make it possible to perform subjective tests on people in their own homes when exposed to simulated sonic boom noise corresponding to aircraft that have not yet been built. This presentation describes an ongoing study directed toward the design of a simulation device that can accurately reproduce sonic booms over the entire exterior surface of a building. The authors will present some plausible strategies that are in the development stages.

However, an additional motivation for this presentation is to make ASA members aware of this research and to solicit feedback from attendees prior to a January 2009 down-selection activity for the design of a system. The strict requirements for sonic boom playback fidelity make this project a Grand Challenge in Audio Reproduction. [Work supported by the FAA through the PARTNER Center of Excellence, www.partner.aero .]

Contributed Paper

9:50

3aSAa5. Broadband duct noise attenuation by a middle-ear mechanism. Lixi Huang (Dept. of Mech. Eng., Univ. of Hong Kong, Pokfulam Rd., Hong Kong)

Low-frequency duct noise is difficult to control passively by any device limited by its total volume due to system stiffness. When a side-branch device is used, cavity air becomes incompressible toward the dc frequency. One way to tackle the relative incompressibility for low to medium frequencies is to introduce acoustic phase difference in the cavity air oscillation, which otherwise behaves like the low-order room acoustics modes. The introduction of a light plate covering the cavity in our previous studies [J. Acoust. Soc. Am. **119**, 2628–2638 (2006)] does essentially that by enhancing the non-volume-displacing modes. Optimization shows that the required stiffness-to-mass ratio is much beyond what is available for existing bulk materials. In the present study, an alternative device that resembles the ossicular structure of the mammalian middle ear is introduced. The acoustic signal is passed along through a rigid pin which simulates the ossicular bones via its longitudinal motion. Contrasting with the usual flexural waves in plate or membrane structures, the bone transmission has almost infinite speed and it creates the phase difference needed to overcome the system stiffness. Theoretical predictions show that the performance of such device is comparable and better than the idealistic plate silencer with interesting comparisons of features.

WEDNESDAY MORNING, 12 NOVEMBER 2008 LEGENDS 1, 10:15 A.M. TO 12:00 NOON

Session 3aSAb

Structural Acoustics and Vibration: Scattering and Propagation

Courtney B. Burroughs, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Philip L. Marston, Cochair
Washington State Univ., Physics and Astronomy Dept., Pullman, WA 99164-2814

Contributed Papers

10:15

3aSAb1. Interfacial and elastic contributions to the backscattering of a right circular cylinder: Experiments and interpretation. Jon La Follett and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, lafollej@mail.wsu.edu)

Synthetic aperture sonar (SAS) images of elastic targets near an interface contain features resulting from the interaction of incident and scattered sound with the interface in addition to those associated with the free-field dynamical response of the target. To improve the understanding of these features, a right circular aluminum cylinder was suspended in a water tank through an air interface and insonified at a grazing incidence. Monostatic SAS images were obtained by scanning the source (and receiver) along a horizontal line. Backscattering measurements were also made as the distance from the target to the interface was varied. Some of the features could be explained using a previously developed ray-based theory [K. Gipson and P. L. Marston, J. Acoust. Soc. Am. **106**, 1673–1680 (1999);, **107**, 112–117 (2000)]. Among these features were responses due to guided generalized Rayleigh waves such as meridional rays and helical rays. Also, ray path calculations for edge diffraction contributions (as a function of target-interface distance) were used to identify certain echoes in the SAS images. [Research supported by ONR.]

10:30

3aSAb2. Numerical study of finite-amplitude source reconstruction in one-dimension. Micah R. Shepherd, Kent L. Gee (Dept. of Phys. and Astr., Brigham Young Univ., N-283 Provo, UT 84602), and Mark S. Wochner (Univ. of Texas at Austin, Austin, TX 78713)

Since its inception in the 1980s, near-field acoustic holography (NAH) has been widely used to determine structural source properties. More recently, it has also been used to determine aeroacoustic source behavior. Although several alternate forms of NAH have been developed, all are based on the (linear) Helmholtz equation and would presumably not function correctly for a finite-amplitude pressure wave propagating nonlinearly. However, when nonlinear effects or reconstruction distances are small, NAH may still be a useful tool for determining radiation characteristics of finite-amplitude sources such as jets and rockets. A one-dimensional numerical propagation scheme has been used to propagate broadband noise with various spectral shapes and amplitudes. A simple NAH-based magnitude reconstruction is used to determine frequency-, amplitude-, and propagation distance-dependent errors indicating in what regions linear NAH methods may produce large errors. The results indicate that the presence of shock coalescence causes relatively large errors in linear sound field reconstructions of broadband noise sources.

10:45

3aSAb3. Comparison of surface velocity measurement using a scanning laser vibrometer and acoustic holography. Pawan Pingle, Scot Bruderer, Christopher Niezrecki, and Peter Avitabile (Dept. of Mech. Eng., Univ. of Massachusetts Lowell, One University Ave., Lowell, MA 01854, christopher_Niezrecki@uml.edu)

A significant amount of interest exists in performing noncontacting full-field surface velocity measurement. For many years traditional surface velocity measurements have been made by using a scanning Doppler laser vibrometer (SDLV). Nearfield acoustical holography is another approach that enables reconstruction of quantities such as the acoustic pressure, surface velocity, intensity, and power radiated from a structure into three-dimensional space, based on the sound pressure measured at a two-dimensional surface. Within this work the surface velocity of a clothes dryer

panel is computed based on the acoustic field pressure measurements by using the Helmholtz equation least square (HELS) method, in which the reconstructed sound field is optimized by using spherical wave functions. The dryer panel is measured using an SDLV as well as a 64-channel microphone array. The reconstructed full-field surface velocity using the HELS method is compared to the measurement from the laser vibrometer during operation and also at discrete resonant frequencies.

11:00

3aSAb4. Scattering of a helicoidal Bessel beam by a sphere: The effect of an azimuthal phase gradient. Philip Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

In prior work [P. L. Marston, J. Acoust. Soc. Am. **121**, 753–758 (2007)] the partial wave series for the scattering by a sphere centered on an ordinary Bessel beam was derived. The present work extends the analysis of the far field scattering to the case of a helicoidal Bessel beam having an angular phase ramp equal to the azimuthal angle. Helicoidal beams possess an axial null and have an azimuthal phase gradient. Some of the resulting modifications of the scattering are illustrated for an empty steel shell in water. In agreement with symmetry arguments given previously [B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **106**, 3313–3316 (1999)] the backward scattering and forward scattering vanish. For some directions the scattering increases when shifting from ordinary to a helicoidal beam illumination. In addition to describing the scattering by acoustic beams, there are potential applications of this analysis to the description of acoustic radiation forces and torques. [Work supported in part by ONR and by NASA.]

11:15

3aSAb5. Pressure and displacement fields inside an absorbing fluid sphere ensonified by a plane harmonic wave. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543) and David T. I. Francis (Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, UK)

Analytic expressions have been developed for the pressure and displacement fields and dilatation in a lossy fluid sphere due to ensonification by a plane harmonic wave in a lossy immersion medium. These quantities have been computed for a 50-mm-diameter fluid sphere in water with a density of 1000 kg/m^3 and a sound speed of 1500 m/s for each of two frequencies, 10 and 100 kHz, hence with wavenumber-radius products of order 1 and 10, respectively. The density and sound speed values in the fluid sphere have been varied to achieve contrast factors in the range of 0.5–2 relative to the respective medium property. Absorption in the fluid sphere has been varied over the range of 0–10 dB/m. Results are presented for each field quantity along the axis of the sphere, as defined by the direction of propagation of the incident wave, transverse to this axis from the center, and along the surface from the forward to reverse directions. [Work partly supported by NOPP through ONR Award No. N000140710992.]

11:30

3aSAb6. Measurement of diffracted pressure fields using an automated measurement system. Simo-Pekka Simonaho and Timo Lähivaara (Dept. of Phys., Univ. of Kuopio, P.O. Box 1627, FIN-70211 Kuopio, Finland, simo-pekka.simonaho@uku.fi)

The theory of sound diffraction has been studied widely. Different modeling methods for sound diffraction have been proposed. Some of these modeling techniques are also used in the room acoustic modeling. In this study, diffracted pressure fields from a thin panel were measured. In the measurements, a loud-speaker was used as a sound source and diffraction from a thin panel was measured. The measurements were performed in small semianechoic chamber using a three-dimensional 3-D automated position system. This system consists of a 3-D position system, a data acquisition hardware, and a control unit. With this system, it is possible to measure diffracted pressure fields with high-spatial accuracy. In addition, numerical simulations are used in the modeling of sound diffraction from a thin edge. Finally, the simulated pressure fields are compared to the measured fields.

11:45

3aSAb7. The invariants of the time-reversal operator for asymmetric response matrices. Peter Simko and Jafar Saniie (Dept. of Elec. and Comput. Eng., Illinois Inst. of Tech., 3301 S. Dearborn, Chicago, IL, psimko@ece.iit.edu)

The decomposition of the time-reversal operator is a signal analysis method applicable to ultrasonic methods of target detection and characterization. In this paper, the eigenmodes of the time-reversal operator are studied for a single rigid cylinder with elliptical cross-section leading to an asymmetric response matrix formed by the interelement response of an active transducer array. The elliptical cylinder is of considerable interest in scattering problems since it encompasses within the limit both the circular cross-section case (very low eccentricity) and the strip (eccentricity approaching infinity). Variable cylinder properties will include parameters such as cross-sectional area, eccentricity, and angle of inclination of the semimajor axis with respect to the axis of a probing transducer array. Exact solutions for determination of the far-field response matrix using the familiar modal expansion with scattering coefficients computed via Mathieu functions will be discussed. The basis functions spanning the range of the time-reversal operator under conditions for which reciprocity no longer holds will also be discussed. Theoretical analysis will indicate that for scatterers near the Rayleigh limit, it is possible to characterize the aspect ratio of the target using a subarray method.

Session 3aSC

Speech Communication: A Quantal Transition: Ken Stevens in "Retirement" I

Helen M. Hanson, Chair
Union College, Electrical and Computer Eng. Dept., 807 Union St., Schenectady, NY 12308

Chair's Introduction—8:00

Invited Papers

8:05

3aSC1. Physical principles behind quantal relations. Helen M. Hanson (ECE Dept., Union College, 807 Union St., Schenectady, NY 12308) and Kenneth N. Stevens (MIT RLE, Cambridge, MA 02139)

A distinctive contrast can be defined by a quantal relation between an articulatory parameter and an acoustic parameter. It is postulated that there are two sources of quantal relations. *Aeromechanical interactions* arise because the nature of the interaction of airflow with the compliant mechanical structures of the vocal tract can change abruptly as an articulatory parameter changes continuously. As a result, the nature of the generated acoustic source changes abruptly. *Acoustic resonator coupling* results from the fact that the vocal tract, together with adjacent structures, can create several cavities that can be coupled and uncoupled, resulting in a transfer function that can show an abrupt discontinuity as a consequence of the rapid movement of a zero. These two principles lead to a natural division of the distinctive features into two groups, articulator free and articulator bound. Because the features in the two groups are defined by different physical principles, relations among the features are quite different: the articulator-free features, being based on aerodynamic conditions in the vocal tract, are constrained hierarchically, while articulator-bound features have fewer constraints. The natural constraints among the features are such that the featural representation of a segment is rather sparse.

8:30

3aSC2. Small vocal-tract cavities revealed by magnetic resonance imaging and their acoustic influences. Kiyoshi Honda (ATR Cognit. Information Sci. Labs, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan and Phonet. Phonology Labs, UMR7018 CNRS-Univ. Paris III, 19 rue des Bernardins, 75005 Paris, France, honda@atr.jp), Hironori Takemoto (ATR Cognit. Information Sci. Labs, Kyoto 619-0288, Japan), and Tatsuya Kitamura (Konan Univ., Kobe 658-8501, Japan)

In natural speaking conditions, the vocal tract exhibits small cavities in the hypopharyx and oral cavity, while their acoustic effects have not been fully recognized. Acoustic studies based on static and dynamic magnetic resonance imaging of the vocal tract in three dimensional have shown that these cavities cause regional resonance and antiresonance in the vocal tract. The hypopharyngeal cavities, including the supraglottal laryngeal cavity and bilateral cavities of the piriform fossa, characterize vowel spectra in the higher frequencies and contribute to determining voice quality and speaker characteristics. The laryngeal cavity functions as a Helmholtz resonator to generate an extra formant in the vicinity of 3 kHz, and the piriform fossa forms a pair of side-branches of the vocal tract to cause spectral zeros in the vicinity of 4–5 kHz. The bilateral interdental spaces form a pair of side-branches in the oral cavity in nonlow vowels, and their geometry varies with articulation. Their antiresonance frequency is below 2 kHz in /i/, while it rises toward /a/ crossing the second formant in transition, and thus causing a small discontinuity of that formant between the two vowels.

8:55

3aSC3. Subglottal coupling as a quantal basis for the feature [back]. Morgan Sonderegger (Dept. of Comput. Sci. Univ. of Chicago, 1100 E 58th St., Chicago, IL 60637, morgan@cs.uchicago.edu) and Xuemin Chi (RLE, MIT, Cambridge, MA 02139)

A model of acoustic coupling between the oral and subglottal cavities is described, which predicts attenuation of vowel formant prominences and discontinuities in formant trajectories near resonances of the subglottal system. The hypothesis that these effects on F2 near the second subglottal resonance (SubF2) are quantal effects for the feature [back] is examined using acoustic and subglottal data from English-speaking adults. Experimental studies of F2 and SubF2 in English vowel production [Chi and Sonderegger, JASA **122**, 1735–1745 (2007)] are reviewed and show that attenuation of second formant prominence and discontinuities in F2 trajectories near SubF2 consistently occur in back-front diphthongs, in accordance with the acoustic model, while for monophthongs front and back vowel F2 values pattern above and below SubF2, as expected under the quantal hypothesis. An additional analysis of the data is presented, showing that breathiness, attenuation, and discontinuity are positively correlated across back-front diphthongs, as predicted by the acoustic model. Lastly, a cross-linguistic survey of F2 data from 45 languages from the literature is presented, showing the cross-linguistic plausibility of the hypothesis that [+back] and [−back] vowel F2 values are separated by SubF2.

9:20

3aSC4. Development of subglottal quantal effects in young children. Youngsook Jung (Speech Commun. Group, MIT, Cambridge, MA 02139, and Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139), Steven M. Lulich, and Kenneth N. Stevens (Massachusetts Inst. of Technol., Cambridge, MA 02139)

Quantal articulatory acoustic relations between formants and subglottal resonances have been shown to define several vowel and consonant distinctive features, including the vowel features (back) and (low). In particular, the *F*2 frequency of front vowels is higher than the second subglottal resonance (Sg2) but is lower than Sg2 for back vowels. Likewise for low vowels, *F*1 is higher than Sg1. Measurements of *F*1 and *F*2 for productions of the vowels /a/ and /ae/ were made on a number of occasions for several children in the age range 2.5–3.7 years. Measurements of Sg1 and Sg2 for each child were obtained from the locations of discontinuities in the *F*1 and *F*2 trajectories in diphthongs and in vowel-consonant transitions. At the earlier ages for these children, there was variability in the utterances, and the *F*1 and *F*2 values deviated from the expected relation. At the later age of 3 years and above, considerable agreement with the expected values of *F*1 and *F*2 in relation to the subglottal resonances was obtained. The transition from the expected quantal relation appears to occur in the age range between 2 and 3 years. [Work supported in part by NIH Grant R01-DC00075.]

9:45—10:00 Break

10:00

3aSC5. Quantal relations from the acoustic coupled resonators of sonorant consonants. Carol Espy-Wilson (ECE Dept., Univ. of Maryland, A. V. Williams Bldg., College Park, MD 20742)

In this talk, we discuss the quantal relations that result from the coupled resonators of sonorant consonants. We have developed three dimensional geometries for light and dark /l/'s and for retroflex and bunched /r/'s based on MRI data from several speakers. In addition, we have used a computer vocal tract program (VTAR) to develop simpler models from MRI-derived area functions for /r/'s, nasal consonants, and nasalized vowels. Our vocal tract modeling shows that zeros occur in all of these cases due to side branches. For /l/, the side branches are the result of the palatal constriction with airflow around the sides of the tongue. For /r/, the side branches are due to the large front cavity, sometimes with space under the tongue. For the nasals and nasalized vowels, zeros occur due to the coupling between the oral and nasal cavities and the paranasal sinuses. We will discuss how the frequencies of the zeros vary as a result of changes in articulatory configurations. In addition, for /r/, we will show the quantal relationship in the frequency of *F*3 and the relative lengths of the cavities anterior and posterior to the palatal constriction.

10:25

3aSC6. Language-independent phonetic definitions of distinctive features. G. N. Clements and Rachid Ridouane (Laboratoire de Phontique et Phonologie, Sorbonne-Nouvelle, 19 rue des Bernardins, 75005 Paris, France, clements@idf.ext.jussieu.fr)

In spite of the fundamental role that distinctive features play in phonology and phonetics, current research continues to raise basic questions concerning how features can be defined in terms of measurable language-independent physical properties in the articulatory and acoustic/auditory domains. It has been proposed that for a feature to be recovered from a speech event, not only must its articulatory condition be met but the acoustic definition must be satisfied, in the absence of which further enhancing attributes must be present [K. N. Stevens, "Acoustic and perceptual evidence for universal phonological features," Procedings of the 15th International Conference of Phonetic Sciences, Barcelona, 2003 (unpublished), pp. 33–38]. This paper addresses this proposal through the study of the phonetic feature (spread glottis), drawing upon acoustic, fiberscopic, and photoelectroglottographic data from Berber and Kabiyé. A definition for this feature is proposed which associates a specific articulatory state with an equally specific acoustic definition, holding for the full set of sounds and contexts in which this feature can be implemented.

10:50

3aSC7. Acoustic characteristics of glides /j/ and /w/: Interactions with phonation frequency. Elisabeth Hunt (Speech Commun. Group, RLE, MIT, 77 Massachusetts Ave., Rm. 36-549, Cambridge, MA 02139, ehon@speech.mit.edu)

As part of a larger study of acoustic characteristics of glides, patterns in the movements of the fundamental frequency of phonation (*F*0) were measured during utterances of the glide segments /j/ and /w/. Two male and two female American English speakers produced intervocalic glides in six different vowel contexts and five different prosodic contexts. *F*0 contours show pronounced valleys during glide segments, in which the frequency of phonation is decreased relative to the surrounding prosodic contour. These valleys occur more frequently for higher fundamental frequencies, as in female speech. Effects of surrounding vowel and prosodic contexts were also observed, with *F*0 valleys more often occurring during glides when the following vowel is pitch accented or when the height of the surrounding vowels is high. The results are interpreted in terms of acoustic loading on the glottal source; the effects of which are most pronounced when a narrow vocal tract constriction (as in a glide segment) lowers the first formant frequency into the range of *F*0. This interaction with *F*0 may provide an enhancing acoustic cue to the presence of glides in the speech stream, particularly for female speakers or high-vowel contexts. [Work supported in part by NIH Grant No. DC00075.]

11:15

3aSC8. F0 control in electrolarynx speech. Yoko Saikachi (Speech and Hearing Bioscience and Technol., Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139, yokos@mit.edu), Kenneth Stevens (MIT, Cambridge, MA 02139), and Robert Hillman (Massachusetts General Hospital, Harvard Med. School, Boston, MA 02114)

Over half of laryngectomy patients use an electrolarynx (EL) to communicate, but current EL devices produce a mechanical sound quality mainly due to the lack of natural *F*0 variation. In order to improve the sound quality of EL speech, the present study aimed to develop and evaluate an automatic *F*0 control scheme, where *F*0 in EL speech is covaried based on modulations in the rms amplitude

of the EL speech signal. Declarative sentences ending with vowels produced by two male laryngectomy patients before and after the laryngectomy were used in order to develop procedures for controlling *F*0. The linear regression coefficients between *F*0 and rms amplitude in the normal speech were first determined and applied to the amplitude variation in EL speech to compute the *F*0 contour for the EL sentences. An analysis-by-synthesis approach was employed to modify the *F*0 contour of the EL speech as computed. Perceptual evaluation showed that the addition of amplitude-based *F*0 modulation resulted in EL speech that was judged to be more natural sounding than EL speech having constant *F*0, supporting the idea of using a simple linear relationship between amplitude and frequency to compute an *F*0 contour. [Work supported by NIH Grant No. R01DC006449.]

WEDNESDAY MORNING, 12 NOVEMBER 2008 LEGENDS 12, 8:55 TO 11:25 A.M.

Session 3aSP

Signal Processing in Acoustics, Underwater Acoustics, and Engineering Acoustics: Autonomous System Acoustic Sensors and Processors

Juan Arvelo, Chair
Johns Hopkins Univ., Applied Physical Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099

Chair's Introduction—8:55

Invited Papers

9:00

3aSP1. Advanced classification techniques for real-time signals in resource-constrained systems. G. Scott Peacock, David Barsic, and Ashley Llorens (Johns Hopkins Appl. Phys. Lab, 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

Automated classification of situational awareness data collected by autonomous vehicles is currently an unmet need in many applications. Classification algorithms developed at JHU/APL extend large margin classification (LMC) machine learning techniques to solve domain-specific problems such as those found in unmanned undersea vehicles systems. Common classification issues for the described systems include the following: (1) Asymmetric binary class membership, that is, a small amount of a target signal must be distinguished in a very large collection of data; (2) no silver-bullet features, i.e., robust classification requires many weak features, are used to distinguish targets from other signals; and (3) limited processing resources. The JHU/APL solution uses existing LMC technology with modifications to solve specific domain issues: (1) addition of a penalty term to address class asymmetry, (2) an iterative training algorithm that yields a sparse solution optimized for a given computational footprint, and (3) featureless classification that requires minimal or no data reduction and can improve classification robustness of the algorithm and cut development costs. Use of these developed techniques is demonstrated using well studied open source data that are representative of that required for autonomous vehicle classification tasks.

9:20

3aSP2. Time-frequency variations of the bistatic scattering response of proud and buried elastic shells in shallow water: Implication for mine counter-measure sonar systems. Shaun D. Anderson, Karim G. Sabra (Woodruff School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW, Atlanta, GA 30332-0405), Manell E. Zakharia (French Naval Acad., BP600, 29240 Brest armees, France), William A. Kuperman (Univ. of California, San Diego, La Jolla, CA 92093-0238), and Henrik Schmidt (MIT, Cambridge, MA 02139)

For underwater sonar, time-frequency analysis has been shown to be a relevant tool for the detection and classification of man-made targets (shell). For instance, for traditional monostatic systems, a key energetic feature of spherical shell is the coincidence pattern, or midfrequency enhancement, that is created by the coherent addition of antisymmetric Lamb-waves propagating clockwise and counterclockwise around the shell. The development by the Navy of mine countermeasure sonar systems, using a network of autonomous systems in unmanned vehicles, provides a mean for bistatic measurements, and thus potentially bistatic enhancement for target detection. We have investigated the time-frequency variations of the bistatic scattering response of elastic shells. The influence of the medium parameters as well as the source-receiver configuration will be investigated in free space and then extended to the case of a shallow water waveguide. In particular, we studied the bistatic variations of the coincidence pattern for classification purposes. Finally, the design of a robust space-time-frequency bistatic sonar system will be discussed to enhance the target detection of shells with the use of multiple sensors.

9:40

3aSP3. Signal processing strategies for autonomous underwater acoustic systems. Paul Hursky, Martin Siderius, Michael B. Porter (3366 North Torrey Pines Court, Ste. 310, La Jolla, CA 92037), Vincent K. McDonald, John M. Stevenson, and Brian Granger (SPAWAR Systems Ctr., San Diego, CA 92152)

Designing autonomous systems for underwater acoustic applications presents a number of competing challenges—these systems must adapt to a wide variety of possible events without human intervention, and they must do so without using much power. To address these challenges, we have explored algorithms borrowed from software-defined radio, octave processing (aka proportional processing),

parametric approaches (that do not require continuous sampling), and algorithms amenable to parallelism. Important functions for realizing autonomy include reducing the dimensionality of the data in stages, associating observations within and across platforms, maintaining multiple hypotheses, and opportunistic specialized processing to resolve outstanding ambiguities. We will describe our resulting embedded processing system and how we adapt this system to two related applications: environmental mapping and marine mammal observation. We will address the following questions: Does mobility help? Do greater numbers of less capable platforms help redress loss of individual platform capability? How much communications bandwidth do we need? We will close by presenting processing results produced by this system during several experiments at sea.

10:00—10:15 Break

10:15

3aSP4. Aggressive adaptive beamforming for ambient noise inversion with a limited-aperture sonar on an autonomous platform. Juan Arvelo, Jr. (The Johns Hopkins Univ. Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

Bottom loss inference from ambient noise inversion has been shown to yield accurate and robust predictions of undersea sound propagation and sonar detection performance using vertical line arrays and a conventional beamformer [J. I. Arvelo, J. Acoust. Soc. Am. **123**, 679 (2008)]. However, any array with a vertical aperture should be good candidates for ambient noise inversion. The main problem with vertical line arrays is that they are not well suited to rapidly survey a large area. Platform translation with a vertical array is expected to cause extreme array shape distortions that could seriously limit the accuracy of the results. To mitigate array shape distortion, a smaller array with vertical aperture could be mounted on the survey platform, which would allow it to survey the area at greater speeds. The platforms' diameter certainly limits the vertical aperture of the array. A limited vertical aperture limits the frequency range, resolution, and accuracy of the estimated bottom loss. Since the array is rigidly fixed to the platform's body, an aggressive adaptive beamformer may be adopted to increase the resolution and accuracy of the inverted bottom loss to very low frequencies. The performance is demonstrated via comparisons against conventional beamformers with numerical simulations and measurements.

10:35

3aSP5. Geoacoustic inversion of ocean surface-wave noise with a littoral glider. Billy D. Jones (Alaska Native Technologies, LLC, 26273 Twelve Trees Ln., Ste. M, Poulsbo, WA 98370, billy@alaskanativetech.com)

Passive acoustics can be used to characterize the ocean bottom in a littoral environment. Ocean surface-wave noise acts as a natural directional source that can be measured to obtain bottom loss and subsequently inverted to obtain bottom density, sound speed, attenuation, and layering structure. A physical ocean surface noise model is introduced and its horizontal/vertical spatial cross-correlations is discussed. Acoustical energy conservation and its relation to bottom loss are discussed. The measurement method is a continuation of the ambient noise inversion work of [J. I. Arvelo, Jr., J. Acoust. Soc. Am. **123**, 679–686 (2008)]. Aggressive adaptive beamforming is used for high-resolution bottom loss measurements with surface-wave noise as a source. Ambient noise measurements with an eight-element wire polyvinylidene fluoride vertical line array mounted in the bow of a littoral glider are presented. Results from summer 2008 measurements in Port Madison, WA are presented and compared against known bottom loss curves for the area. [Work supported by SPAWAR.]

Contributed Papers

10:55

3aSP6. Using the holographic property of synthetic aperture sonar images to enable real aperture sonar navigation. Richard Rikoski and Jose Fernandez (Naval Surface Warfare Ctr., Panama City, Panama City, FL 32407)

Synthetic aperture sonar (SAS) images contain broadband multiaspect information about a scene; this leads to several holographic properties. The most commonly understood property is that a complete image can be reconstructed at lower resolution from any subband of the Fourier transform; this effect is used for speckle reduction. A less commonly understood property is that a SAS image contains all possible real aperture images over some range of angles and frequencies. This allows real aperture images to be coherently correlated with SAS images. The ability to coherently correlate sonar images permits the use of a broad class of probabilistic feature-based navigation algorithms. This paper describes the method of coherently correlating SAS and real aperture images for navigation, and presents results from St. Andrew Bay and the Gulf of Mexico using the small synthetic aperture minehunter.

11:10

3aSP7. Geometrical distortion correction in attitude estimating Hough transformation. Hisashi Shiba (Radio Application Div., NEC Corp., 1-10, Nisshin-cho, Fuchu, Tokyo 183-8501, Japan, h-shiba@aj.jp.nec.com)

In the previous meeting, a new underwater vehicle attitude estimation procedure was presented. It is based on a three dimensional expansion of the Hough transformation detecting the sea surface. It is robust for noise and does not require *a priori* knowledge whether the sea surface and/or the seafloor are in the obtained sonar image. After a rough investigation, it was found that two types of geometrical distortions sometimes affect the Hough transformation results. The distortions are intrinsic in the Hough transformation. One of them is caused by nonlinear relationship between the real space and the parameter space. Another originates from digital geometrical measurements. The former is able to be suppressed by the inverse Hough transformation. The latter is resolved by applying a Euclidean distance area measurement. The two methods are explained in detail in this presentation. These problems tend to be distinguished especially in severe noise environments and it is possible to emerge in usual line detections on two dimensional images. After explaining the two methods, the Hough transformation effectiveness for a real data is also shown by tank experiment results.

Session 3aUW

Underwater Acoustics and Signal Processing in Acoustics: Robust Array Processing

Claire Debever, Chair
Scripps Inst. of Oceanography, Marine Physical Lab., 8820 Shellback Way, La Jolla, CA 92093-0238

Chair's Introduction—8:30

Invited Papers

8:35

3aUW1. Spatial coherence? Henry Cox (Lockheed Martin, 4350 N. Fairfax Dr., Ste. 470, Arlington, VA 22203, harry.cox@lmco.com)

The term spatial coherence is frequently used to summarize the impact of the propagation environment on array performance and to suggest a limitation on useful array aperture. A source of considerable confusion is that there is no single widely accepted precise meaning of the term. In general, coherence implies a fixed phase relationship and involves time as well as space. The time duration over which the measurement is made or for which the phase relationship appears fixed is central to the measurements of coherence. A number of different measurements types and associated data analyses are described and the results are interpreted in terms of the pertinent aspects of acoustic propagation, such as multipath, surface and internal waves, etc. The effects of source and receiver motion are also considered. The implications for array performance and the robustness of beamforming approaches in the face of environmental variability and uncertainty are also discussed. Examples are given from recent shallow water experiments. [Work supported in part by NAVSEA Contract No. N00024-07-C-5210.]

8:55

3aUW2. Image-based refocusing of ultrasound arrays in the presence of strongly scattering objects. John Ballard and Emad Ebbini (Dept. Elec. and Comput. Eng., Univ. of Minnesota, 200 Union St. SE, Minneapolis, MN 55455, ball0250@umn.edu)

Ultrasound phased arrays are currently being investigated for a dual-mode (imaging/therapy) operation in targeting deep-seated abdominal tumors for selective destruction noninvasively. In some cases, e.g., liver tumors, the target is partially obstructed by the ribcage, which limits the array gain at the target. In addition, the intensity at the ribs may be high enough to cause treatment-limiting pain and/or collateral damage to the tissues surrounding the ribs in the path of the therapeutic beam. Using the imaging capabilities of these dual-mode arrays, we have formulated and experimentally verified a robust minimum variance beamforming algorithm for adaptive refocusing in the presence of the ribs. The algorithm utilizes image data of the treatment region formed using conventional beamforming in estimating the array steering vector at the target and the weighting matrix that maximizes the array gain at the target. In this paper, we will describe the mathematical formulation and present experimental data to demonstrate its robustness. We will also discuss the implications of the method, the use of ultrasound phased arrays for imaging, and therapy of trans-thoracic targets such as the liver and the heart.

9:15

3aUW3. Bayesian localization and tracking with environmental and array-element uncertainties. Stan Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada, sdosso@uvic.ca), Dag Tollefsen (Norwegian Defence Res. Establishment, 3191 Horten, Norway), and Michael Wilmut (Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper considers matched-field source localization and tracking when environmental parameters and/or array-element positions are not well known. A Bayesian formulation is applied in which source, array, and environmental parameters are considered unknown random variables constrained by noisy acoustic data and by prior information on parameter values (e.g., physical limits for environmental properties and element positions) and on interparameter relationships (e.g., limits on source speed and interelement spacing). The goal then is to extract source information from the posterior probability density (PPD). One approach is based on maximizing the PPD over all parameters to obtain optimal source locations. A key to solving this problem efficiently is that the VITERBI algorithm is applied to compute the highest-probability source track for each environment/array realization: this provides the optimal track, while requiring the optimization is applied only over the nuisance parameters. A second approach involves integrating the PPD over unknown environmental and array parameters to represent source-location information as a series of joint marginal probability surfaces over range and depth. Given the strong nonlinearity of this problem, marginal PPDs are computed numerically using efficient Markov-chain Monte Carlo importance sampling methods. The approaches are illustrated with examples based on simulated and measured acoustic data.

9:35

3aUW4. Robust adaptive beamforming: Evolution of approaches, analysis, and comparison. Sergiy A. Vorobyov (Dept. of ECE, Univ. of Alberta, 9107-116 St., Edmonton, AB T6G 2V4, Canada)

A so-called worst-case-based robust adaptive beamforming approach has been proposed a few years ago. This approach has been followed up by a number of works, which significantly improved the implementation issues. Recently, two new approaches have been developed: the probability constrained based and sequential quadratic programing based approaches. The probability constraint based

design requires the distortionless response constraint to be kept with a certain selected probability, while the worst-case-based design requires this constraint to be kept for the worst-case operational conditions. The sequential quadratic programing based approach uses the iterative estimation of the signal steering vector and, therefore, is significantly different from two two aforementioned designs. In this paper, we overview, analyze, and compare all the aforementioned approaches to robust adaptive beamforming design.

9:55

3aUW5. Robust processing techniques for underwater acoustics. Lisa Zurk (ECE Dept., Portland State Univ., P.O. Box 751, Portland, OR 97207, zurkl@pdx.edu)

The performance of sonar signal processing methods in shallow water waveguides is highly sensitive to the effects of the acoustic wave propagation, and the nature of this propagation is often critically dependent on the exact nature of the underwater channel. Signal processing formulations that exploit the structure of the channel have been shown to increase detection and localization performance when the channel properties are known. However, for many tactical situations, channel parameters are poorly unknown or "uncertain," thus motivating the need for robust processing techniques that do not require exquisite knowledge of the environment but may still exhibit the increased performance of full field methods. The goal of devising robust processing approaches remains of high interest and is an ongoing topic for signal processing research. This presentation first provides some metrics of environmental uncertainty and the effect on the sonar signal processing then provides a review of several proposed approaches for achieving robustness. The algorithms are grouped into the following three categories: robust matched field processing algorithms, guide source calibration techniques, and invariance processing (both passive and active). The presentation concludes with a discussion of ongoing research and future areas for investigation.

10:15—10:30 Break

Contributed Papers

10:30

3aUW6. Broadband high frequency matched-field processing. Claire Debever and William A. Kuperman (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238, cdebever @ucsd.edu)

Adaptive matched-field processing (MFP) is extremely sensitive to environmental uncertainties. While conventional and adaptive techniques may work for low-frequency signals in well-studied environments, the localization performance usually degrades rapidly as frequency increases, such that MFP in the 3.5 kHz regime, in shallow water environments, is typically problematic. A broadband coherent method [IEEE J. Ocean. Eng. **21**, 384–392], combined with white noise constraint and principal component techniques, is implemented to construct robust replicas from experimental data. Matched-field tomography is then used to better understand the origin of the frequency dependent MFP mismatch (uncertain bottom structure, sound speed fluctuations, or both), and the results are compared with simulations. Ultimately, we want to gain some insights into how to implement a robust matched-field processor for high-frequency scenarios without using experimental data to create replica vectors. In particular, we are seeking to understand the thresholds for adaptive processors as function of signal to noise ratio and frequency dependent environmental mismatch. [Work supported by ONR.]

10:45

3aUW7. Geoacoustic tracking. Caglar Yardim, Peter Gerstoft, and William Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr.-0238 9500 Gilman Dr., La Jolla, CA 92093-0238)

This paper shows how to incorporate tracking techniques such as the extended Kalman, unscented Kalman, and particle filters into geoacoustic inversion problems. This enables not only the inversion of environmental parameters but also the spatial and temporal tracking of them, making geoacoustic tracking a natural extension to geoacoustic inversion techniques. Water column and seabed properties are tracked in simulation and using the MAPEX2000 experimental data for both vertical and horizontal line arrays. Filter performances are compared in terms of filter efficiencies using the posterior Cramér–Rao lower bound. Tracking capabilities of the geoacoustic filters under slowly and quickly changing environments are studied in terms of divergence statistics. The suitability of each filter in geoacoustic tracking is discussed in terms of the track quality, complexity of the probability density function of environmental parameters, nonlinearity of the geoacoustic propagation, and computational cost. The results show that geoacoustic tracking can provide continuously environmental estimates and their uncertainties using only a fraction of the computational power of classical geoacoustic inversion schemes. [Work supported by ONR-N00014-05-1-0264.]

11:00

3aUW8. Inversion of ocean environmental variations via time reversal acoustics. Wen Xu and Jianlong Li (Zheda Rd. 38, Dept. of Information Sci. and Electron. Eng., Zhejiang Univ., Hangzhou 310027, China)

Time reversal (TR) processing is derived from the invariance of the wave equation for lossless medium to change in the sign of the time variable. By retransmitting the TR version of the time-dispersed received signal propagated from a probe source (PS) to a source-receiver array, one can reacquire the transmitted pulse at the PS location (time compression and spatial focusing) when the waveguide environment is time invariant. However, if some environmental variations occur between the two transmissions, the retrofocusing signal will be defocused. This paper presents a new method of environmental parameter inversion by comparing the difference of the focused signals measured at the PS location with/without environmental perturbations. Because the sound speed profile (SSP) plays a critical role in an uncertainty ocean environment, inversion of the SSP represented in terms of the empirical orthogonal functions is developed and discussed in detail. Simulations demonstrate some advantages of environmental inversion via TR: (1) great signal-to-noise ratios can be obtained at the PS location; (2) variations of the environment at different times can be directly inverted by repeatedly retransmitting either the same received signal generated by one PS transmission or the updated received signals generated by updated PS transmissions.

11:15

3aUW9. Arrival time estimation from sound signals in the ocean: A particle filtering approach. Rashi Jain and Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102, rj45@njit.edu)

The focus of this work is on accurate arrival time estimation from measured time-series at an array of vertically separated hydrophones in the ocean. We develop a particle filtering approach that treats arrival times as "targets", dynamically modeling their "location" at arrays of spatially separated receivers. Using Monte Carlo simulations, we perform an evaluation of our method and compare it to conventional maximum likelihood estimation, whereas we also compare our errors to Cramer–Rao bounds. The comparison demonstrates an advantage in using the proposed approach, which can be employed for minimization of uncertainty in arrival time estimation. Improved arrival time estimates can then be used for accurate geoacoustic inversion and source localization. [Work supported by ONR.]

Meeting of Accredited Standards Committee (ASC) S3 Bioacoustics

C. A. Champlin, Chair S3
University of Texas, Department of Communication Sciences & Disorders, CMA 2-200, Austin, TX 78712

D. A. Preves, Vice Chair S3
Starkey Laboratories, Inc., 6600 Washington Ave., S., Eden Prairie, MN 55344

Accredited Standards Committee S3 on Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note - those meetings will be held in conjunction with the Standards Plenary meeting at 9:15 a.m. on Tuesday, 11 November 2008.

Scope of S3: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S3/SC 1, Animal Bioacoustics

D. K. Delaney, Chair S3
USA CERL, 2902 Farber Drive, Champaign, IL 61822

W. Au, Vice Chair S3
Hawaii Institute of Marine Biology, P. O. Box 1106, Kailua, HI 96734

Accredited Standards Committe S3/SC 1 on Animal Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

Scope of S3/SC 1: Standards, specifications, methods of measurement and test, 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.

Session 3pAA

Architectural Acoustics: The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

David Lubman, Cochair
DL Acoustics, 14301 Middletown Ln., Westminster, CA 92683-4514

William J. Cavanaugh, Cochair
Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776-3027

Chair's Introduction—1:00

Invited Paper

1:05

3pAA1. Aural architecture: The missing link. Barry Blesser (Blesser Assoc., P.O. Box 155, Belmont, MA 02478, bblesser@alum.mit.edu)

While acoustic architecture focuses primarily on the acoustic physics of objects and geometries, aural architecture emphasizes the experience of space in terms of behavior and emotions. Because auditory spatial awareness, which is the basis for aural architecture, depends on a social value system, the role of acoustics varies among individuals and cultures. When evaluating the aural experience of space, two independent phenomena must be simultaneously considered: Space changes our experience of sound and sound changes our experience of space. Sound sources and spatial acoustics are inseparable. This bilateralism creates an interdisciplinary complexity that fuses physical and social sciences. Hearing is a means by which people acquire a sense of where they are, connecting them to dynamic events and spatial geometry. Auditory spatial awareness allows people to sense the elegance of a plush office, the emptiness of an uninhabited house, the depth of a dark cave, the quiet of a city covered in snow, the vastness of a railroad station, and the openness of a beach front. Each of these situations can be described in the language of aural architecture, which includes at least five types of experiential spatiality: navigational, social, aesthetic, symbolic, and musical.

2:05—2:15 Question and Answer
2:15—2:45 Book Signing

Session 3pED

Education in Acoustics: Acoustics Education Prize Lecture

Uwe J. Hansen, Chair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Chair's Introduction—2:15

Invited Paper

2:20

3pED1. From the sublime to the scientific: What musicians and acousticians can learn from each other. Murray Campbell (School of Phys. and Astronomy, Univ. of Edinburgh, Edinburgh EH9 3JZ, UK)

Many university music programs include an acoustics module, often taught by a physicist. At the University of Edinburgh, such a module has existed since the 1850s; taking over this course as a junior lecturer was my introduction to the fascinating world of musical acoustics. It rapidly became clear that a meaningful communication between scientists and musicians required humility and willingness to learn from both sides. This lecture explores aspects of that mutual learning process, focusing on some controversial areas in which the reconciling of scientific and musical viewpoints has not always proceeded in a spirit of humility.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

David R. Dowling, Chair
Univ. of Michigan, Dept. of Mech. Eng., 1231 Beal Ave., Ann Arbor, MI 48109-2133

Chair's Introduction—1:00

Invited Papers

1:05

3pID1. Hot topics in engineering acoustics. David A. Brown (BTech Acoust. LLC, and Electro-Acoust. Res. Lab., Adv. Tech. Manuf. Ctr. and ECE Dept, Univ. of Massachusetts Dartmouth, 151 Martine St., Fall River, MA 02723)

The maturation of single crystal piezoelectric materials in the past decade has spawned a renewed interest in traditional and new transducer designs for underwater acoustic and medical imaging applications. The engineering of acoustic devices based on piezoelectric crystals began in the first half of the 20th century and was overtaken by developments in ferroelectric ceramic compounds such as barium-titanate and lead-zirconium-titanate (PZT). These materials are now being challenged by engineered relaxor piezoelectric single crystals materials such as lead-magnesium (or zirconium or indium)-niobate-lead-titanate (PMN-PT, PZN-PT, and PIN-PT) in many applications. The new materials have tremendous improvements in piezoelectric properties including electromechanical coupling coefficients exceeding 90%, which can double the power factor bandwidth for underwater projectors, sound speeds that are a factor of 3 lower than PZT that enable compact low frequency sources, and strain levels as high as 1% for high drive and actuator applications. The hot topics related to engineering acoustics involve the development and commercialization of devices based on new and traditional transducer designs that exploit the novel properties of the new crystal materials. Examples of broadband underwater communications transducers, tonpiltz sonar projectors, pressure and pressure-gradient hydrophones, mechanical actuators, and medical devices are presented.

1:25

3pID2. Uncertainty in ocean acoustics. Steven Finette (Acoust. Div. Naval Res. Lab., Washington, DC 20375-5320, steven.finette@nrl.navy.mil)

When modeling ocean-acoustic systems, the environmental information necessary to compute the acoustic field properties is assumed to be accurately specified by various parameters, fields, and boundary conditions. However, in real world applications, these quantities are subject to uncertainties due to our incomplete knowledge of the waveguide environment. This form of uncertainty involves errors that are quite distinct from the numerical errors that can arise when a mathematical model is discretized, implemented on a computer, and solved with finite precision arithmetic. In effect, the environmental uncertainty introduces spurious degrees of freedom into the system. In order to make reliable simulation-based predictions, this uncertainty needs to be quantified and incorporated in the simulation process itself. The idea of embedding uncertainty into the simulation framework and elevating its status to a subject worth studying on its own merits represents a paradigm shift that has stimulated research in several disciplines. This talk will overview recent approaches to this problem in an ocean-acoustic context and give examples of computations that have incorporated environmental uncertainty in the numerical computation of acoustic field properties. [Work supported by the Office of Naval Research.]

1:45

3pID3. Hot topics in physical acoustics. Joseph Gladden, III (Dept. of Phys., Univ. of Mississippi, University, MS 38677, jgladden@phy.olemiss.edu)

The field of physical acoustics touches a broad range of technical areas important to fundamental science and society. This "Hot Topics" presentation will reflect the breadth of this impact by discussing the following three topics: sound waves in the early cosmos, acoustics in slip-stick friction systems, and acoustic metamaterials. The early universe, composed of hot ionized matter, was able to support acoustic waves until the temperature cooled enough to allow the formation of neutral atoms. The imprint of these relic acoustic waves is still evident in the cosmic microwave background and yields new information about key cosmological constants and dark matter. Earthquakes are perhaps the most destructive of natural disasters, as was painfully demonstrated by recent events in China. Studies on the effects of acoustic vibrations in slip-stick friction systems have begun to shed light on triggering mechanisms for earthquakes and may lead to better early warning systems. Metamaterials are man-made materials in which precise geometric arrays of structures are engineered to produce coherent scattering effects on scales much larger than the structures themselves. These systems can exhibit such exotic properties as a negative index of refraction, band gaps, and a negative effective elastic modulus.

Session 3pSC

Speech Communication: A Quantal Transition: Ken Stevens in "Retirement" II

Helen M. Hanseon, Chair
Union College, Electrical and Computer Eng. Dept., 807 Union St., Schenectady, NY 12308

Invited Papers

1:00

3pSC1. Enhancing the left edge: The phonetics of prestopped sonorants in Australian languages. Andrew Butcher (School of Medicine, Flinders Univ., GPO Box 2100, Adelaide SA 5001, Australia, andy.butcher@flinders.edu.au) and Debbie Loakes (Univ. of Melbourne, Parkville, VIC 3010, Australia)

The consonant systems of Australian Aboriginal languages are very similar to one another but very different from those of most other languages of the world. They have unusually few contrasts in manner of articulation and an unusually large number of places of articulation. Previous research has shown that speakers appear to employ a number of strategies to preserve place of articulation distinctions, particularly in intervocalic (coda) consonants. One such strategy is that in vowel + nasal sequences speakers avoid lowering the velum until the latest possible instant, presumably to preserve spectral clarity at the VC boundary. This often results in a brief homorganic oral stop occurring before the nasal. Phonetically prestopped nasals occur in a large number of languages across Australia and have become distinctive phonemes in a number of languages in the center and south. Less well documented is the parallel phenomenon of prestopped laterals, which is taken to be the outcome of a similar coarticulation avoidance strategy. This paper describes the wide distribution and distinctive phonetic characteristics of prestopped nasals and laterals in a number of Australian languages and proposes that both strategies are aimed at the enhancement of the left edge of the sonorant consonant.

1:25

3pSC2. Quantal events generated by the structural and temporal variation of the vocal tract. Brad Story (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721)

For connected speech, the time-varying vocal tract shape can be represented as a consonant superposition function that imposes constrictions and expansions on an underlying vowel substrate at specific points in space and time. The resulting flow of continuous speech sounds is a combination of characteristics of both the vowels and the consonants. A question is how speakers choose specific spatial locations and temporal patterns with which to execute particular consonants. In this study, acoustic sensitivity functions and formant nomograms based on a vocal tract area function model were used to determine the optimal locations for consonantal constrictions. Specifically, these techniques indicate the points within the vocal tract at which the acoustic effect of a constriction, in terms of formant frequency transitions, will rapidly change. These points are suggestive of quantal events that may occur during production of speech. [Work supported by NIH R01-DC04789.]

1:50

3pSC3. Consonant landmarks: Automatic detection and interpretation. Chiyoun Park and Nancy Chen (MIT, 77 Massachusetts Ave., Rm. 36-525, Cambridge, MA 02139)

Consonant landmarks are acoustic discontinuities in the speech signal that correspond to the closures and releases in speech production, and have been proposed as critical elements in speech processing [Stevens (2002)]. The three types of consonant landmarks represent the onset and offset of salient acoustic events: glottal vibration, turbulence noise, and sonorancy (e.g., nonvocalic voicing). While earlier work [Liu, (1996)] evaluated the success of identifying single candidate landmarks of all three types, this work focuses on two tools for evaluating strings of landmark candidates. First, a bigram model representing the physiologically feasible sequences of consonant landmarks is used to evaluate candidate strings. Second, a graphical method is used to identify the regions where the landmarks are reliably detected versus where they are ambiguous. Together these tools substantially improve the performance for landmark detection, identify regions in need of further acoustic analysis, and model the 'grammatical' structure of landmark sequences. Furthermore, the reliable regions in the proposed representation often correspond to structural elements such as lexical stress and word boundaries. Thus, the proposed representation is potentially useful in analyzing speech not only at the phoneme level but also at the word and phrase levels. [This work was supported by NIH/NIDCD DC02978 and T32DC00038.]

2:15

3pSC4. Point process models of distinctive feature landmarks for speech recognition. Aren Jansen and Partha Niyogi (Dept. of Comput. Sci., Univ. of Chicago, 1100 E 58th St., Chicago, IL 60637, aren@cs.uchicago.edu)

Several interrelated strands of research in linguistics, acoustic phonetics, and cognitive neuroscience suggest a host of new directions for the development of end-to-end computational models of speech perception and recognition. Natural candidates for exploration include (i) phonological representations in terms of distinctive features; (ii) nonlinear detectors for distinctive feature landmarks (or any other set of perceptually salient acoustic events), which define a sparse point process representation of the speech signal; (iii) syllable-metered temporal processing and/or syllable-sized integration windows; and (iv) point process models and hierarchical strategies for

recognizing words, syllables, phonemes, and features. A computational framework around these ideas has been developed and has led to phonetic recognition and keyword spotting performance that is competitive with equivalent hidden Markov model-based systems. This framework thus connects a computational platform for benchmarking competing scientific theories with simultaneous advancement toward a viable technological solution to the speech recognition problem.

2:40—2:50 Panel Discussion

WEDNESDAY AFTERNOON, 12 NOVEMBER 2008 GRAND BALLROOM A/B, 3:30 TO 5:30 P.M.

Plenary Session and Awards Ceremony

Mark F. Hamilton, Chair
President, Acoustical Society of America

Business Meeting of the Acoustical Society of America

Motion to approve the Plan of Merger of the Acoustical Society Foundation, Inc. into the Acoustical Society of America, Incorporated

Presentation of Certificates to New Fellows

DAVID A. BERRY
GEORGE A. BISSINGER
JOHN A. FAWCETT
DENNIS M. FREEMAN
BRUCE R. GERRATT
FRANK H. GUENTHER
KEITH R. KLUENDER
YIU W. LAM
MARSHALL LONG
CHRISTIAN LORENZI
BRYAN E. PFINGST
JOE W. POSEY
STUART ROSEN
ARMEN SARVAZYAN
MICHAEL A. STONE
ANN K. SYRDAL
JOE WOLFE

PRESENTATION OF SCIENCE WRITING AWARDS

SCIENCE WRITING IN ACOUSTICS FOR A JOURNALIST

HAZEL MUIR FOR ''NOISY NEIGHBOURS'' PUBLISHED IN *NEW SCIENTIST MAGAZINE*, AUGUST 2007

SCIENCE WRITING AWARD FOR MEDIA OTHER THAN ARTICLES

KATHLEEN VIGNESS RAPOSA, GAIL SCOWCROFT, CHRISTOPHER KNOWLTON, PETER WORCESTER
FOR ''DISCOVERY OF SOUND IN THE SEA'' WEBSITE

PRESENTATION OF ACOUSTICAL SOCIETY AWARDS

ROSSING PRIZE IN ACOUSTICS EDUCATION TO D. MURRAY CAMPBELL

SILVER MEDAL IN MUSICAL ACOUSTICS TO GABRIEL WEINREICH

SILVER MEDAL IN PHYSICAL ACOUSTICS TO PETER J. WESTERVELT

SILVER MEDAL IN SPEECH COMMUNICATION TO WINIFRED STRANGE

WALLACE CLEMENT SABINE MEDAL TO JOHN S. BRADLEY

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in Musical Acoustics

Gabriel Weinreich

2008

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Carleen M. Hutchins	1981	Thomas D. Rossing	1992
Arthur H. Benade	1984	Neville H. Fletcher	1998
John G. Backus	1986	Johan E. F. Sundberg	2003
Max V. Mathews	1989		

CITATION FOR GABRIEL WEINREICH

. . . for contributions to violin and piano acoustics.

MIAMI, FLORIDA • 12 NOVEMBER 2008

If anyone ever deserved the title "Renaissance man" it is Gabriel Weinreich. Physicist, theologian, linguist, philosopher, musician, humanitarian, inventor, and acoustics researcher: Gabi has distinguished himself in so many ways. He has published books on the theory of condensed matter (1965), thermodynamics (1968) and extended vector mathematics (1998), as well as extensive notes that have guided physics instruction at Michigan and elsewhere. However, it is his contributions to musical acoustics for which we honor him today.

Gabi was born in Vilna, Poland (now Vilnius, the capital of Lithuania) in 1928. His father, Max Weinreich, was a linguistic scholar of note and a founder of the YIVO Institute of Jewish Social Science. When the Second World War began, Gabi's family came, in serial fashion, to New York City—Gabi and his mother escaping in 1940 via train across Siberia. Gabi studied physics at Columbia University and received a Ph.D. in 1953 for a thesis on atomic physics directed by the legendary Nobel laureate, I. I. Rabi. He subsequently worked on fundamental properties of semiconductors, first at Bell Labs, then, starting in 1960, at the University of Michigan. Perhaps most significant among this early work was the theory and (subsequently) the observation of the acoustoelectric effect, where an ultrasonic wave in a semiconductor gives rise to a direct electrical current.

In 1977 he turned his attention to the acoustics of musical instruments, mainly the piano and bowed strings. What a fortunate day it was for acoustics when he became interested in the problem of super radiance in coupled piano strings! He studied all phases of the physical elements: string excitation, string vibration, coupling, and radiation. Later he turned his attention to violins, and his creativity has brought renewed life to this longstanding field of study.

Gabi brought his special style to acoustics—a combination of theory and experiment that imaginatively imports ideas and techniques from one area of physics into another, a willingness to attack traditional problems afresh by returning to first principles, and the ability to present ideas with incisive wit and charm so that information is not only informative but is also entertaining. His papers at meetings of the Acoustical Society of America (ASA) are classics, both for the scholarly content and the passion with which he delivers them.

His seminal papers on violin acoustics have steered the way we think about the subject and helped to guide the research of others. One such paper in the *Journal of the Acoustical Society of America* (JASA) was "Sound hole sum rule and the dipole moment of the violin" (1977) which defined radiativity and showed how a vibrating hollow shell with a sound hole would radiate as a dipole at low frequencies. Another was his paper on "Directional tone color" (1997) which formed the basis of his patented invention of the directional tone color (DTC) loudspeaker for accurate reproduction of the sounds of violins and other musical instruments. His Klopsteg Memorial Lecture to the American Association of Physics Teachers (1992) on "What science knows about violins—and what it does not know" was published in the *American Journal of Physics.*

His *Scientific American* cover story on "The Coupled Motion of Piano Strings" (1979), based on his earlier JASA paper (1977), is one of the most widely quoted articles on musical acoustics. It explained the initial sound and the decaying sound of a piano in terms of the horizontal and vertical vibrations of a single string and the coupled vibrations of two or three strings. Most of the research on the piano was done in his own living room before he had developed an acoustics laboratory at the University of Michigan.

Gabi's fame is truly international. His linguistic accomplishments include fluency in French, German, and Yiddish. One of his hobbies is translation of scripture from Hebrew to English (he is a priest in the Episcopal Church). He has collaborated extensively with French scientists at Institut de Recherche et Coordination Acoustique/Musique (IRCAM) and Laboratoire d'Acoustique Musique in Paris. With René Caussé, he created the electronic violin bow, which excites the violin string via the Lorentz force on a current in a magnetic field. The new science entered because the current was computed through

feedback from the string motion itself in such a way as to simulate the stick-slip motion of the bow. After months of work Gabi and René succeeded in achieving the rotating-kink Helmholtzian motion, which led to new insight on the stability of this motion. With Xavier Boutillon, Gabi invented a new way to measure mechanical admittance and applied it to the bridge of a violin. As a result of this and similar work, Gabi received the Foreign Medal of the French Acoustical Society in 1992, and was awarded the first Carleen Hutchins Gold Medal for Lifetime Achievement in Musical Acoustics in 2002.

It is typical of Gabi's style that while working on musical acoustical problems, he should take a little break and invent some pure mathematics. Everyone "knows" that a vector is an arrow—characterized by length and direction, but in his book, *Geometrical Vectors*, Gabi showed that in order to make a logically consistent vector calculus, arrows need to be supplemented with other beasts, which he called "sheaves," "stacks," and "thumbtacks."

In spite of his many accomplishments, Gabi retains his modesty and remains a friend of all of us in the musical acoustics community. MacArthur Foundation awardee violin maker Joseph Curtin wrote in *The Strad* (2000) "Weinreich is one of those joyfully articulate people who speak in publishable sentences. He can carry an idea, in the face of digression and interruption, with the agility of a star football player carrying a ball down the field." He is never too busy to use his remarkable physical insight to help a colleague with a difficult problem. He is always there when we need him.

THOMAS D. ROSSING
WILLIAM M. HARTMANN

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in Physical Acoustics

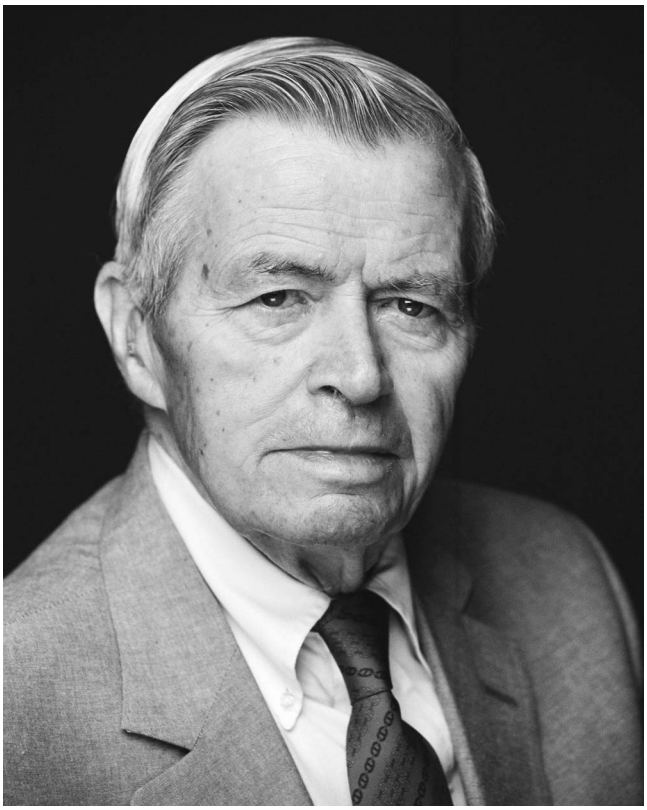
Peter J. Westervelt

2008

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Isadore Rudnick	1975	Julian D. Maynard	1994
Martin Greenspan	1977	Robert E. Apfel	1997
Herbert J. McSkimin	1979	Gregory W. Swift	2000
David T. Blackstock	1985	Philip L. Marston	2003
Mack A. Breazeale	1988	Henry E. Bass	2006
Allan D. Pierce	1991		



CITATION FOR PETER J. WESTERVELT

. . . for fundamental contributions to nonlinear acoustics.

MIAMI, FLORIDA • 12 NOVEMBER 2008

Peter J. Westervelt, Professor of Physics Emeritus at Brown University, was born in Albany, New York, on 19 December 1919. He lived in Paris from ages 4 to 8 and had to learn English to begin school in the U.S. His interest in physics began at age 6, when he built a single-tube radio receiver.

Like many others his age, Peter found his college education interrupted by World War II. Having entered the Massachusetts Institute of Technology (MIT) in 1938, he was recruited to the nascent MIT Radiation Laboratory in 1940, but transferred to the MIT Underwater Sound Laboratory within months, serving as a technician. When the war ended, he resumed his studies, earning a BS degree in 1947, an MS in 1949, and a PhD in 1951, all at MIT and in physics.

Peter joined Brown University in 1951, but was immediately asked to serve as consultant to the Naval Attaché at the American Embassy in London, where he worked with Ed David. Over the next year, he made biweekly trips between his office in Grosvenor Square, London, and Loch Goyle to measure the radiated noise of submarines. This routine was interrupted one time when he was asked to deliver a steak-and-kidney pie to a retired naval officer consulting for the Admiralty. Thus Peter came to meet Captain H. J. Round. While visiting Round's laboratory in a London garret, with experiment in progress, Peter walked in front of a small 18-kHz underwater projector that was actively pinging in air. He heard a powerful low-frequency signal of order 100 Hz, which was also very directional. This caught Peter's attention, prompting the initial reflection that it was due to one of two causes: broadcast of noise, as from an unfiltered power supply, or demodulation occurring in his ears. Peter later recognized a third potential cause: demodulation in air. Thus was born the idea he subsequently developed as the parametric acoustic array.

Within the acoustics community, Peter is famous today for a number of seminal contributions to the theory of finite-amplitude, or nonlinear, waves. His work on radiation pressure and streaming corrected much earlier work and laid the foundations for all subsequent explorations of this branch of nonlinear acoustics.

Since the first work on the subject by Euler in the 1750s, developments in nonlinear acoustics had been restricted to one-dimensional propagation, primarily progressive plane waves. Through the "Westervelt" equation, two- and three-dimensional problems became accessible. By making reasonable approximations, Peter reduced the equations of motion in a compressible fluid medium to a single partial differential equation for pressure. This analytic achievement was seized upon immediately by other researchers, principally to solve problems in which diffraction is important. Generalization of the equation has led to the solution of many other problems in nonlinear acoustics.

The parametric acoustic array, Westervelt's most famous contribution, was first presented at the ASA Meeting in Providence in 1960. Nonlinear interaction of two directional, high-frequency primary waves propagating in the same direction generates difference-frequency sound all along the path of the primaries. The result is a very narrow beam of difference-frequency sound having no side lobes. By contrast, generation of such a low-frequency narrow beam by conventional means requires an enormous transducer. Experiments confirmed Westervelt's theory, and many applications followed, first in underwater sound and later in airborne acoustics, especially in audio engineering. A symposium held in 1968 provided a forum for parametric array research and exploitation. That meeting initiated the International Symposia on Nonlinear Acoustics (ISNA), a series that continues.

All of Peter's major works in acoustics have been published in the *Journal of the Acoustical Society of America* (JASA). These are models of pithiness, witness the seminal paper on the laser-excited broadside array (1973) by Westervelt and his student Richard Larson. Its length is 1 2/3 pages, compared with 2 1/3 pages for the classic expounding the parametric acoustic array (1963).

There is another side to Peter's intellectual life: general relativity. The connection to nonlinear acoustics is the wave equation, one acoustical, the other gravitational. Through

a linearization of the fundamental equations of general relativity, Peter solved these consistently with the three classical tests of Einstein's theory. These notes remain unpublished, but have been circulated at the European Organization for Nuclear Research.

Peter has supervised a number of students in their postgraduate studies. He has always been generous in sharing his keen physical insight. In one case, the supervision was inspired laissez faire, more like a consultation in which Peter gave the general direction, but without prescribing the approach. Once the occasional meeting was concluded, there was time to discuss current events, extracurricular reading, and sports: all a pleasure given Peter's wide reading and his swimming at MIT.

In a profound sense, Westervelt is modest. He has said, in effect, about his inhomogeneous wave equation, paraphrasing Einstein, "Since the mathematicians have invaded the theory, I hardly understand it myself any more." Paraphrasing further, "The main thing is the physical content, not the mathematics." Peter's wife Alice Brown Westervelt has said, "Peter never minded too much about a factor of two here or a factor of π there."

In addition to being named fellow of the ASA in 1953, Peter has acquired other honors. He is a fellow of the American Physical Society. He received the Rayleigh Medal of the Institute of Acoustics in 1985. The present award of the Silver Medal in Physical Acoustics to Peter Westervelt recognizes his pioneering achievements, which have stimulated an enormous growth in nonlinear acoustics as an academic discipline, and which have inspired new applications and devices for use in both air and water.

KENNETH G. FOOTE
DAVID T. BLACKSTOCK
ALLAN D. PIERCE

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Speech Communication

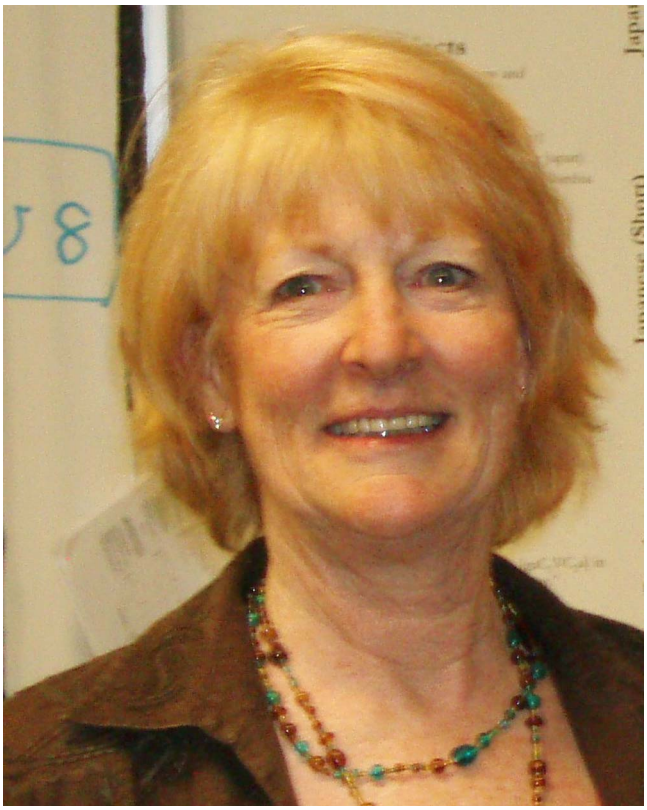
Winifred Strange

2008

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Frank S. Cooper	1975	Peter Ladefoged	1994
Gunnar Fant	1980	Patricia K. Kuhl	1997
Kenneth N. Stevens	1983	Katherine S. Harris	2005
Dennis H. Klatt	1987	Ingo R. Titze	2007
Arthur S. House	1991		



CITATION FOR WINIFRED STRANGE

. . . for contributions to understanding speech preception.

MIAMI, FLORIDA • 12 NOVEMBER 2008

Winifred Strange was born in Red Bluff, California, the younger of two girls born to Verdon Earl Strange and Winifred Rose Strange. Winifred is known by many as 'Pinky,' a nickname coined when her older sister commented on her rosy newborn appearance, and a name that persisted due to the flaming red hair and spirited personality that soon emerged. After graduating from Chico State College in California, Winifred became a graduate student in the Department of Psychology in the early 1970s, where she was acknowledged as one of the University of Minnesota's best teaching assistants. Pinky was the teaching assistant for an intense, yearlong graduate seminar taught by Professor James J. Jenkins, himself a legend at the University of Minnesota. An interdisciplinary group of graduate students interested in the new field of 'cognitive science' adopted the seminar's goal, which was to dissect, sentence-by-sentence, one of the most exciting new theories in cognitive psychology, described in MIT Professor Noam Chomsky's book *Syntactic Structures*. Pinky's highly spirited repartees with the graduate students convinced many of them to alter their careers and devote their lives to this field.

Winifred later married Jenkins, and became step-grandmother to 7, step-great grandmother to 5, and spiritual grandmother to one Christopher Chisolm, son of Richard Chisolm and Teresa Hnath-Chisolm. At the University of South Florida, where Winifred and Jim took faculty positions in 1982, they mentored a new generation of students. In 1998, Winifred took a faculty position at the City University of New York–Graduate School and University Center, where she continues today to play an important role in mentoring doctoral students and bringing them into the ASA fold.

Winifred's scientific contributions are in the area of adult speech perception. She was a pioneer in research on cross-language speech perception, in which she focused on the effects of perceptual training and nonnative language experience. Her work on vowel perception changed our notions of the extent to which speech sounds are processed using dynamic as opposed to static acoustic information. Most experiments conducted in the first 30 years of research on speech perception used stimuli that were generated by computer and contained steady-state formant frequencies. In experiments published in the *Journal of the Acoustical Society of America* beginning in the late 1980s, Strange showed that one could remove the steady-state centers of vowels, leaving only the formant transitions into and out of the vowel, and perception of the vowel remained excellent. These findings on "vowel-less" syllables were surprising, and were initially met with skepticism, but further research produced a body of evidence that weakened alternative interpretations and led to the acceptance of the idea that the dynamic portions of the onsets and offsets of coarticulated vowels contributed importantly to vowel perception. From that point on, researchers had to take into account the fact that listeners were attuned to the dynamic properties of speech sounds, not just their static steady-state values.

One of the biggest challenges of this work is Winifred's continuing commitment to understanding how vowels are perceived in natural co-articulated syllables. Research taking the dynamic specification view in other languages is beginning to show how vowel spectral dynamics and their perception vary across languages. Recent work compares the influence of phonetic context, prosodic context, and speech style on the acoustic characteristics and the cross-language perception of languages such as English, German, Japanese, and French. Her detailed studies of how acoustic similarity relates to perceptual similarity has contributed importantly to an in-depth, and ecologically more valid, understanding of cross-language perception of vowels and of the language-specific influences of consonantal context on acoustic vowel characteristics.

Winifred Strange has contributed greatly to the Acoustical Society of America (ASA) as a past Associate Editor for Speech for the *Journal of the Acoustical Society of America*, as a Member of the Executive Council from 2001-04, and as an active member of the Speech Technical Committee. Winifred has worked behind the scenes on numerous projects, including the ASA's 75th Anniversary Celebration Committee. Many students and junior colleagues have been exposed to her contagious enthusiasm for our field, and she continues to introduce new students to the Acoustical Society of America. Awarding

the Silver Medal in Speech Communication to Winifred Strange recognizes her contributions to the science of speech communication as well as her many contributions to the Acoustical Society of America.

LINDA POLKA
OCKE-SCHWEN BOHN
PATRICIA KUHL

WALLACE CLEMENT SABINE AWARD OF THE ACOUSTICAL SOCIETY OF AMERICA

John S. Bradley

2008

The Wallace Clement Sabine Award is presented to an individual of any nationality who has furthered the knowledge of architectural acoustics, as evidenced by contributions to professional journals and periodicals or by other accomplishments in the field of architectural acoustics.

PREVIOUS RECIPIENTS

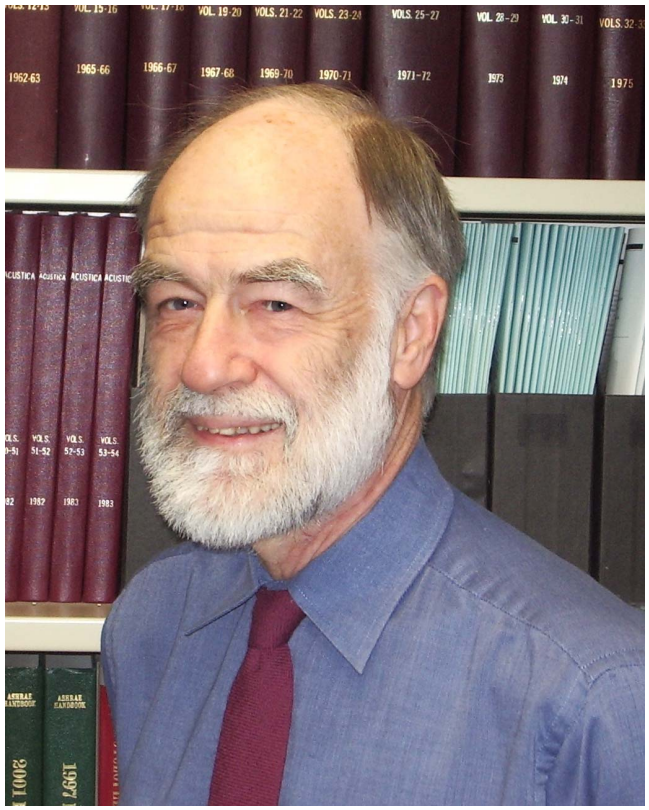
Vern O. Knudsen	1957	Thomas D. Northwood	1982
Floyd R. Watson	1959	Richard V. Waterhouse	1990
Leo L. Beranek	1961	A. Harold Marshall	1995
Erwin Meyer	1964	Russell Johnson	1997
Hale J. Sabine	1968	Alfred C. C. Warnock	2002
Lothar W. Cremer	1974	William J. Cavanaugh	2006
Cyril M. Harris	1979		

SILVER MEDAL IN ARCHITECTURAL ACOUSTICS

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENT

Theodore J. Schultz 1976


1962-63
1965-66
1966-67
1967-68
1969-70
1970-71
1971-72
1973
1974
1975
ACUSTICA
ACUSTICA
ACUSTICA
ACUSTIC
VOLS. 51-52
VOLS. 53-54
1982
1983
1983
ASHRAE HANDBOOK
ASHRAE HANDBOOK

CITATION FOR JOHN S. BRADLEY

. . . for advancing measurement techniques in spaces for speech and music.

MIAMI, FLORIDA • 12 NOVEMBER 2008

John Bradley is a Principal Research Officer at the National Research Council of Canada (NRCC). His career in acoustics began when, after earning a B.Sc., he studied violin response curves for a Masters degree at the University of Western Ontario. His next two degrees from the University of London in England produced theses on sonic booms and the response of electroacoustic feedback channels. At an early stage in his career he demonstrated the breadth of his interest and his ability to tackle diverse research topics.

Following his Ph.D. work at the University of London, John returned to Canada and spent 8 years at the University of Western Ontario where his research focused mainly on traffic noise. In 1980 he joined the acoustics group in the Division of Building Research (now the Institute for Research in Construction) at NRCC where he still is.

During his career at NRCC, he has made significant contributions in several areas of architectural acoustics. As the citation suggests, he is best known for his investigations of the characteristics of different types of rooms, ranging from concert halls, lecture rooms, class rooms, and meeting rooms to offices and rooms in homes. As well as measuring the characteristics of these spaces, John conducted many subjective studies to relate the measurements to probable occupant reactions.

In the 1980s it was evident that using only reverberation time to describe the acoustics of a space was inadequate. The increasing availability of portable computers and digital instruments offered ways to obtain more complicated quantities, such as early-/late-arriving sound energy and useful to detrimental sound ratios. Before these new tools could be put to good use, the effects of different sound sources—pistols and various loudspeakers—had to be determined and software had to be developed. John led the development of consistent measurement procedures and room acoustics software at NRCC that made it easier to measure acoustical characteristics quickly at many places in a room. The software was eventually offered for sale for use with a commercial real-time analyzer.

Using his computer-based measurement systems and the newly developed quantities, he investigated the acoustics of several concert halls and other spaces where the quality of speech or music is important. Much of this work was done in cooperation with researchers from other countries and universities in particular the Concert Hall Research Group.

The acoustical properties of a concert hall are given prominence because of the size and cost of these facilities where the acoustical behavior is fundamental to success. More mundane spaces are of greater importance in everyday life—classrooms, meeting rooms and such. John was the leader of a project investigating the acoustics in classrooms conducted under the auspices of the Canadian Language and Literacy Research Network. Measurements using his techniques in 43 classrooms were related to the fraction of speech understood by children in the same classrooms. The project provided answers to major research questions about classroom acoustics and improved understanding of the different signal-to-noise conditions necessary for children of different ages to understand speech.

John well understands the importance of relating physical measures to subjective reactions. He has conducted many projects to obtain the reactions of listeners to sounds with different early-, late-, and laterally arriving sound energy ratios and related these ratios to the sense of spatial impression and envelopment experienced by the listeners. More recently he has been investigating how well people can understand speech in different kinds of offices. It was often assumed that when the standardized speech intelligibility index was close to zero, speech could not be understood. John's experiments showed that this was not so and that many listeners could still understand speech in such cases. This work will have wide application in all situations where speech privacy is a concern.

John did not confine his research to room acoustics. Changes in construction materials and techniques in Canada created a need for new information about the penetration of

aircraft noise into wood-frame buildings. John led an extensive investigation of sound transmission through typical exterior building elements. Constructions were tested in the laboratory, in a test house near Ottawa airport and in actual homes near Vancouver airport. The study led to an extensive database of sound transmission information and a computer program that simplified the calculation of the aggregate sound isolation that results from choosing different building elements. Users of the program are able to hear simulated aircraft flyovers appropriate to their choice of building elements.

This energetic approach to research led to more than 70 papers in refereed journals and more than 200 other publicly available publications. Despite his prolific, packed career, John found time to contribute to acoustical societies. He has served on the ASA Technical Committee on Architectural Acoustics, organized special sessions and helped organize ASA meetings held in Ottawa. He has been President and Editor-in-chief of the Canadian Acoustical Association (CAA) and is an Editorial Board Member for the *Journal of the Audio Engineering Society*.

John Bradley communicates his research knowledge clearly and appropriately at levels that can be understood by architects, builders and the general public. He is a much sought-after speaker. His reputation and energy have attracted many guest workers to NRCC. He has cooperated well with all of them over the years and has supervised many students during his career. He claims to play the tuba but his colleagues are probably fortunate that he has never demonstrated that skill. He has an evident sense of humor and an informal manner that makes him easy to work with. He is without doubt a worthy recipient of the Sabine medal.

ALF WARNOCK

Session 4aAA

Architectural Acoustics and Noise: Acoustical Issues of Green Buildings

Brandon D. Tinianov, Chair
Serious Materials, 1250 Elko Dr., Sunnyvale, CA 94089-2213

Chair's Introduction—8:00

Invited Papers

8:05

4aAA1. Comparing acoustical requirements of green building assessment systems. Amy Costello and Kenneth Roy (Armstrong World Industries, ABP Technol., 2500 Columbia Ave., Lancaster, PA 17604, aacostello@armstrong.com)

As the Green Building movement gains momentum, more buildings are being evaluated and rated using the environmental assessment systems. A number of these systems now include acoustical requirements or award points for achieving a certain level of acoustical performance. However, how do these environmental assessment systems compare in terms of acoustical requirements? This paper examines and compares some of the leading national and international environmental assessment systems including LEED, BREEAM, Green Globes, and Collaborative for High Performance Schools, and evaluates the importance that acoustical performance is playing in green building rating systems.

8:25

4aAA2. Why has it been so difficult to add acoustics to the criteria of the Leadership in Energy and Environmental Design green building rating system? Ralph Muehleisen (Civil, Architectural, and Environ. Eng., Illinois Inst. of Technol., Chicago, IL 60616, muehleisen@iit.edu)

Leadership in Energy and Environmental Design (LEED) is the most popular green building rating system in use in the United States. There are LEED rating systems for a variety of different building types. The LEED system rates the sustainability and performance in several categories including the indoor environmental quality (IEQ). While LEED incorporates many important IEQ factors including air quality and daylighting in all the rating systems, acoustics has been included only in LEED for schools and the soon to be released LEED for healthcare. The development and subsequent modification of the acoustics portion of LEED for schools highlights some of the problems involved with modifying LEED to include acoustics and some of the issues involved in implementing a LEED rating system that includes acoustics. This paper will discuss why including acoustics within LEED rating systems has been and will continue to be problematic.

8:45

4aAA3. Leadership in Energy and Environmental Design for schools: Unintended consequences? Going "green" and the implications on classroom acoustics. Molly Norris (Threshold Acoust., 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604, mnorris@thresholdacoustics.com)

The advent of the LEED for Schools Minimum Acoustical Performance criteria is a welcome step in the development of green building design. Recent experiences with trying to meet these criteria have led to frustration on the part of architects and the subsequent relaxation of the criteria by the USGBC into a more attainable structure. While the movement to incorporate green design is imperative, the relaxed acoustic criteria may actually be counterproductive in achieving a reasonable acoustic environment for learning. Has good acoustic practice been sacrificed for the green label and is it worth it? A recent case study will be presented.

9:05

4aAA4. A Leadership in Energy and Environmental Design study of Philadelphia Schools; What is old is new again. Felicia Doggett (Metropolitan Acoust., LLC, 40 W. Evergreen Ave., Ste. 108, Philadelphia, PA 19118)

The Philadelphia School District has a stock of over 90 school buildings listed on the National Register that were constructed between 1920 and 1940, most of which are still in use today. With a minimum size of 50 000 sq ft per building, this resource of over 5×10^6sq ft of building area is likely to remain in use for at least another generation. The classrooms in the buildings are very similar in construction consisting of large windows, narrow floor plates, and concrete construction. The School District has also adopted the policy that every one of their new and renovated schools will be LEED certified under LEED For Schools 2007. Since it is not practical to raze these buildings, what can be done to meet the LEED certification? It is surprising how "green" they are. What is old is new again.

9:25

4aAA5. Case study: Leadership in Energy and Environmental Design Platinum office building with innovation credit for acoustic design. Kenneth Roy, Amy Costello, and Anita Snader (Armstrong World Industries, ABP Technol., 2500 Columbia Ave., Lancaster, PA 17604, kproy@armstrong.com)

During the fall of 2006, an existing corporate office building in Lancaster, PA was registered with USGBC for Leadership in Energy and Environmental Design (LEED) certification. In the spring of 2007, this building was awarded the highest achievement, a LEED-EB Platinum rating, only the sixth such award nationwide at that time. This level of performance was due in part to the achievement of an Innovation Credit 1.3 for superior acoustics. The building IEQ postoccupancy survey was evaluated by using the UC Berkeley, Center for the Built Environment, on-line survey tool. The survey results indicated problems with speech privacy performance in the open plan spaces, and a balanced design approach was taken to resolve that issue. Since this was the first such application for an innovation credit on the basis of acoustic design, it was first necessary to provide USGBC with proof that the acoustic environment does affect the health and performance of occupants and that a superior acoustic design can be developed compared to what is generally considered a typical design.

9:45

4aAA6. Green hospitals: Acoustical challenges and opportunities. Kenric Van Wyk (Second Fl., 124 Fulton St. East, Grand Rapids, MI 49503, kvanwyk@acousticsbydesign.com)

The United States Green Building Council and Green Guide for Healthcare have each adopted portions of the new American Institute of Architects (AIA) Interim Sound and Vibration Design Guidelines as the Reference Standard as points for their rating system. This paper gives a brief overview of the AIA guidelines, how they are addressed by these two leading sustainability rating systems, and how they impact the design of future healthcare facilities. Sustainable practices are changing to include acoustics and vibration design. However, even the best intent has some of its own challenges. How will healthcare facility design benefit from these design changes to include sound and vibration and what are the potential pitfalls?

10:05—10:15 Break

Contributed Papers

10:15

4aAA7. Case study of the acoustical analysis for three leadership in energy and environmental design for school applications. Threcia Robinson (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607, trobinson@siebeinacoustic.com), Gary Siebein (Univ. of Florida, Gainesville, FL 32611-5702), Adam Bettcher, and Chris Jones (Siebein Assoc., Inc., Gainesville, FL 32607)

Case studies of the leadership in energy and environmental design (LEED) analysis of an elementary school, a high school, and a college academic building were conducted. The case studies included field measurements of reverberation times within core learning spaces, noise insulation class ratings of walls between classrooms and other rooms as well as between classrooms and the exterior, and noise from building systems. The field measurements demonstrated that reverberation times generally met the criterion as designed. The walls between classrooms and walls between classrooms and other spaces generally required improvements to meet the LEED criteria. Noise from building systems generally met the 0 EQ point criterion. Few if any rooms met the +1 or +2 point requirement without special acoustical design including reduced air velocities, adding silencers in supply and return ductwork and moving terminal boxes and fan coil units out of the classroom spaces. Rooms adjacent to the mechanical rooms also generally exceeded the 45 dBA criterion for background noise due to both sound transmitted through the walls and through short duct runs. The study shows that standard school construction practices will have to be modified to meet the LEED for school criteria.

10:30

4aAA8. Leadership in energy and environmental design for schools 2009 acoustics prerequisite and credit evolution and future direction. Alexis Kurtz (Arup, 155 Ave. of the Americas, New York, NY 10013, alexis.kurtz@arup.com), Daniel Bruck (BRC Acoust. Technol. Consulting, Seattle, WA), David Lubman (DL Acoust., Westminster, CA), and Charles Salter (Charles M. Salter Assoc., Inc., San Francisco, CA)

In May 2008, the U.S. Green Building Council (USGBC) released an updated acoustics prerequisite and credit for inclusion in the Leadership in Energy and Environmental Design for Schools 2009 Rating System. Guidance for the update was provided to the USGBC by the four acousticians in the Indoor Environmental Quality Technical Advisory Group, which focused on identification of key acoustical metrics and development of the credit for greater acceptance within the design and construction community. This paper discusses the process and challenges of balancing market factors, cost, and effective acoustical design for classrooms. The resulting prerequisite and credit are intended to be easily attainable within reasonable cost constraints for classrooms in all geographic regions of the United States. Future direction for classroom acoustics in green design, implications arising from poor classroom design, and the need for greater involvement of the acoustics community in credit development will also be discussed.

10:45

4aAA9. Classroom acoustics in green schools. Pamela Harght and Robert C. Coffeen (School of Architecture and Urban Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

Green schools are becoming increasingly common as school districts continue to experience a rapid growth in population. These schools promote academic success and higher retention rates as well as healthier environments with cleaner air and more daylight as compared to their nongreen counterparts. However, green buildings have become synonymous with poor room acoustics because many of these green goals often result in compromising acoustics and noise control with green building implementing natural ventilation, radiant cooling, and green materials, to name a few. This paper will examine the current conditions of room acoustics in classrooms based on field tests for reverberation, background noise levels, speech intelligibility, and noise isolation from adjacent spaces in addition to the overall design of the classroom. In addition to testing, results from an electronic questionnaire issued to those who teach in these green environments on a daily basis will be discussed. This questionnaire addreses common problems in classroom acoustics in additon to seeking feedback on this new teaching environments. Finally, the role of acoustics in Leadership in Energy and Environmental Design for schools, the process for acoustical consultants for certifying a school, will be discussed.

11:00

4aAA10. Geothermal heat pumps and acoustics for people who do not know much about geothermal heat pumps. Michael Ermann (School of Architecture + Design, Virginia Tech, 201 Cowgill Hall, Blacksburg, VA 24061-0205, mermann@vt.edu)

Geothermal heat pumps operate more efficiently than traditional heat pumps and air conditioners–much more efficiently–and have no noisy outdoor equipment. Yet, after inquiries to mechanical system manufacturers and to the Technical Committee on Architectural Acoustics Green Building Acoustics Subcommittee listserv, I have not found instances where these systems are specified for their acoustic performance. This paper presents a short tutorial on the basics of geothermal systems (also called geoexchange systems or ground-source-coupled heat pumps) geared to those without significant prior knowledge.

11:15

4aAA11. Leadership in Energy and Environmental Design (LEED) for schools: Providing room acoustics for classrooms that is affordable, green, and meets requirements. Joseph F. Bridger, Steven S. Stulgin, and Mathew M. George (Stewart Acoust. Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607)

Leadership in Energy and Environmental Design (LEED) for schools has added classroom acoustics (based on ANSI S12.60) to the prerequisite for all K-12 schools seeking the U.S. Green Building Council LEED certification and several points for achieving additional measures. LEED for schools' room acoustics requirements are the same as the classroom acoustics standard. The U.S. Green Building Council has made it a prerequisite to receive certification for all K-12 schools. This paper explores how to achieve the requirements, while keeping it affordable and green.

11:30

4aAA12. Leadership in Energy and Environmental Design for schools: Designing mechanical system noise control to be green, affordable, and meet requirements. Joseph F. Bridger, Steven S. Stulgin, and Mathew M. George (Stewart Acoust. Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607)

Leadership in Energy and Environmental Design (LEED) for schools has added classroom acoustics (based on ANSI S12.60) to the prerequisite for all K-12 schools seeking the U.S. Green Building Council LEED certification and several points for achieving additional measures. The Background Noise Level requirements of LEED for schools (U.S. Green Building Council) have several different compliance paths to meet the requirements. This paper explores the pros and cons of each and shares experience on how to design the HVAC noise system to meet the requirements for typical green building designs.

THURSDAY MORNING, 13 NOVEMBER 2008 LEGENDS 2, 8:30 A.M. TO 12:00 NOON

Session 4aAB

Animal Bioacoustics: Acoustics of Manatees, Alligators, and Other Topics

Jennifer L. Miksis-Olds, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Chair's Introduction—8:30

Invited Papers

8:35

4aAB1. What can we learn from studying manatee vocalization patterns? Jennifer L. Miksis-Olds (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, jlm91@psu.edu)

Vocalizations are assumed to form the basis of most long range communication in manatees, so studying how animals naturally use their vocalizations can provide insight into many aspects of the manatee acoustic communication system. Vocalizations were recorded from manatees in Sarasota Bay, FL over two nonwinter seasons. Recordings were made under a variety of conditions spanning different habitats, social groupings, behavioral states, environmental sound levels, and observation methods. Analysis of vocalization patterns indicates that manatee vocalization usage is highly context dependent. Patterns were observed to be a function of calf presence, behavioral state, and environmental sound levels. Additionally, vocalizations tended to separate into two distinct call types. Observed patterns will be discussed in the context of communication theory. Further analysis of the dataset has the potential to offer information relating to signal evolution, call function, and impacts of noise on manatee communication and behavior.

8:55

4aAB2. Vocal behavior of Florida manatees during vessel approaches. Douglas Nowacek (Nicholas School of the Environment and Earth Sci., Pratt School of Eng., 135 Duke Marine Lab Rd., Beaufort, NC 28516, dpn3@duke.edu) and Athena Rycyk (Florida State Univ., Tallahassee, FL 32306)

Florida manatees (*Trichechus manatus latirostris*) spend much of their lives in an "urbanized" ocean with high levels of vessel traffic being common during much of the year. To explore the detailed acoustic and motor behavior of manatees during vessel approaches, digital archival tags were attached to animals in the waters of southwest Florida in 2007 and 2008. These tags record sounds produced

by the manatee and other sounds in the environment up to 32 kHz as well as sampling a suite of behavioral and environmental sensors at 50 Hz. The vocal behavior of manatees before, during, and after vessel approaches was investigated. Vocalizations were assigned to the tagged manatee by the higher amplitude and the presence of higher frequency harmonics compared to signals produced by nearby animals. Dolphins have been found to increase their vocalization rate at the onset of vessel approaches, and we have numerous vessel approaches to 20 individuals to evaluate manatee vocal behavior in these contexts. We will present results from single as well as multiple boat approaches and compare these vocal rates to periods without boats. In addition to reporting vocalization rates during vessel approaches, we are investigating the types of signals used by manatees in these situations.

9:15

4aAB3. Technical challenges in the acoustic detection of manatee vocalizations. Christopher Niezrecki (Dept. of Mech. Eng., Univ. of Massachusetts Lowell, One University Ave., Lowell, MA 01854, christopher_Niezrecki@uml.edu)

Over the past several years there has been much interest in detecting the West Indian manatee (*Trichechus manatus latirostris*) in heavily used waterways to reduce the number of collisions with watercraft. The successful detection of manatee vocalizations at an appreciable range in the natural environment is largely dependent on the background noise levels. The background noise consists primarily of boat and snapping shrimp noise as well as other animals and natural sources of noise. In order for effective detection to be realized, advanced signal processing techniques must be used. Within this work, the scope of the problem is presented and some of the technical hurdles are described. Several advanced signal processing algorithms (to reduce background noise) are described and their effectiveness is presented.

9:35

4aAB4. Overall hearing abilities of manatees and ecological acoustical challanges. Edmund Gerstein, Laura Gerstein, Joseph Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), and Steven Forsythe (U.S. Naval Undersea Warfare Ctr. Div. Newport, Newport, RI)

A comprehensive series of underwater psychoacoustic tests was conducted to measure the hearing abilities of West Indian manatees. Pure tones, complex and real world sounds were presented to manatees under controlled acoustical conditions. The results from 30 000 threshold trials measured their audiogram, temporal integration, critical ratios, MMAs and directional hearing. Complementing these investigations, underwater acoustical measurements of manatee habitats and vessel noise propagation were conducted to evaluate acoustical factors that render Florida manatees vulnerable to repeated collisions with vessels. Both low-frequency cutoffs in shallow water and near surface boundary effects limit the propagation of the dominant low-frequency spectra from slow moving boats. Slow speed zones implemented to protect manatees do not address this underlying acoustical challenge. Ironically, the strategy can be counterproductive in turbid waters and can exacerbate the problem, making vessels more difficult or impossible for manatees to detect while increasing transect times and the opportunities for collisions. While manatees are not adapted for hearing the dominant low-frequency spectra from a slow watercraft, they are well equipped to detect and locate higher-frequency modulated sounds. This provides a narrow sensory window through which to alert manatees of approaching vessels. A specially designed alarm to alert manatees is being tested. [Work funded by the U.S. Department of Defence Legacy Resource Management Program, Navy, Florida Inland Navigation District, and Florida Fish and Wildlife Conservation Commission.]

9:55

4aAB5. Problems and issues in low-frequency aquatic communication. Neil Todd (Faculty of Life Sci. Univ. of Manchester, Manchester M60 1QD, UK, neil.todd@manchester.ac.uk)

A recent study concerning the production, transmission and reception of alligator vocalizations has indicated that there are significant gaps in our knowledge of the physiology and physics of low-frequency sound communication in air and water [Todd, J. Acoust. Soc. Am. **122**, 2906–2915 (2007)]. In this paper these issues are discussed, and some possible solutions are considered for aquatic communication by both amphibians and crocodilians in the light of some other recent data on frog vocal behavior in water.

10:15—10:30 Break

Contributed Papers

10:30

4aAB6. Field tests of a directional parametric acoustic alarm designed to alert manatees of approaching boats. Edmund Gerstein, Laura Gerstein, Joseph Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), Josiah Greenewald, and Narayan Elasmar (Florida Atlantic Univ., College of Sci., 777 Glades Rd., Boca Raton, FL 33431)

The efficacy of the alarm was documented during controlled slow boat approaches under two experimental conditions: (1) approaches without an alarm and (2) same boat approaches with an alarm. Experiments conducted in a NASA security area provided controlled environmental conditions and minimized anthrogenic acoustical variables. An array of GPS instrumented buoys were deployed to acoustically and visually grid sites. HD aerial video synchronized with these buoys documented behavior and acoustic conditions at focal animal positions and throughout sites. Bathymetry and propagation measurements indicate shallow water constraints along with Lloyd's mirror effect resulting in significant transmission losses at frequencies of 1 kHz. This could account for the lack of response observed during 94% of the boat approaches without the alarm. In contrast, 100%, of the alarm approaches elicited overt avoidance responses. The change in behavior during approach trials was significantly greater during alarm trials $F=76.74$, df$=1$, $p<0.01$). The distance at which manatees responded was significantly greater during alarm trials ($F=143.42$, df$=1$, $p<0.01$). Applying conservative critical ratio estimates for wideband noise, manatees could not detect no-alarm approaches at distances of 9 m, while manatees responded to alarm

approaches up to 35 m away at levels 18 dB above their estimated critical ratios. [Work funded by the U.S. Department of Defence Legacy Resource Management Program, (Navy), Florida Inland Navigation District, and Florida Fish and Wildlife Conservation Commission.]

10:45

4aAB7. Ultrasonic hearing and vocalizations are used in communication by West Indian manatee mothers and calves. Edmund Gerstein, Laura Gerstein, Joseph Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), and Steve Forsythe (U.S. Naval Undersea Warfare Ctr., Div. Newport, Newport, RI)

Ultrasonic hearing thresholds above 46 kHz were first measured with a test-sophisticated manatee in 1997. Using staircase and method of constants paradigms, pure tones (38–96 kHz) were presented in force-choice two-alternative tests. Repeatable detection thresholds were measured at 38, 46, 56, 66, and 76 kHz. These extended ultrasonic hearing results precipitated additional tests using the CHP-87-L acoustic tag, which regulatory biologists routinely attach to manatees. The tag produces a 75 kHz pulsed tone at 155 dB. The received SPL at manatee ears approximates 149 dB. Over a tag's 2 year life expectancy, permanent selective hearing loss may occur. This is a concern because ultrasonic hearing may be important for directional hearing and communication between mothers and calves. Using a four hydrophone array, calls between wild mothers and calves were sampled at 100 kHz. Individual callers were localized, and some individuals could be identified. Peak source levels of 121 dB *re* 1 Pa at 2 m were directly recorded with significant energy (103 dB) at 46 kHz. These calls are highly directional and were only documented when manatee callers were positioned on axis with a hydrophone's acoustic center. Vigorous duets between calves and mothers had call rates up to 20 calls/min.

11:00

4aAB8. Detection and classification of underwater movements with an in-air biosonar system. Maosen Wang (School of Mech. Eng., Nanjing Univ. of Sci. Technol., Nanjing, 210094, Peoples Republic of China, maosen.wang@gmail.com) and Andreas Zell (Tuebingen Univ., 72076, Tuebigen, Germany)

The aim of this work is to take advantage of the perceived properties of fishing bats. An in-air biosonar system on a mobile robot is presented for the detection and classification of tiny movements under the water surface. Detecting and classifying underwater turbulence with in-air biosonar system allows mobile agent to sense underwater movement without dipping transducers into water. In this work, both fishes—4-cm long goldfish—and waterproof motors (3-W power, 2-cm diameter) are used as sources to generate underwater turbulence. According to real-time feedback results from digital signal processing, the in-air biosonar system on an autocontrolled robot scans water surface with task oriented chirps and sensing strategies. Fishes and motors can be detected and classified in time and frequency domain with a high percentage rate. Experimental result indicates that a mobile robot with in-air biosonar system can achieve the ability of underwater movement detection and classification. However, a precise localization depends heavily on the selection of sensing strategy. [Work supported by NSFC.]

11:15

4aAB9. Water flow detection by larval bullfrogs (*Rana catesbeiana*): Learning and group effects. Sarah A. Stamper and Andrea Megela Simmons (Dept. of Psych., Brown Univ., Providence, RI 02912)

The lateral line system in bullfrog tadpoles (*Rana catesbeiana*) detects water current but mediates orientation away from, rather than toward, the source of current flow. In these experiments, we examined the effects of repeated testing and of group formation on flow detection. Tadpoles (Gosner stages 24–28) were placed in a laminar flow tank in which visual, olfactory, and auditory cues were controlled. Each animal was tested multiple times per day and across multiple days. All animals oriented away from the flow. In addition, the mean latency to adopt a stable position in the tank significantly decreased as a function of repeated testing both within and across days. In a second experiment, animals were tested in groups. Under these conditions, interanimal distances declined significantly at higher flow rates. These experiments suggest that orientation away from water flow in the larval bullfrog may function to facilitate avoidance of current flow produced by approaching predators. In this case, tadpoles should show a preference for locations with reduced water flow and they would benefit from remembering these locations and finding them quickly. Tadpoles would also benefit from forming aggregations in the presence of flow, if potentially generated by a predator, because of dilution effects.

11:30

4aAB10. Sound source localization by the plainfin midshipman fish (*Porichthys notatus*). David Zeddies (Marine Acoust. Inc., 4100 Fairfax Dr., Arlington, VA 22203, davidzeddies@gmail.com), Richard Fay (Loyola Univ. Chicago, Chicago, IL 60626), Peter Alderks, Kiel Shaub, and Joseph Sisneros (Univ. of Washington, Seattle, WA 98195)

Sound source localization of the midshipman fish (*Porichthys notatus*) was studied using the phonotactic response of gravid females to synthetic advertisement calls. Playback experiments were conducted in a 12-ft-diameter outdoor concrete tank at the Bodega Marine Laboratory using a J-9 transducer placed at the center of the tank. The sound field in the tank was measured at 5-cm intervals using an eight-hydrophone array to measure the pressure gradients from which particle motion vectors were calculated. The acoustic measurements confirmed that the J-9 projector was operating as a monopole source. Animals were released 90 cm away from the sound source, and 60 positive phonotactic responses from naïve gravid females were video taped and analyzed. The phonotactic responses consisted primarily of straight to somewhat curved tracks to the sound source. Abrupt changes in trajectory to the sound source were rarely observed. The results confirm that fish can locate sound sources in the near field.

11:45

4aAB11. Recognizing holographic perception. Richard Rikoski (Naval Surface Warfare Ctr., Panama City, 110 Vernon Ave., Panama City, FL 32407)

Recent robotics work has demonstrated that the holographic property of synthetic aperture sonar images can be used to enable object or terrain recognition using real aperture sonars. For instance, it was shown using Gulf of Mexico data from the small synthetic aperture minehunter that terrain can be recognized using as little as a single sonar element and a single ping. Similar capabilities have been ascribed to marine mammals, raising questions about whether the techniques are similar. This paper describes the basics of holographic perception and a series of experiments, which can be used to rule out its use.

Session 4aEA

Engineering Acoustics: Acoustics for Battlefield Operations and Homeland Security II

Michael V. Scanlon, Chair
U. S. Army Research Lab., 2800 Powder Mill Rd., Adelphi, MD 20783-1197

Chair's Introduction—8:00

Contributed Papers

8:05

4aEA1. A biomimetic acoustic system for threat detection and localization. Socrates Deligeorges, Christian Karl, Leah Field, Shuwan Xue, Aaron Soloway, Lee Lichtenstien (BioMimetic Systems, 810 Memorial Dr. Ste. 106, Cambridge, MA 02139), Aleks Zosuls (Boston Univ., Boston, MA 02215), Tyler Gore (BioMimetic Systems, Cambridge, MA 02139), and Allyn Hubbard (Boston Univ., Boston, MA 02215)

As part of our development efforts to transition cutting edge algorithms to practical devices for use in the field, hardware and software systems using the biomimetic approach are being designed for real world battlefield conditions. A new digital system has been developed that not only improves on existing sniper detection and localization technology but also enables many other capabilities useful for enhanced situational awareness. Capabilities include detection of vehicles and personnel sounds such as speech and footsteps. Systems in development include soldier-worn, vehicle-mounted, and robot-mounted systems, as well as unattended ground sensor systems. Our systems can integrate acoustic target data with GPS and other sensor information using simple GUIs. These systems are modular with multiple interface ports for USB, Ethernet, RS-232, and standard audio jacks. As these systems are transitioned to dismounted soldiers, power, weight, and size become the driving factors in design choices. Additional capabilities such as tracking/identifying a variety of acoustics targets, enhancing warfighter hearing, and fusing information with other sensors (for no additional increase in size, power, and weight) can make these acoustics systems an even more valuable asset. Authors associated with BMS may realize financial profit with the success of this work.

8:20

4aEA2. Enhancing the functionality of a biomimetic acoustic direction finding system with back-end algorithms. Yirong Pu (VLSI and Neural Net Systems Lab., Elec. and Comput. Eng. Dept., College of Eng., Boston Univ., Boston, MA 02215, yrpu@bu.edu), Sarah Kelsall, and Allyn Hubbard (Boston Univ., Boston, MA 02215)

Increasing battlefield awareness can improve both the effectiveness and timeliness of response in hostile military situations. The front-end of the existing biomimetic acoustic direction finding system provides the back-end system with spike trains. The back-end algorithms, transportable to FPGA platforms and other general-purpose computers, are tailored to extract specific information for a variety of possible applications. The gunshot classification and localization (C&L) neural network algorithm can recognize and trace supersonic and muzzle-blast acoustic signals in an environment with a minimum SNR value of 1dB and a temporal resolution of 10 s. The C&L algorithm outperforms cross-correlation algorithms in computational efficiency, memory requirements, and noise robustness. Algorithms under development for speech processing are aimed at determining whether speech is present, identifying the speakers' gender and classifying the language spoken. The gender identification algorithm is composed of single pitch-related measures that are weighted based on their estimated performance. Using a variety of speech record lengthens a type I error of 6–16%.

8:35

4aEA3. Propagation of impulsive sound in the nocturnal boundary layer. Roger Waxler, Carrick Talmadge, Xiao Di, and Gilbert Kenneth (NCPA, Univ. of Mississippi, University, MS 38677, rwax@olemiss.edu)

On clear, dry nights a sound channel, the nocturnal duct, forms in the lowest few hundred meters of the atmosphere. Impulsive sound propagating in a nocturnal duct undergoes severe distortion due to multipath effects and dispersion. At long ranges from the source, the signal from a single impulse is received as an extended wave train beginning with a series of distinct arrivals and ending with a low frequency tail. The nature and form of each of the arrivals will be described. The influence of elevated wind jets and the associated convergence zones will be discussed. Possible applications to signature identification and source ranging will be presented.

8:50

4aEA4. Source-height and frequency dependence in sensitivity analysis of near-ground sound propagation. Chris Pettit (Aerosp. Eng. Dept., U.S. Naval Acad., 590 Holloway Rd., M.S. 11-B, Annapolis, MD 21402, pettitcl@usna.edu) and D. Keith Wilson (U.S. Army Cold Regions Res. and Eng. Lab., Hanover, NH 03755)

Computational forecasts of the near-ground sound pressure level (SPL) are key features of proposed frameworks for designing acoustic sensor networks. In many cases the local weather and terrain will not be known precisely enough to justify high confidence in forecasts of the probability of detection. The sensitivity of SPL forecasts to parameter variations therefore must be known in order to assess the probability of detecting acoustic disturbances in poorly characterized environments. To facilitate these assessments, we have expanded a recent framework for full-field sensitivity analysis (FFSA) throughout the parameter space. This new version continues to employ sampling methods, proper orthogonal decomposition, and cluster-weighted models to develop robust surrogate models for sensitivity analysis. Enhancements shown here include (i) locally linear functions in the cluster-weighted models, (ii) analytical differentiation in place of local response surfaces for computing sensitivities, and (iii) estimation of the forecast uncertainty in the sensitivities. These capabilities are used for FFSA of the near-ground SPL due to a harmonic point source operating at several frequencies. The governing parameters and source height are assumed to vary across wide but physically realistic ranges. The dependence of the forecast uncertainty on various factors is examined.

9:05

4aEA5. Acoustic wave characteristics of coarse and fine grain ground, frequency dependence of material properties. Hasson M. Tavossi (Dept. of Phys., Astronomy Geosciences, Valdosta State Univ., 1500 N. Patterson St., Valdosta, GA 31698.)

Acoustic wave propagation in the ground with both coarse and fine grain materials is investigated. The distinction between the coarse and fine grains is made by the value of the wave number kd, where k is inversely proportional to the wave length and d is associated with the grain size. Wave velocity-dispersion and attenuation are studied at low- and high-frequency limits separately, that is, for kd values much greater and much less than 1, as well as for $kd=1$. It can be shown that the frequency dependence of elastic

wave-velocity, dispersion, and attenuation are related not only to the grain-size and the nature of the contact areas, such as roughness and friction, but also these wave characteristics depend on the grain material. It can be shown that the grain material properties, represented by the elastic moduli, are frequency dependent. This frequency dependence of material properties in turn has effects on the ground wave behavior. Experimental and theoretical findings are presented to show these variations with frequency grain size and grain material. Among other applications, these findings on ground material elastic wave properties could be applied to the battlefield operations such as localization of the near surface explosions.

9:20

4aEA6. Improved time reversal focusing in complex media with inverse filter. Thomas Callaghan (Inst. for Computational and Mathematical Eng., Stanford Univ., Durand Bldg., 496 Lomita Mall, Stanford, CA 94305, tscallag@stanford.edu) and George Papanicolaou (Stanford Univ., Stanford, CA 94305)

We analyze the problem of improving time-reversal refocusing in a complex medium through the use of a noninvasive filter. A source at an unknown location emits a pulse, which is recorded by an active set of, possibly distributed, transducers called the focusing array. The goal is to use only the information available at the focusing array to refocus at the source location while minimizing the received energy away from the source. It has been shown that time-reversal in a random medium exhibits superresolution. However, we investigate further improvement of the resolution by first applying an inverse filter. In general, the construction of such a filter requires invasive measurements on a control array in the vicinity of the source, which we assume are not available. The noninvasive inverse filter utilizes only backscattered signals recorded when the focusing array probes the medium. The result is sharp refocusing at the source comparable to the improvement from use of the invasive filter. Possible applications include communications in complex environments where the goal is to send information to a particular location with low probability of intercept. Temporal compression is also important in communications and we investigate the effect of the inverse filter on this as well.

9:35

4aEA7. Comparison of three wind noise reduction strategies. Richard Raspet, Jeremy Webster, and Jiao Yu (Dept. of Phys. and Astronomy and the Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Box 1848, University, MS 38677)

Wind noise is a serious adverse factor in acoustic detection/location systems. Three types of wind noise reduction devices have been used in the past: streamlined probes, spherical wind screens, and surface mounted devices. When these systems are used outdoors near the surface of the ground, the largest contribution to the wind noise is the interaction of the device or surface with the atmospheric turbulence. The goal of our research program is to develop theories to relate measured wind and turbulence properties to the measured wind noise so that wind noise reduction methods can be optimized for any particular application. In this paper we will present measurements of wind noise reduction with devices using the three strategies listed and compare the reduction achieved. Advantages and disadvantages in a variety of applications will be discussed based on the theoretical predictions.

9:50—10:05 Break

10:05

4aEA8. Development of a portable therapeutic ultrasound system for military, medical and research use. George Lewis, Jr and William Olbricht (Dept. of Biomedical Eng., Cornell Univ., 108 Olin Hall, Ithaca, NY 14853, george@cornellbme.com)

In the past two decades therapeutic ultrasound has obtained attention by the medical community as a tool to relieve arthritis, improve rehabilitation, and enhance wound healing processes. In the research realm, therapeutic ultrasound and its effects on tissue properties are currently being studied to great lengths. For example, researchers are assessing the ability of ultrasound for large molecule transdermal drug delivery, in targeted chemotherapy delivery to brain cancer, and cellular gene-transfer applications. Even though many applications of therapeutic ultrasound exist, the basic instrumentation has not changed much in the past 50 years. Here, we present a novel therapeutic ultrasound system we developed in our laboratory that is capable of producing acoustic power outputs well over the therapeutic range (greater than 50 W), lightweight (under 6 lb), portable (a foot print of $4\times6\times2$ in.^{3}), and rechargeable battery powered. The portable therapeutic ultrasound unit has the potential to replace plug-in medical systems and rf amplifiers used in research, both of which are bulky and burdensome. The portable system is capable of between-office and field service on its long lasting internal battery, making it especially useful for military, ambulatory, and house-call medical applications. [This research was supported by the National Science Foundation, National Institutes of Health, and Transducer Engineering Inc.]

10:20

4aEA9. Phased array for acoustic hail and warning. Timothy McDevitt, Timothy Brungart, Andrew Barnard, and David Jenkins (Penn State Univ., Appl. Res. Lab., P.O. Box 30, State College, PA 16804)

Background is given for the need of a high output acoustic hailing device (AHD) for military use. Those detrimental effects, which require high output power of an AHD, such as refraction, absorption, and insertion loss are briefly reviewed along with associated experimental data. The background for intelligibility is touched to address the importance of frequency response. A prototype phased array AHD, THOR, is then described and characterized via experimental results. [Work supported by the DOD.]

10:35

4aEA10. Fabrication and measurement of polypeptide-based piezoelectric composite polymer film. Dawnielle Farrar, James E. West (Dept. of Mater. Sci. and Eng., Johns Hopkins Univ., Baltimore, MD 21218), Ilene J. Busch-Vishniac (Johns Hopkins Univ., Baltimore, MD 21218), and Seungju M. Yu (McMaster Univ., Hamilton, ON L8S 4K1, Canada)

A new class of polymer composite piezoelectric materials based on piezoactive biopolymer, poly(-benzyl,L-glutamate) (PBLG), and a matrix polymer, poly(methylmethacrylate) (PMMA), is presented. The composite polymer offers the possibility of decoupling the electrical and mechanical properties, therefore improving the overall system efficiency. By simultaneously poling and curing (at ambient conditions) PBLG and methylmethacrylate (MMA) mixture solutions, we fabricated a flexible composite film with a significant portion of the PBLG molecules oriented normal to the film surface. This film exhibited moderate piezoelectricity ($d33=20$ pC/N), and its mechanical characteristics were similar to those of low molecular weight PMMA. Present samples are about 40 μm thick with tensile strength of 13 MPa and Young's modules of 450 MPa. The frequency response is ± 3 dB from 0.1 to 7 kHz and at 1 kHz is linear with acceleration more than a 10 dB range.

10:50

4aEA11. Analitical modeling of piezoelectric ceramic transducers based on coupled vibration analysis. S. Boris Aronov (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Dartmouth, 151 Martine St., Fall River, MA 02723)

The energy method for analyzing piezoelectric ceramic transducers [B. S. Aronov, J. Acoust. Soc. Am. **117**, 210–220 (2005)] is applied to the treatment of transducers with mechanical systems that can be considered as two-dimensional. Analysis is made following the general outline of the theory of coupled vibration in two degree-of-freedom systems and its extension to calculating resonance frequencies of elastic bodies, as suggested by Gibbe and Blechshmidt [Ann. Phys. **18**, 417–485 (1933)]. The approach to the problem is illustrated with examples of piezoelectric rectangular plates, stripes, and thin-walled cylinders at various orientations relative to crystallographic coordinate system. For all of the examples, the resonance frequencies and effective coupling coefficients are presented as functions of the dimensional aspect ratios. Equivalent electromechanical circuits are introduced, which permit calculation of the transducers performance under different acoustical loading conditions. Practical recommendations for improving the electromechanical properties of the transducers by use of appropriate aspect ratios are discussed.

Session 4aED

Education in Acoustics: Hands-on Experiments for High School Students

Uwe J. Hansen, Chair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Approximately 20 acoustics experiments will be set up, ranging in complexity from simple resonance on a string to ultrasonic levitation. Around 60 students from local area high schools will perform the experiments with the help from ASA scientists and students. Regular ASA conference participants are welcome to the session as long as they do not interfere with student experimentation.

Session 4aNS

Noise, Architectural Acoustics, and ASA Committee on Standards: Workshop on Standardization for Soundscape Techniques: Soundscape and Sound Quality—Measurement and Lexicon

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Fluids Mechanics and Eng., Einsteinufer 25, 10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corp., 30 Lafayette Square, Ste. 103, Vernon, CT 06066

Chair's Introduction—9:00

Invited Papers

9:05

4aNS1. Better soundscapes for all workshops on continuing development of soundscape techniques standardization: Workshop introduction. Brigitte Schulte-Fortkamp (TU-Berlin, Inst. of Fluid Mech. and Eng. Acoust., Einsteinufer 25, D-10587 Berlin, Germany) and Bennett Brooks (Brooks Acoust. Corp., Vernon, CT 06066)

The perception of the soundscape can provide comfort, tranquility, and needed information to the person concerned or may be a source of annoyance. The combination of physical acoustical measurements with the scientific evaluation of perceptual responses to environmental sound, known as soundscaping, is an essential method for the assessment and actualization of positive outdoor environments. The characterization of the acoustical environment includes identifying the nature of the sound sources and the reactions of the perceivers. Soundscaping provides for the measurement, analysis, and design of environmental sound by applying the knowledge of both science and community experts. Much fundamental and practical research has been conducted to establish the bases for the soundscape field. In recent sessions and workshops researchers and practitioners have begun to standardize the available soundscape techniques to allow for more comparison of test and survey results and wider application in design. This workshop continues the discussion, evaluation, and standardization of proposed methods and techniques for soundscape analysis and design. The areas of focus will be a catalog of correlations between physical parameters and perceptual responses, the standardization of a terminology lexicon of soundscape descriptors, and the standardization of measurement procedures.

9:30

4aNS2. Experiments to develop soundscape design methods. Gary Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702, gsiebein@siebeinacoustic.com), Joshua Fisher, Adam Bettcher, Threcia Robinson, Robert Lilkendey, Hyun Paek, Chris Jones, and Reece Skelton (Siebein Assoc., Inc., Gainesville, FL 32607)

A proposed soundscape design method to evaluate acoustical impacts of planning and building projects on the surrounding community as a way to work toward net zero noise impacts is summarized. The concept of net zero impacts is borrowed from ecological planning where buildings are designed to produce as much energy as they use. The idea of a net zero noise impact is one where a building or planning project does not increase the existing ambient noise level during its use. The first step in this process is to define the nature of the existing ambient soundscape in the vicinity of the project. The method includes long term acoustical measurements of ambient sounds, short term detailed measurements of specific acoustic events, focus group meetings with residents and stakeholders to identify and evaluate acoustical issues, methods to map qualitative aspects of the soundscape, computer modeling of various types to

assist in developing design alternatives, evaluation of possible weather effects, qualitative and quantitative assessments of calibrated aural simulations of design alternatives, and postconstruction verification methods. Examples from multiple case studies of large scale infrastructure and building projects are used to document strengths and weaknesses of the proposed method.

9:55

4aNS3. Soundscape standardization of measurement procedures. André Fiebig (HEAD acoustics GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany, andre.fiebig.head@head-acoustics.de)

A wide range of measurement systems, questioning techniques, and analysis methods were already applied in previous soundscape studies over the past decades, but a common consent about meaningful procedures and required measurement systems is missing so far. However, a standardization of measurement procedures with respect to a uniform investigation, documentation, and description of soundscapes is required to allow for the valid comparison and analysis of different studies. Up to now, the diversity of used measurement procedures and technologies hinders the in depth derivation of cross-cultural similarities and dissimilarities in the physics of soundscapes and in the perception and assessment of soundscapes. In this paper, an overview of applied measurement technologies and procedures will be given as well as advantages and disadvantages discussed concerning the main questions: (1) how, (2) where, (3) when, and (4) how long must be measured. An initial step in the direction of a common basis of measurement procedures available for soundscape researchers was taken on the occasion of the workshop in Salt Lake City, 2007. This process must be continued. It will promote new ideas in the field of environmental noise research and enhance the acceptance of soundscape concepts in noise policy.

10:20—2:00

Working Groups, Part 1

Working groups will be organized around the following areas of interest, as expressed by persons responding to the workshop announcement or at meeting registration. Working groups will develop presentations on problems, solutions and recommended actions.

To further develop and refine the methods of soundscaping the following are needed:

Working Group A—Catalog of correlations between physical parameters and perceptual responses
Working Group B—Standardization of a terminology lexicon of soundscape descriptors
Working Group C—Standardization of measurement procedures

Working Groups, Part 2

Working group discussions will continue, focusing on (1) what has been done in the past, (2) what should be done now, and (3) how may this be accomplished? There may be some realignment of the participants in the groups.

Luncheon

Working Groups, Part 3

The combination of physical acoustical measurements with scientific evaluation of perceptual responses to environmental sound, known as soundscaping, is an essential method for the assessment and actualization of positive outdoor environments.

Final Group Presentation

The final plenary session of the Workshop will include presentations by the moderators/recorders of all the working groups on their findings. Discussion will include proposed methods, means and possible venues for follow-up and further action. The final set of recommendations will be published and distributed to all of the participants.

Session 4aSA

Structural Acoustics and Vibration: Causality in Acoustics

J. Gregory McDaniel, Chair
Boston Univ., Aerospace and Mechanical Eng., 110 Cummington St., Boston, MA 02215

Invited Papers

8:30

4aSA1. Causality and mathematical models in vibration and acoustics: A realistic perspective. Allan D. Pierce (Dept. of Mech. Engrg., Boston Univ., Boston, MA 02215, adp@bu.edu)

Acoustic and vibrations applications require stricter causality conceptions than primitive causality (effect never precedes cause). Relativistic causality (nothing travels faster than light) is irrelevant; there is no practical reason that a satisfactory continuum-mechanical model, holding for low to moderate frequencies, be relativistically invariant. Modifying the requirement so that the speed of light is replaced by some acoustic speed is not satisfactory, as the insertion of dissipative mechanisms into any continuum-mechanical model invariably results in small precursors which may precede sonic-velocity wavefronts. These are small, but they are formally nonzero at arbitrarily large distances in advance of the front. The present paper follows Ginzberg (1955) and advocates acoustic causality: the requirements that (i) the precursors die out rapidly with distance, (ii) up to some frequency of interest acoustic disturbances genuinely propagate, with the attenuation per wavelength being substantially less than unity, and (iii) vibrations and propagation are governed by coupled partial differential equations. With these principles as a guide, approximate relations involving extrapolations into the complex plane are derived, and it is shown how error bounds can be placed on applications of various members of a derived family of Kramers–Kronig relations. Very low-frequency relaxation processes account for proportional damping in vibrations.

9:00

4aSA2. Implications of causality on acoustic propagation in highly dispersive bubbly media. Gregory Orris, Dalcio Dacol, and Michael Nicholas (Naval Res. Lab., 4555 Overlook Ave., SW, Washington, DC 20375)

Causality in nearly all physical systems has been a recurrent subject often causing apparent paradoxes since before the 20th century. Linear acoustic propagation through subsurface bubble clouds in the ocean offers an especially challenging physical system within which to investigate issues of causality, and has in the past had several competing fundamental theories. Signal travel times and absorption in such a system exhibit enormous variations depending on the acoustic signal frequency, bubble size distribution, void fraction, and other ambient physical parameters. In a recent work [Orris *et al.*, J. Acoust. Soc. Am. **121**, 3349–3362 (2007)] we presented a correction to some contemporary theories of acoustic propagation in bubbly media to bring them into compliance with the physical law of causality. We will discuss this theory within the context of complex analysis and some of its implications to acoustic signal propagation; also present experimental data, which commensurate with archival data, suggest that the theories are still far from offering a complete description of the physical phenomena surrounding acoustic propagation in bubbly media. Possible modifications will be explored that may lead to a complete self-consistent causal theory of acoustic signal propagation in bubbly liquids. [Work supported by the Office of Naval Research.]

9:30

4aSA3. Implicitly Causal Expansions in the Frequency Domain. J. Gregory McDaniel (Mech. Eng. Dept., 110 Cummington St., Boston, MA 02215)

While the causality statement in the time domain is quite simple, the corresponding statement in the frequency domain involves Hilbert transform relations between the real and imaginary parts of the relevant transfer function. If the transfer function is minimum phase, one can go further and develop Hilbert transform relations between the amplitude and phase. Such relations are important and useful in at least two classes of problems: analysis of frequency-sweep data and active control. In both problems, one usually wishes to simultaneously enforce the causality statement while achieving some other goals. This lecture will explore various ways of accomplishing this and will focus on implicitly causal expansions of transfer functions in the frequency domain. The expansion functions are built directly from Hilbert transform pairs and therefore the series is guaranteed to be causal. One example is a complex Fourier series in frequency in which the coefficients are required to be real valued. Examples will be developed in which the causal series expansions are used to improve the quality of data, by eliminating noncausal noise and to achieve active control. [This work was supported by the Office of Naval Research under Grant N00014-99-1-1017.]

10:00—10:15 Break

10:15

4aSA4. Acoustic Kramers–Kronig relations in the ultrasonic frequency band. Joel Mobley (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Physical manifestations of the principle of causality, Kramers–Kronig (KK) relations, have proven to be adaptable to a wide array of tasks which include measuring fundamental material parameters, establishing the consistency of laboratory data, and building causally consistent physical models. This talk is concerned with the use of finite bandwidth KK relations between and among the components and derivatives of the complex wavenumber in the low megahertz ultrasound band. Special focus is given to their applicability to data from suspensions of resonant scatterers and media exhibiting attenuation with a power-law dependence on frequency. One recent development is the validation of a KK relation for the direct prediction of the group velocity from the attenuation coefficient, which has demonstrated great utility for suspensions with resonant type dispersion. The roles of KK analysis in other recent issues are also discussed, including the apparent negative dispersion in cancellous bone.

10:45

4aSA5. Is an impedance operator necessary causal, and is this an issue of complexity? John J. McCoy (School of Eng., The Catholic Univ. of America, Washington, DC 20064)

An impedance operator describes the mapping of a velocity field across a part of a boundary surface, to the traction field across the same part. Understood to represent the solution of a "direct" problem, i.e., the velocity field describes the problem forcing and the traction field part of the solution, the impedance operator is necessary causal. On the other hand, understood to represent the general solution of an "inverse" problem, i.e., the velocity field is part of the observed solution with the traction field representing the problem forcing, the operator need not be causal. Continuing, a uniqueness theorem that applies to the direct problem assures that the impedance operator thusly defined is unique. The lack of a corresponding theorem for the inverse problem suggests that the impedance operator thusly defined need not be unique. This further suggests requiring causality selects from *multiple* impedance operators, representing multiple solutions to the inverse problem, the one that is unique. This raises two questions. Is the causality that makes the operator unique a requirement of the governing physics? What impact does this have on the concept of impedance as a tool for addressing complexity in dynamical systems?

THURSDAY MORNING, 13 NOVEMBER 2008 LEGENDS 7, 9:00 A.M. TO 12:00 NOON

Session 4aSC

Speech Communication: Production (Poster Session)

Ewa Jacewicz, Chair
Ohio State Univ., Speech and Hearing Sci., 1070 Carmack Rd., Columbus, OH 43210

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbers papers will be at their posters from 10:30 a.m. to 12:00 noon.

4aSC1. Development of temporal characteristics in the speech of hearing impaired preschoolers. Mark VanDam, Nicholas A. Smith, Dana Ide Helvie, and Mary Pat Moeller (Boys Town Natl. Res. Hospital, 555 N 30 St, Omaha, NE 68131, vandamm@boystown.org)

This longitudinal study examined the development of temporal speech properties in hearing impaired and normal hearing children at 4 and 5 years of age. Children repeated a list of target words following the experimenter's model. Measures of duration for onset, nucleus, coda, and syllable were collected for children with normal hearing and children identified early (mean = 3 months) versus late (mean = 30 months) as hearing impaired. Main effects of age and group were observed for nucleus, coda, and syllable but not onset duration. Early-identified children were more similar to normal-hearing children than late-identified peers, despite early-identified children having much less hearing (mean diff. = 40 dB HL). Age and group differences were examined using (1) relationships among acoustic duration measures, (2) phone-by-position accuracy, and (3) spoken word intelligibility to assess the influence of hearing experience on development. Although there was a wide individual variation, results suggest benefits of early identification: early identified hearing impaired children performed more like normal hearing children. Results favor a "delayed-acquisition" over "different-mechanism" model for the development of temporal speech properties in children with hearing loss. [Work supported by NIH-NIDCD T32 DC00013-26; R01 DC006681; and P30 DC04662.]

4aSC2. Children's articulatory constraints inferred from acoustic output: How some speech-sound errors arise. Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420) and Susan Nittrouer (The Ohio State Univ., Columbus, OH)

Our investigations of children's speech acoustics indicate that normally developing children exhibit speech behavior that is not simply scaled adult behavior. Specific anatomical differences between children's and adults' vocal tracts cause these effects: For example, the size of the tongue in relation to the size of mouth decreases with age. The effects of these age-related anatomical differences have been identified for preschool children as young as 1 year of age in phonetic segments, such as /ɹ/, /s/, and /ʃ/. These anatomic differences may lead children to produce subphonemic distinctions, as between /r/ and /w/ in English. Children's anatomical features can also affect sounds generally not considered to be difficult for normally developing children to produce, such as syllable-initial stops. We have noted children 12–18 months old producing intended /g/ that is often transcribed as /d/ by adult listeners. However, fronted velar /g/, which could be the result of a large

tongue size in relation to palate size, exhibits subphonemic distinctions with children's /d/. The behaviors arising from early anatomical features can be preserved into school age and be labeled as aberrant articulation.

4aSC3. Developmental study of the relationship between F0 and formant frequencies. Peter F. Assmann (Univ. of Texas at Dallas, Richardson, TX 75083), Terrance M. Nearey (Univ. of Alberta, Edmonton, AB T6G 2E7, Canada), Sneha V. Bharadwaj, Daniel Hubbard, and Anu Jayaraman (Univ. of Texas at Dallas, Richardson, TX 75083)

There is a systematic relationship between fundamental frequency (F0) and formant frequencies in natural speech across adult talkers, associated with anatomical differences in the length of the vocal tract and vocal folds. In the present study we examined developmental trends in this relationship using a database of vowel recordings from adults and children ranging from 5 to 18 years of age from the Dallas, Texas region. A moderate correlation was found between the geometric mean of the formant frequencies (F1–F3) across all of the vowels for a given speaker (a measure related to vocal tract length) and the geometric mean F0. Overall the correlation was higher in males (r=0.87) than in females (r=0.46) and increased as a function of age. The presence of systematic covariation between F0 and formant frequencies is consistent with studies showing a perceptual contribution of F0 to vowel identification.

4aSC4. Formant lowering in spontaneous crying speech. Donna Erickson (Showa Music Univ., 1-11-1 Kamiaso, Asaoku, Kawasaki 215-8558, Japan, ericksondonna2000@gmail.com), Takaaki Shochi (Gipsa-Labs, Grenoble, France), Hideki Kawahara (Wakayama Univ., Wakayama 640-8510, Japan), Albert Rilliard (LIMSI-CNRS, France), and Caroline Menezes (Univ. of Toledo, Toledo, OH)

Acoustic and articulatory recordings were made at the EMA facilities of NTT Research Laboratories, Atsugi, Japan, for an American English speaker producing (a) spontaneous crying speech and (b) imitation of phrasing of the original crying speech, as control data. Articulatory analysis indicates differences in jaw, lip, and tongue positions for crying speech versus control speech. Acoustic analysis also shows that for crying speech compared with control speech, not only *F*0 increases but also higher formants tend to be lowered. Results of perception tests using the copy-synthesis program STRAIGHT (Kawahara) to morph a continuum of stimuli, keeping *F*0, duration, and intensity constant, suggest listeners to use cues of lowered formants to perceive emotional intensity of an utterance. Recent biophysiological modeling studies suggest that lowered formants may be due to a lowered larynx along with an expanded hypopharyngeal region [e.g., D. Honda, J. Acoust. Soc. Am. (1966); Kitamura *et al.*, Acoust. Sci. Tech. (2004)]. This hypothesis as it applies to crying speech is currently being explored. [This work was supported in part by Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007–2010): 19520371 and SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.]

4aSC5. Cross-linguistic acquisition of vowels: English, Korean, Greek, and Cantonese. Hyunju Chung, Jan Edwards, and Gary Weismer (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, Goodnight Hall 1975 Willow Dr., Madison, WI, 53706, hchung23@wisc.edu)

It has been widely claimed that vowels are acquired earlier than consonants. This is why relatively few studies have focused on investigating acquisition patterns of vowels. However, the acquisition patterns of consonants are not separable from those of vowels and a few studies [e.g., work of MacNeilage (1990)] suggest that vowel acquisition is more complicated than previously assumed. Furthermore, there is relatively little cross-linguistic research on how children master language-specific characteristics of vowels. While children show preferences for native-language vowels by 6 months of age, language-specific patterns in production emerge considerably later. This study examines cross-linguistic variation in the location of shared vowels in the vowel space across four languages: English, Korean, Greek, and Cantonese for 2-year-old, 5-year-old, and adults. The vowels /i/, /u/, and /a/ were elicited in familiar words using a word repetition task. Productions of target words were recorded and transcribed by native speakers of each language. First and second formant frequencies were measured for correctly produced vowels. Language-specific differences in the location of shared vowels were observed in the formant values of both children and adults. [This work was supported by a Fulbright Fellowship to H.C., and NIDCD Grant 02932 and NSF Grant 0729140 to J.E.]

4aSC6. Psychoacoustic measures of stop production in Cantonese, Greek, English, Japanese, and Korean. Timothy Arbisi-Kelm (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr. Madison, WI 53706, arbisikelm@wisc.edu), Mary Beckman, Eunjong Kong (The Ohio State Univ., Columbus, OH 43210), and Jan Edwards (Univ. of Wisconsin-Madison, Madison, WI 53706)

Spectral analyses of stop bursts have revealed that the place of articulation can be predicted based on both invariant and time-varying cues present within the acoustic signal (e.g., Stevens and Blumstein, 1978; Forrest *et al.*, 1988). While prior studies have had some success in uncovering such cues for American English, it is not clear whether these parameters are equally pertinent in stoping consonant classification in other languages. Furthermore, one of the main limitations of a linear acoustic analysis is that it imposes different scales of loudness and frequency on the acoustic signal than does the human ear, thus generating power spectra with different frequency distributions than are produced by the auditory system (e.g., Zwicker 1961; Kewley-Port, 1983). In the present study of word-initial stops produced in Cantonese, English, Greek, Japanese, and Korean, measures derived from a psychoacoustic model of auditory perception were developed in order to more accurately isolate the front cavity resonances of the burst transient. For all languages, peak amplitude frequency and a peak compactness measure successfully distinguished alveolar from velar stops with at least 74% accuracy and were more accurate at stop classification than were spectral moments measures.

4aSC7. The effect of secondary labialization on stop burst spectra. Yunju Suh (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794, yunjusuh@gmail.com)

Secondary labialization increases the size of the cavity before /t/ and /k/ constrictions, rendering their burst spectral energy concentrated in the lower-frequency region than when they are not labialized. However, the peak location and center of gravity of the Spanish and the Korean stop burst spectra before /(w)e/ and /(w)a/ only captured the frequency lowering effect of labialization on /k/ bursts and failed to distinguish between the plain and labialized /t/ bursts. Spectral shapes of both /t/ and /k/ bursts were altered by /w/ but in different ways: The single prominent peak of the compact /k/ burst spectra was located in significantly lower frequency when it co-occurred with /w/, whereas the two prominent peak locations did not differ between the plain and labialized /t/ burst spectra. What differentiated /tw/ from /t/ instead were the nonexistence of the high-frequency peaks (above 4 kHz) and the amplitude increase of the peak around 3 kHz in /tw/. As a result, /tw/ burst had spectral energy concentrated below 4 kHz, unlike diffuse /t/ spectra whose energy was spread over a wider frequency range. The amplitude ratio between the mid-, and high-frequency regions (A_{mid}-A_{hi}) successfully captured this difference between the plain and labialized coronal stop bursts of both Korean and Spanish.

4aSC8. Voice onset time of Korean stops as a function of speaking rate. Eunjin Oh (Dept. of English Lang. and Lit., Ewha Womans Univ., Seoul 120-750, Korea, ejoh@ewha.ac.kr)

This study aimed to explore the effects of varying speaking rate on voice onset time (VOT) of stops in Korean and to see whether contrasts among stop categories are maintained despite these variations. Plain, tense, and aspirated stops in /CVn/ words (C = three stop places, V = three vowels) in a carrier phrase were examined at both normal and fast rates, which were controlled by alternating phrases on a computer screen at intervals of 2 s and 1 s, respectively. Results revealed that plain and aspirated stops moved toward smaller VOT values, while tense stops did not significantly change as a function of rate. These results are in accordance with Kessinger and Blumstein [J. Phonetics **25**, 143–168 (1997)], who found that long-lag and prevoiced categories shifted toward short-lag values at a fast rate in Thai, French, and English. Overlap in VOT ranges between plain and aspirated stops often occurred even at the normal rate, although the degree of overlap was larger at the fast rate, indicating that phonetic cues other than VOTs

play a critical role in characterizing the stop categories in Korean. However, VOTs of tense and plain stops generally did not overlap under either rate condition.

4aSC9. Acoustic features of American-English vowels for English, Chinese, and Korean talkers. Chang Liu and Sangeeta Kamdar (Dept. of Commun. Sci. and Disord., Univ. of Texas, 1 University Station, A1100, Austin, TX 78712)

Sixteen American-English vowels including 12 monothongs and 4 diphthongs were recorded in a phonetic context of /hvd/ from young English, Chinese, and Korean talkers. The Chinese and Korean talkers were bilingual and had stayed in US from 6 months to 6 years. Results of acoustic analysis showed that there was no significant difference in F1×F2 vowel space among the three groups of talkers. In addition, the three groups of talkers showed great similarity in F2/F1 ratio across the 12 monothongs. Vowel durations had significantly greater variability across vowel categories for the Chinese and Korean talkers than for the English talkers, indicating that, besides producing spectral differences among vowels, Chinese and Korean talkers also attempted to generate durational difference among vowels to make each vowel distinguishable from others. More acoustic features such as spectral tilt and formant transition in the diphthongs will be discussed as well as the effects of the second language experience.

4aSC10. Acoustics of epenthetic vowels in Korean loanwords. Hyun-ju Kim (Dept. of Linguist., Stony Brook Univ. (SUNY), S205 SBS, Stony Brook, NY 11794-4376, hyunjkim@ic.sunysb.edu)

Several previous studies reported that epenthetic vowels in other languages were phonetically not the same as lexical vowels: for example, English speakers produce inserted schwas as transitional, which are shorter in duration and lower in *F*1 than lexical schwas [L. Davidson, J. Phonetics **34**, 104–137 (2006)]. This study examines whether Korean epenthetic vowels in the production of non-native phonotactics are real vowels or are transitional elements that appear due to failure to coordinate the two consonant gestures. In the experiment, Korean speakers were asked to produce illegal consonant clusters contained in pseudowords written in English and the same consonants but with an intervening lexical vowel in native words written in Korean. Results showed that epenthetic vowels were not different phonetically (in duration, *F*1, and *F*2) from lexical vowels in native words, which is different from what the previous study of epenthetic vowels in English found. Given the fact that complex codas are never possible in Korean phonology, whereas certain types of complex codas are permissible in English, this finding suggests that distinct native phonology may generate language-specific phonetics and force Korean speakers to produce epenthetic vowels as real.

4aSC11. Phonetic status of Cj combinations in Korean and Spanish. Young-ran An, Jiwon Hwang, and Yunju Suh (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794-4376)

Korean allows combinations of consonant and glide (CG) at onset position but forbids obstruent+liquid (OL) clusters. This is a potential problem for the universal sonority dispersion principle, which can be avoided if the Korean CG combinations are secondary-articulated consonants rather than clusters. To see if this hypothesis is phonetically supported, we compared Korean with Spanish, which allows both CG and OL onsets. *F*2 at the vocoid onset of (C)jV syllables varied as a function of the backness of the following vowel in Korean, whereas it stayed constant in Spanish. The vocoid duration increase from CV to (C)jV was also smaller in Korean. This shows that the Korean /j/ is weak in that it lacks a target frequency and its tongue position is decided by the following vowel. However, this property is not confined to the Cj combination, as CjV and jV syllables behaved in the same way. Our results thus support the idea that the Korean CG is close to secondary-articulated consonants, but the weak realization of /j/ is not a contextual variation after a consonant. One possible explanation for this pattern may come from the inherent property of CV coarticulation in Korean.

4aSC12. The influence of different tones of Mandarin Chinese on the temporal characteristics of nasalization. Wenlang Zhang and David Kuehn (Dept. of Speech and Hearing Sci., Univ. of Illinois at Urbana-Champaign, 901 South Sixth St., Champaign, 61820, wzhang26@uiuc.edu)

This study was designed to investigate whether the four different tones in Mandarin Chinese would influence the temporal characteristics of nasalization. Seven female and seven male subjects were enrolled in this study. Four nasal syllables/ma mi na ni/ in all four tones were embedded in two syllables with initial pressure consonants to form the stimuli. Then the nasal onset intervals, nasal consonant durations, nasal offset intervals, and the whole nasalization durations were segmented in the speech samples. The corresponding durations and ratios for the same syllable were compared pairwisely among the four tones by paired means *t*-test. Multiple statistically significant differences did exist among the four tones in both the durations and the ratios. The most important differences were as follows: Tone 3 has shorter nasal consonant durations than Tones 1 and 4, and shortest nasal offset intervals and whole nasalization durations among the four tones; it also has a larger nasal onset and nasal consonant duration ratios but least nasal offset ratios among all four tones. Different tones of Mandarin Chinese can influence the temporal characteristics of nasalization, and they may serve as extra cues that can help improve perception.

4aSC13. Palatalization in Romanian: Acoustics, perception, and the role of place of articulation. Laura Spinu, Irene Vogel (Dept. of Linguist. and Cognit. Sci., Univ. of Delaware, 46 E Delaware Ave., Newark, DE 19711), and Timothy Bunnell (Nemours Biomedical Res., Wilmington, DE 19803)

Departing from the crosslinguistic generalization whereby the contrast between the plain and palatalized consonants is favored at the coronal, as compared to the labial, place of articulation, recent perceptual studies show native speakers of Romanian displaying higher sensitivity to this contrast in labials. To investigate this unexpected behavior, a production study was conducted with 31 subjects. Five plain and palatalized fricatives (/f, v, z, S, h/) were analyzed in terms of average duration and spectral properties (coefficients of the Bark Cepstrum). A linear discriminant analysis was run using duration and the Cepstral coefficients to predict segment type (plain/palatalized). 78.2% of 3674 tokens were classified correctly; however, the contribution of the duration and spectral properties of each segment showed interesting asymmetries. To summarize, /h/ was most successfully distinguished, /S/ was least successfully distinguished, and /f/ and /v/ were distinguished better than /z/. Our acoustic analysis supports the perceptual findings regarding palatalization in Romanian. Perceptually, the distinction between the plain and palatalized consonants was strongest with labials. This is paralleled by the greater acoustic difference between the plain and palatalized /v/, /f/ versus /z/, /S/. Thus, Romanian exhibits palatalization patterns that appear at odds with the claim that the least marked place is coronal.

4aSC14. Acoustic properties of clear speech in dysarthria. Kris Tjaden, Joan Sussman, Erin Szjata, Grace Liu, Katrina Fulcher, Beth Hilczmayer, Miranda Crumb, and Ken Johnson (Dept. of Communicative Disord. and Sci., Univ. at Buffalo, 3435 Main St., 122 Cary Hall, Buffalo, NY 14214, tjaden@buffalo.edu)

Clear speech produced by neurologically normal speakers has been shown to be more intelligible than conversational or habitual speech, although the magnitude of the effect varies widely among speakers. Speech production strategies underlying the improved intelligibility associated with clear speech have also been described in a number of acoustic studies. Although clear speech effects would be of interest in speech disorders such as dysarthria, where habitual intelligibility may be compromised, few dysarthia studies have examined clear speech [see A. Goberman and L. W. Elmer, J. Comm. Disord. **38**, 215–230(2005)]. The current study reports acoustic characteristics of clear speech for individuals with dysarthria secondary to multiple sclerosis and idiopathic Parkinson's disease. A group of healthy control speakers was studied for comparison purposes. As part of a larger study, all speakers produced a list of 25 Harvard sentences in habitual and

clear speaking conditions. Both segmental and suprasegmental acoustic measures will be reported. Group effects as well as individual speaker trends will be examined. [Work supported by NIH]

4aSC15. Acoustic characteristics of Lombard speech in Parkinson's disease patients. Firas Al-Fwaress, Suzanne Boyce (Dept. of Commun. Sci. and Disord., Univ. of Cincinnati, 3202 Eden Ave., Cincinnati, OH 45267-0379, boycese@email.uc.edu), Kathy Groves-Wright (Cincinnati Veterans Medical Ctr., Cincinnati, OH 45220), Jean Neils-Strunjas (Univ. of Cincinnati, Cincinnati, OH 45267-0379), and Angel Ball (Texas A & M Univ.-Kingville, Kingsville, TX 78363)

Lombard speech refers to the well-known effect by which speakers talk differently in a noisy environment than they would in a quiet environment. This difference involves both increased loudness and acoustic phonetic characteristics that enhance intelligibility. Parkinson's disease (PD) patients who show reduced intelligibility of speech have also been shown to exhibit impaired sensory integration of auditory feedback [Kiran Larsen, JSLHR, 2001]. In this paper, we examine the Lombard effect on speech in noise by PD patients and healthy age-matched controls. The results of acoustic phonetic measures suggest that the Lombard speech behavior of PD patients without dysarthria resembles that of control subjects both in quiet and in noise. However, patients with mild-to-moderate dysarthria show fewer effects of Lombard speech, suggesting that their ability to adjust the intelligibility of their speech in noise is correlated with their level of dysarthria.

4aSC16. Consonant-vowel and diphthong transitions in the speech of persons with dysarthria and of healthy controls. Christina Kuo and Gary Weismer (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706 and Waisman Ctr., 1500 Highland Ave., Madison, WI 53705, kuo2@wisc.edu)

Formant transitions have been of interest to researchers for their contribution to speech perception, sound identification, and inferences to articulatory behavior. In previous work diphthong transitions have been shown to be sensitive to speech motor control deficits found in dysarthria; comparable data regarding transitions between obstruents (i.e., consonants) and vowels are not available. Here consonant-vowel (CV) and diphthong transitions in speakers with dysarthria and in healthy controls are examined to understand whether the effects for diphthong transitions observed in speakers with dysarthria—reduced extent and slope and occasionally lengthened duration—are present for CV transitions. The first part of this study extends work by Weismer *et al.* [J. Acoust. Soc. Am. **121**, 3135 (2007)], who reported a tendency for shallower CV transitions in speakers with dysarthria when compared to healthy controls, but only for syllables where C=dorsal. Given the limited number of dorsal CV transitions in the previous work, more utterances are included to better balance the sample sizes for different places of articulation (i.e., bilabial, alveolar, and dorsal) and also between the two speaker groups. Distributional analyses for several transition measures will be presented for 8 speakers with dysarthria and 18 healthy controls. [Work supported by NIDCD R01 DC003723.]

4aSC17. Acoustic variability and speaker discriminative power of liquids in American English. Xinhui Zhou, Daniel Garcia-Romero, and Carol Espy-Wilson (Elec. and Comput. Eng. Dep., Univ. of Maryland, College Park, MD 20740)

In American English, liquid sounds /r/ and /l/ are the most articulatorily variable and complex sounds. In previous work [Zhou *et al.*, J. Acoust. Soc. Am. **123**, pp. 4466–4481 (2008)], we found that the acoustic variability of /r/ due to two different tongue shapes ("retroflex" versus "bunched") is reflected at the fourth and fifth formants. In order to understand how this articulatory diversity contributes to the intra- and interspeaker variabilities in speech signals of liquid sounds, an analysis of variance (ANOVA) analysis has been performed to investigate the intra- and interspeaker variabilities of liquid sounds across different frequency subbands. The Buckeye database was analyzed in this study, which includes about 30–60 min of broadband conversation speech for each of 40 speakers. The ANOVA analysis showed that liquids have the maximum F-ratios (interspeaker variance/intraspeaker variance) in the range of about 3–5 kHz, and /l/ has a larger maximum F-ratio than /r/. In general, liquids have larger F-ratios than stops and fricatives and have smaller F-ratios than vowels and nasals. The results of phoneme-based speaker identification task will be presented to show the speaker discriminative power of liquids along with the results for other phonemes such as vowels and nasals.

4aSC18. On the relation between locus equations and subglottal resonances. Steven M. Lulich (Speech Commun. Group, MIT, 77 Massachusetts Ave., Rm. 36-595, Cambridge, MA 02139, lulich@speech.mit.edu)

In consonant-vowel transitions, it is well known that the frequency of the second formant at the onset of voicing ($F2_{\text{onset}}$) is linearly correlated with the frequency of the second formant in the middle of the vowel ($F2_{\text{vowel}}$). This correlation, which holds across the vowel contexts for a given consonant place of articulation (POA), is characterized by a regression line (locus equation) with a slope and y-intercept that depend on the consonant POA. Furthermore, for a given POA, slopes and y-intercepts of collections of locus equations are inversely related. The cause of this inverse relation has not yet been explained. In this presentation, we will show that the inverse relation implies that the $F2_{\text{vowel}}$ vs. $F2_{\text{onset}}$ coordinate system is most naturally described as a mathematical translation away from the origin, and that the size of the translation is a function of the POA and of the second and third subglottal resonances.

4aSC19. Does vowel inventory or inventory size condition the articulation of /i/ and /y/? Michel T.-T. Jackson and Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420)

Wood [J. Acoust. Soc. Am. **80**, 391–401 (1986)] and dispersion-focalization theory (DFT) [Schwartz *et al.*, J. Phonetics **25**, 255–86 (1997)] have made differing predictions about variation in the articulation of /i/ and /y/. Wood claimed that there are "language-specific tendencies to either prepalatal or midpalatal tongue positions for palatal vowels … languages contrasting [i] with [y] preferring the prepalatal position for both vowels" [Wood (1986), p. 392]. However, DFT predicts that the high-front vowel /i/ is always driven to near the high-front corner of the acoustic space in order to maximize perceptual-acoustic distinctness from the rest of the vowels in the inventory of the language. In addition, it is possible that languages with smaller vowel inventories have less extreme articulations for "corner" vowels such as /i/ and /y/ overall. In this study, we compare articulatory measures of constriction location, extent, and degree in four languages (with /y/ versus without /y/ and small inventory versus large inventory).

4aSC20. Final consonant voicing and vowel height contrasts in whispered speech. Yana Gilichinskaya and Winifred Strange (Univ. of New York-Grad. Ctr., 365 Fifth Ave., New York, NY 10016-4309, ygilichin-skaya@gc.cuny.edu)

Whispered speech is a naturally distorted speech signal. Whereas it preserves some characteristics of fully phonated speech, some important acoustic cues are removed, diminished, or altered. The prominence of acoustic cues in whispered speech may change due to the physical properties of the whispered speech signal, i.e., decreased intensity, the absence of periodic vibration of the vocal folds, damping of $F1$, shift of the formants, and flattening of the amplitude envelope. Such changes affect the acoustic cues both for vowels (e.g., vowel height) and consonants (e.g., voicing contrasts). The objective of the present project was to explore the acoustic cues for post-stressed syllable-final consonant voicing contrasts and the vowels preceding them in continuous whispered speech of American English speakers and to compare the results with those in fully phonated speech. The stimuli were recorded in the carrier sentence "I'll utter /habVC/ off the list." The consonant pairs included voiced/voiceless bilabial stops /b-p/ and labiodental fricatives /f-v/, each combined with 11 AE vowels /i, ɛe, æa, ʌo, ɔu/. Preliminary results showed that vowels had longer duration in whispered speech than in fully phonated speech. Spectral dispersion, temporal contrastiveness of vowels, $F1$, and vowel duration cues will be reported in the presentation.

4aSC21. Acoustic characteristics of sonorant consonants. Phoebe Allen, Marios Fourakis, and Gary Weismer (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706)

This is a report on the acoustic characteristics of sonorant consonants in Midwestern American English. Ten speakers (five male and five female) of south central Wisconsin origin were recorded producing these consonants. The sonorants [l,r,m,n] plus the velar nasal, when possible, were elicited in the word initial, medial, and final positions. In the initial and final positions,

the immediately adjacent vowel varied over three different stressed vowels. In the final position, the sonorants also occurred postconsonantally, as in the word bottom resulting in their syllabic versions. All but 4 of the 37 words used were real words. Five repetitions of each word were recorded. Measurements of durations, amplitudes relative to the adjacent stressed vowel, the first three formants at the halfway point of each sonorant, and of the first antiformant for the nasals will be presented. In addition, formant transitions were tracked going into and/or coming out of the sonorant consonant and will also be presented and discussed. [Research supported by UW-Madison Graduate School.]

4aSC22. The effect of talker image on phonetic convergence. Molly Babel (Dept. of Linguist., UC Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, mbabel@berkeley.edu)

This paper reports on an experiment that examines the socially motivated status of phonetic convergence. This is done by comparing social and asocial conditions in a lexical shadowing task. The social condition includes a photo of the talker while the asocial condition does not. The lexical shadowing task consists of the presentation of 50 low-frequency monosyllabic words with the vowels /i ae a o u/ six times each. Participants are also recorded reading the word list in pretask and post-task readings. Acoustic analyses are underway to identify differences in level of phonetic convergence in participants' productions in the two conditions ($n = 20$ in each condition) by comparing productions in the shadowing task to the pretask base line recordings. The results of this experiment contribute to the discussion regarding the status of phonetic convergence as a socially motivated process or a natural reflex that stems from the relationship between speech perception and speech production. This work has clear implications for theories of language change in addition to addressing issues regarding exemplar-based theories of speech production, the relationship between speech perception and production, and the perception and encoding of talker-specific characteristics.

4aSC23. Variation in stop consonant voicing in two regional varieties of American English. Ewa Jacewicz, Robert Allen Fox, and Samantha Lyle (Speech Percept. and Acoust. Labs, Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, jacewicz.1@osu.edu)

This study is an acoustic investigation of the nature and extent of consonant voicing of the stop /b/ in two dialectal varieties of American English spoken in south-central Wisconsin and western North Carolina. The stop /b/ occurred at the juncture of two words such as *small bids*, in a position between two voiced sonorants, i.e., the liquid /l/ and a vowel. Twenty women participated, ten representing the Wisconsin and the North Carolina variety, respectively. Significant dialectal differences were found in the voicing patterns. The Wisconsin stop closures were usually not fully voiced and terminated in a complete silence followed by a noisy and voiceless closure release, whereas North Carolina speakers produced mostly fully voiced closures. Further dialectal differences included the proportion of closure voicing as a function of word emphasis. For Wisconsin speakers, the proportion of closure voicing was smallest when the word was emphasized and it was greatest in nonemphatic positions. For North Carolina speakers, the degree of word emphasis did not have an effect on the proportion of closure voicing. The results are discussed in terms of differences in the way voicing is maintained during the closure by the speakers of respective dialects. [Work supported by NIH..]

THURSDAY MORNING, 13 NOVEMBER 2008 LEGENDS 8, 7:45 A.M. TO 12:00 NOON

Session 4aUW

Underwater Acoustics and Acoustical Oceanography: Acoustics of Harbors, Ports, and Shallow Navigable Waterways

Kyle M. Becker, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Chair's Introduction—7:45

Contributed Papers

7:50

4aUW1. Acoustic characterization of harbors and ports. Kyle M. Becker, David L. Bradley, D. Christopher Barber (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804-0030), Michael L Zucker (Penn State Univ., State College, PA 16804-0030), and Andrew T. Kankey (Penn State Univ., University Park, PA 16802)

The use of acoustics has become increasingly important in shallow waterways including ports and harbors. Applications span a range of military, commercial, and recreational interests including mine detection, port protection, and diver interaction. Harbor environments present challenges to acoustic propagation that are not encountered in the deep ocean or more open shelf environments. Characteristics that contribute to the unique acoustic propagation characteristics include typically very shallow depths, highly variable ambient noise levels, the presence of strong scatterers, and tidal effects. In a typical harbor, even a 1- or 2-m tidal variation can represent a change of 10% or more in the water depth. In this work, *in situ* measurements of low-frequency acoustic propagation are presented for a harbor near Newport, RI. The objective was to assess the propagation channel from a finger pier in a confined harbor out to the main shipping channel in Narragansett Bay and incorporate the findings into acoustic propagation models for this and similar environments. Of particular interest are the effects of tidal variations and changing background noise levels on the ability to detect different signals. [Work supported by ONR.]

8:05

4aUW2. Linear array beamforming in harbor environments. Christopher Barber, Kyle M. Becker, David L. Bradley (Penn. State Univ., Appl. Res. Lab., P.O. Box 30, State College, PA 16804-0030), Michael L. Zucker (Penn. State Univ., State College, PA 16804-0030), and Andrew T. Kankey (Penn. State Univ., University Park, PA 16802)

The use of linear arrays of sources to focus low-frequency sound on a target zone of limited spatial extent in a very shallow water environment is a topic of current interest for applications such as harbor defense. While linear array beamforming provides a starting point, the analytical expressions for beam response of a line array found in standard acoustics texts are developed for arrays of compact simple sources in a free-field environment. In contrast, the typical harbor environment is characterized by very shallow water depths, varying bottom conditions, irregular vertical boundaries, and the presence of multiple scattering bodies. In addition to environmental conditions, the size and cost of low- frequency sources limit the validity of the simple source assumption and impose practical limits on the number of

sources and maximum aperture length. Standard beamforming calculations are compared to measured data from a June 2008 harbor acoustic propagation experiment to quantify limitations of the linear array theory in estimating the beam response and three-dimensional sound field generated by an array of a small number of noncompact low-frequency sound sources in a real harbor environment. The relative impacts of environmental factors and source characteristics are also discussed.

8:20

4aUW3. Acoustic propagation in a shallow water duct with large obstructions. Michael Zucker, David Bradley, and Kyle Becker (ARL-Penn State Acoust., Appl. Sci. Bldg., North Atherton St., State College, PA 16802)

The characteristics of shallow water harbor or mooring environments present unique challenges to an acoustic propagation modeler. Established acoustic propagation codes were developed to address problems in deep water or more open shallow water environments where spherical or cylindrical geometries and one-way propagation approximations can easily be employed. In shallow harbors, large structures, including pier pilings and deep-draft vessels, that represent a significant blockage of the water-column propagation path may need to be accounted for. It is of interest to predict the received levels of the acoustic field transmitted past these obstructions. In order to more easily and accurately represent the complicated geometry of the boundary conditions encountered in this environment, a finite element model (COMSOL) is used. Using this model, predictions of the impact on low-frequency acoustic propagation by structures typically encountered in a harbor are made.

8:35

4aUW4. Focusing sound in coastal environments. Andrew T. Kankey, Gary H. Koopmann (Dept. of Mech. Eng., Penn State Univ., 157 Hammond Bldg., Univ. Park, PA 16802, atk127@psu.edu), Kyle M. Becker, and David L. Bradley (Penn State Univ., University Park, PA 16802)

An investigation of the ability to focus low-frequency acoustic energy in very shallow water was carried out in a semiprotected harbor in June 2008. From the end of a finger pier, a linear array of seven acoustic sources was suspended 2 m from the harbor bottom. The mean water depth in the harbor was 11 m with surficial sediments characterized by silts, sandy silts, and clay. The acoustic field was measured by an array of 11 hydrophones oriented in a cross pattern in the harbor. The phones ranged from 100 to 325 m from the center of the array. The objective of this work was to explore alternative methods to time-delay beamforming for focusing the sound field in the harbor. The approach is based on optimizing the phase of the individual sources to provide maximum sound pressure levels at a particular location in the harbor. The optimization scheme requires *a priori* interrogation of the harbor with a number of hydrophones in the area of interest. However, it is designed to be insensitive to vertical boundaries and other obstructions that can be found in typical working harbors. Results from recent field work are discussed.

8:50

4aUW5. Impulse response and acoustic fluctuation of high-frequency signals in a shallow water estuary. Sreeram Radhakrishnan and Alexander Sutin (Ctr. for Maritime Systems, Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030)

Measurements of channel impulse response and its fluctuations in shallow water areas can be used for the prediction of sonar performance, the estimation of active and passive acoustic diver detections, and the prediction of channel characteristics relevant to underwater communication. A shallow water high-frequency propagation experiment was conducted in the Hudson River near Hoboken, NJ by the Maritime Security Laboratory at Stevens Institute of Technology. Acoustic propagation from an omnidirectional emitter to a receiver was investigated in the frequency band from 20 to 100 kHz for distances up to 100 m in a water depth of 8 m. Frequency sweep signal application was allowed for optimal analysis of the channel impulse response in a wide frequency band. Eigenrays corresponding to direct, surface-reflected, and bottom-reflected arrival paths were traced using the BELLHOP program. The fluctuation statistics of narrowband impulse responses for direct, surface-reflected, and bottom-reflected arrivals in different frequency bands was estimated. It was found that signals corresponding to the surface-reflected arrival show considerably higher temporal variation than direct and bottom-reflected arrivals. Comparison of impulse response for near-surface and near-bottom source depths indicates that the bottom-reflected arrival undergoes significantly high attenuation. [Work supported by ONR Project N00014-05-1-0632: Navy Force Protection Technology Assessment Project.]

Invited Papers

9:05

4aUW6. Acoustic research in the Hudson River Estuary at the Stevens Maritime Security Laboratory. Alexander Sutin and Barry Bunin (Stevens Inst. of Techn., Hoboken, NJ 07030, asutin@stevens.edu)

Stevens Institute of Technology has established a Maritime Security Laboratory (MSL) as a national laboratory resource for government, industry, and universities to advance technologies for the protection of USN maritime infrastructure. Experiment instrumentation includes research vessels, a multiplicity of hydrophones and emitters, stand alone acoustic buoys, diver acoustic simulators, unmanned underwater vehicles (UUVs), and precision instrumentation placement capabilities. The in-river experiments are controlled remotely from a Visualization Center on campus. Acoustic research is supported by sound speed profile measurements, integrated video and acoustic tracking of surface events, and global positionlng system tracking of live divers. Recent results include determination of parameters defining the detection distance of a threat: source level, transmission loss, and ambient noise. The combination of acoustic noise with video data for different kinds of ships in the Hudson River enables estimation of sound attenuation in a wide frequency band. The establishment of a library of various estuarine signatures, including divers, boats, ships, UUVs, construction equipment, and so forth, is underway. This knowledge can be used in a variety of intruder detection scenarios and for optimal methods of threat detection. [This work was supported by ONR Project N00014-05-1-0632: Navy Force Protection Technology Assessment Project.]

9:25

4aUW7. High-frequency one-way propagation experiments in Portsmouth Harbor, NH. Thomas C. Weber, Michelle Weirathmueller, and Larry Mayer (Cent. for Coastal and Ocean Mapping, Univ. of New Hampshire, Durham, NH 03824, weber@ccom.unh.edu)

Portsmouth Harbor is a shallow water estuary with 3-m tide heights and 2-m/s tidal currents, exhibiting strong mixing between ocean water and several fresh water inputs. In order to help characterize the limitations for underwater acoustic positioning systems in this environment, high-frequency (40 kHz) one-way acoustic propagation measurements were made at ranges up to 1 km in an area where the maximum depth reached 25 m. Synchronized acoustic transmissions were made from a bottom mounted projector as well as from a near-surface projector mounted on a moving research vessel. The signals were received at a pier-mounted hydrophone. Mea-

surements of signal levels and arrival times show strong signatures from a variety of phenomenon including changes in tide height, changes in sound speed gradients in response to the tidal forcing functions, and turbulent mixing in the water. Each of these will be discussed in relation to signal fluctuations and constraints on measuring pulse arrival times.

9:45

4aUW8. Impact of the harbor environment on the performance of a swimmer. Peter Stein (Sci. Solutions, Inc., 99 Perimeter Rd., Nashua, NH 03063, pstein@scisol.com)

Over the past several years, Scientific Solutions has been developing the Swimmer Detection Sonar Network (SDSN). It is a novel system that networks individual air-backed parabolic transducers which both transmit and receive in narrow beams. In this paper, system performance issues as related to operating in the port/harbor environment will be discussed. This includes effects due to clutter, noise, multipath propagation, refraction, and tides. Data collected from deployments in two different harbor environments will be used to explore the issues. Both harbors have similar water depths and bottom properties, but one is in a tropical climate, where the sound speed is essentially uniform year-round, and the other is a temperate (midlatitude) climate where the sound speed varies significantly depending on the time of year. The tests in the latter environment were conducted during the summer when the sound speed was strongly downward refracting. The methods by which the SDSN design overcomes the performance issues discussed will also be explored [Work supported by ONR.]

10:05—10:20 Break

10:20

4aUW9. Detection and localization performance using a forward scattering tripwire concept for harbor surveillance. Thomas Folegot (NATO Undersea Res. Ctr., Appl. Res., Viale San Bartolomeo 400, La Spezia, Italy)

An innovative technique has been developed to localize intruders crossing an acoustic underwater tripwire surveillance system in shallow water chokepoints. In the context of defence against terrorism, an sample application would be the permanent autonomous detection and localization of small underwater vehicles or divers surreptitiously entering a friendly harbor. The proposed method uses forward propagation and produces an acoustic image of the whole underwater entrance of a harbor, scanning and checking for intrusion in the area of interest. It is anticipated that the system could provide a high image rate and improved performance in noisy, reverberating, and changing environment. This paper discusses results obtained from a scaled tank experiment and shows preliminary results at sea.

10:40

4aUW10. Acoustical characteristics of muddy sediments. William M. Carey and Allan D. Pierce (Dept. of Mech. Eng., College of Engineeering, Boston Univ., 110 Cummington St., Boston, MA 02215)

Muddy sediments found in rivers, deltas, and harbors are classified as slow bottoms and pose a problem for the detection of buried ordnance. The questions for mud discussed here are as follows: First, can the frequency dependence dispersion characteristic be predicted and verified by measurements in areas where buried object detection is required? Second, what is the optimum yet simplest method of calculating sound scattering from buried compact but complex shaped objects? Wood and Weston [Acustica, V**14**, (1964)] indicated that muddy sediments have a compressional speed 3% less than that of water with a linear frequency dependent attenuation (less than that of sand) in the kilohertz range. This paper reviews experiments performed on muddy sediments at frequencies greater than a kilohertz and compares them to recent experimental measurements in pond sediments. The importance of gas content and bubbles is discussed. Gas content and bubbles are shown to be important factors in the reverberation from these slow bottoms. A theoretical treatment of "muddy sediments" is used to estimate the dispersion characteristic mud and to place limits on the use of sonar to find buried objects such as unexploded ordnance in the presence of reverberation due to gaseous inclusions. [Work sponsored by the Strategic Environmental Research and Development Program.]

Contributed Papers

11:00

4aUW11. Card-house theory of mud sediments containing kaolinite and its acoustical implications. Allan D. Pierce, William M. Carey, and Tyrone M. Porter (Dept. of Mech. Eng., Boston Univ., Boston, MA 02215, adp@bu.edu)

A major component of harbor mud is the clay mineral kaolinite, $Al_2Si_2O_5(OH)_4$. Isomorphous substitution, Al atoms occasionally replacing Si atoms, causes a net positive charge of roughly one electronic charge per 400 Si atoms. Kaolinite platelets have diameters and thicknesses of the order of 1 μm and 20 nm. The platelet's positive charge induces a thin negative charge layer of suspended negative ions on both sides, layer thickness dependent on temperature, creating a continuous sheet of transversely aligned longitudinal electrical quadrupoles. Platelets repel each other weakly when placed face to face. The natural configuration is where the only connection between platelets are with the edge of one touching the center of the face of another. The resulting configuration resembles a house of cards. Because the platelets are thin, the porosity is close to unity, and the sound speed is slightly less than the sound speed in water. This type of mud weakly resists shear because any shearing leads to an increase in electrostatic energy within the card house. Theories are presented to answer why bubbles in mud tend to have a flattened shape and to explain the small but nonzero attenuation of sound in mud. [Work sponsored by Strategic Environmental Research and Development Program.]

11:15

4aUW12. Automated passive acoustic intruder detection based on correlation approach. Alexander Sedunov, Laurent Fillinger, and Alexander Sutin (Davidson Lab., Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030, asedunov@stevens.edu)

We consider the problem of passive acoustic detection of intruders in shallow waters using the signals acquired from multiple hydrophones. The suggested approach uses correlation of the acoustic signals recorded by different hydrophones as well as other features such as periodicity. Cross correlation of signals is conducted within a sliding window after passing through a multiband filter bank. This multiband filtering allows leveraging

the intrinsic spectral properties of emitted signals to discriminate intruder emission from acoustic noise sources such as ship traffic typically present in the river. Further heuristic processing exploits the periodic nature of signals such as human breathing, by detecting the beginnings and endings of the diver's inhalation signals. The detector assesses whether the recognized features are consistent with previously documented breathing rates in order to automatically and robustly detect the presence of a diver. The developed approach was applied in diver detection tests conducted in the Hudson River estuary where cross-correlation technique allowed finding the line of bearing to a diver. The challenges to implementation of such method in a standalone acoustic buoy under development at Stevens are also discussed. [This work was supported by ONR Project No. N00014-05-1-0632: Navy Force Protection Technology Assessment Project].

11:30

4aUW13. Finite difference time-domain simulation of Scholte wave generation and propagation along a shallow waterway bottom containing an anomaly. Thomas Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, One Coliseum Dr., University, MS 38677, tmuir@olemiss.edu) and Dwynn Lafleur (Univ. of Louisiana at Lafayette, Lafayette, LA 70504)

Results are presented on a two-dimensional simulation of the generation and propagation of a Scholte wave pulse along a water-sediment interface in a shallow water waveguide, consisting of 12.5 m of water, overlying a like depth of sand sediment. An underwater sound source, 2.5 m above the sediment, generates a low-frequency Ricker wavelet, which propagates as "modal tone bursts" in the waveguide and the sediment. Some of the energy is partitioned into a propagating Scholte wave mode at the sediment interface, a planar surface containing an "anomaly," i.e. a moundlike feature. Acoustic leakage into the elastic sediment is demonstrated, as is the propagation of the Scholte wave impulse moving along the interface and traveling over the anomaly, at a much slower velocity than its acoustic mode counterpart in the water column. The results are presented as a movie sequence, composed of snapshots of the spatial amplitude distribution of the Scholte wave as it propagates. Also presented are the modeled seismometer signals, yielding a two-dimensional particle velocity (hodographs) in the vertical plane at the anomaly, and somewhat beyond it. The significance of low-frequency Scholte waves in the acoustics of extremely shallow water is discussed.

11:45

4aUW14. A model of distant shipping noise. Cathy Ann Clark, Randall T. May, and Kristy A. Moore (NUWC-DIVNPT, 1176 Howell St., Newport, RI 02841)

Shipping noise from high-density areas is believed to migrate down the coastal shelf and propagate to long ranges via the deep sound channel, resulting in an increase in noise level at low angles for submerged receivers. A propagation calculation, which is applicable when a high degree of cycle mixing results in nearly random summation of path effects, is introduced and used to predict low-frequency (1–300 Hz), long range shipping noise. The vertical directionality of the noise is computed by fitting a source distribution within the deep sound channel by the method of least squares. Comparisons to a limited number of low-frequency measured data sets are presented.

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 9, 1:30 TO 4:25 P.M.

Session 4pAA

Architectural Acoustics: Innovative Integration of Acoustic Treatment into Modern Architecture

Scott D. Pfeiffer, Cochair
Threshold Acoustics LLC, 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604

Molly K. Norris, Cochair
Threshold Acoustics LLC, 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604

Chair's Introduction—1:30

Invited Papers

1:35

4pAA1. Development of special acoustic materials for the Experimental Media and Performing Arts Center at Rensselaer Polytechnic Institute: Part 1. Carl Giegold (Threshold Acoust., 53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604, cgiegold@thresholdacoustics.com)

A series of project-specific interior acoustic elements were developed for the eMPAC project. Most prominent is the fabric ceiling in the concert hall, which was developed in tandem with the canopy of a similar material recently installed in London's Royal Festival Hall. Other concert hall finishes were also the subject of much mathematical and physical study during the design process. In the two studios, the finishes respond to the client's acoustic metaphor of a forest clearing, where the acoustic response is relatively alive but highly diffuse at mild and high frequencies. A system of GFRG panels, each of which has several different scales of shaping, was devised and empirically tested in several iterations to achieve the desired result.

1:55

4pAA2. Development of special acoustic materials for the Experimental Media and Performing Arts Center at Rensselaer Polytechnic Institute: Part 2. Zackery Belanger (Kirkegaard Assoc., 8th Fl., 801 W. Adams St., Chicago, IL 60607, zbelanger@kirkegaard.com)

Modern facilities for experimental performance create rigorous acoustical challenges because of the extreme and peripheral nature of their use. Spaces firmly rooted in tradition have acoustic criteria, which are largely known, but experimental spaces must accommodate the traditional, the contemporary, and the not yet conceived. The Experimental Media and Performing Arts Center at Rensselaer Polytechnic Institute, opening this year in Troy, NY, is one such facility. Five major venues—a concert hall, a visual performance studio, a musical performance studio, a theater, and a production suite—are implemented in a complementary harmony of flexibility. The concert hall includes a lightweight Nomex ceiling, which selectively reflects and transmits sound, and the studios utilize adjustable diffusers and absorbers of custom design. Isolation of the critical venues is extensive in anticipation of the use of extreme levels and low frequency sound.

2:15

4pAA3. Case study: Exposed structure acoustical solutions. Kenneth Good and Sean Browne (Armstrong W. I., 2500 Columbia Ave., Lancaster, PA, 17601)

Many new buildings employ an exposed structure design aesthetic and traditional construction elements such as ceilings and walls are being eliminated. These materials perform an acoustical function that is sacrificed in order to achieve visual and sustainability requirements. This presentation will explore several problematic exposed structure spaces and the acoustical solutions that restored proper performance.

2:35

4pAA4. Is that "acoustic?" Unexpected finishes used to shape the acoustic environment. Gregory Miller, Evelyn Way, Byron Harrison, and Richard Talaske (TALASKE, 1033 South Boulevard, Oak Park, IL 60302)

The visual aesthetic desired by many contemporary architects includes surfaces that are incompatible with acoustic requirements for sound absorption and sound diffusion—particularly concave walls and large expanses of smooth, flat surfaces. Additionally, the necessity of sustainable design requires the creative use (and reuse) of unexpected construction materials. Case studies to be presented include the Jewish Reconstructionist Congregation of Evanston, Illinois, the Arena Stage in Washington, DC, The Harman Center for the Arts in Washington, DC, the Steppenwolf Theater in Chicago, Illinois, and the North Campus Auditorium at the Walgreen Drama Center of the University of Michigan at Ann Arbor. These projects will be used to highlight the use of woven textures, shaped precast, and poured-in-place concrete, screens, and repurposed construction waste to solve common (and some uncommon) acoustic challenges.

Contributed Papers

2:55

4pAA5. Stretched, perforated, and auralized: A case study of integrating acoustic treatment into the modern museum to control sound propagation and room acoustics. Ryan Biziorek (Arup, 155 Ave. of the Americas, New York, NY 10013)

Modern museum designs typically require exhibits with AV components. The museum spaces must also be flexible for various types of internal and external events. While the current architectural trend to use glass, concrete, and other sound reflecting materials presents acoustic challenges, there are various techniques to integrate a sound absorbing treatment into a museum facility to allow for flexibility and control of room acoustics. Using materials integrated into multifunction architectural elements, acoustic treatment can be seen as a necessary element within a space. Further modeling and auralization can help the client understand the amount and necessity of an acoustic treatment within various museum spaces. Two case studies will be presented as to how acoustic treatment can be incorporated into the architectural design and use auralization to allow the client to be involved in the acoustic design process.

3:10

4pAA6. Acoustical optimization of shapes and materials used in modern architecture. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

Classic architecture or classically inspired architecture benefits from the fact that scattering surface, in the form of columns, statuary, and relief ornamentation, is an integral part of the architecture. As architecture evolved into using less ornate surfaces in smooth rectilinear and more recently curvilinear forms, this created a need to design scattering surfaces that complement modern architecture. When contemporary scattering surfaces are required, shape optimization, using any reflective material, has proven to be mutually useful for the acoustician and the architect. A shape drawn by the architect is parametrized and then optimized maintaining the desired motif. An iterative computer program that combines the benefits of boundary element and multidimensional minimization techniques will be described and examples will be shown. When absorption is required, microperforation and microslit designs offer a novel solution, using light transmitting plastics and metals. Microperforated wood veneer panels are also available in which the perforations are barely visible at normal viewing distances. A summary of the theory and application of these ideas can be found in work of Cox and D'Antonio [*Acoustic Absorbers and Diffusers: Theory, Design and Application* (Spon, 2004)].

3:25—3:40 Break

3:40

4pAA7. Modified modal analysis for damped enclosures. Buye Xu and Scott Sommerfeldt (Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84602, buye.xu@gmail.com)

Modal analysis (MA) has been widely used to model and understand acoustic fields in undamped or lightly damped enclosures. When additional damping is introduced, MA is known to fail in predicting acoustic quantities close to the boundaries and to exhibit a much slower speed of convergence. Modifications have been introduced into a standard MA approach to partially match the damped boundary conditions and source conditions. Notable improvements in both accuracy and convergence speed are observed when using this modified modal analysis. The theoretical derivation of the model will be discussed briefly, and several examples will be shown to demonstrate the improved accuracy of the model.

3:55

4pAA8. Mitigating low-frequency sound with tunable acoustic chimneys. Bonnie Schnitta and Roy Freeedman (SoundSense LLC, 46 Newtown Ln., Ste. One, East Hampton, NY 11937, bonnie@soundsense.com)

A way to absorb and filter low-frequency sound that is generated by sound sources in one or more enclosures is described. A sound generating room is encapsulated within an outer acoustic absorbing room that contains various openings to several acoustic ducts. The acoustic ducts function as an acoustic waveguide to the open air. At each end of the acoustic duct is an impedance matching tapered acoustic horn. In the sound generating room the horn serves as an acoustic absorber that tapers sound into the acoustic duct. At the opposite end, the horn serves as an acoustic radiator of sound energy to the open air. Both radiator and absorbers may be tuned to transmit frequencies past a certain cutoff. The effect is similar to a pressure release valve. The acoustic waveguides may be coupled to tunable passive or active noise mitigation devices such as Helmholtz resonators and low-frequency noise sources that utilize phase information.

4:10

4pAA9. Innovative green acoustic techniques to control noise ingress and sound propagation and to reduce energy consumption. Ryan Biziorek (Arup, 155 Ave. of the Americas, New York, NY 10013)

The current LEED rating system currently does not give credit for any elements of acoustic design within the point system. Too often, projects can become too focused on obtaining the points within the defined categories while sacrificing the basic elements of acoustic design (i.e., noise ingress from natural ventilation). The acoustic consultant must often find innovative ways to incorporate acoustic design into the building that will support the sustainability goals of the building and can be used to support and/or obtain a LEED point. Three innovative methods of green acoustic design in project case studies will be presented: natural ventilation techniques to maintain privacy, acoustic comfort in indoor working environments, and energy saving techniques for performing arts facilities. If applied to a project in an innovative method, the acoustics of a project need not be compromised and could be used as a LEED credit for innovation in design.

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 2, 1:30 TO 4:00 P.M.

Session 4pAB

Animal Bioacoustics: General Topics

David C. Swanson, Chair

Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Contributed Papers

1:30

4pAB1. Frequency overrepresentation found in the inferior colliculus of Japanese house bats, *Pipistrellus abramus*. Kazuhiro Goto and Shizuko Hiryu (Life and Medical Sci., Doshisha Univ., 1-3 Tatara Miyakotani, Kyotanabe 610-0321, Japan)

In this study, we examined the best frequencies of the inferior collicular neurons of Japanese house bats, *Pipistrellus abramus*. The bats emit long (>10 ms) and shallow FM pulses (quasi-CF) during their search phase, narrowing the frequency range at around 40 kHz. If the narrowed frequency range is important for detecting a small frequency change or frequency modulation, the bats may need a finer frequency resolution than other frequency ranges. So, the neurons responsive to the frequency range at around 40 kHz must be over-represented somewhere in the auditory system, typically in the inferior colliculus. In order to examine this hypothesis, the tonotopic organization of the inferior colliculus of *P. abramus* was measured electrophysiologically by metal microelectrodes. Results show that the frequency range corresponding to their quasi-CF frequency, at around 40 kHz, appears to be over-represented in the inferior colliculus. [Research supported by an ONR grant.]

1:45

4pAB2. Biosonar features extracted from natural frequencies for natural landmark classification. Jinsong Dai, Maosen Wang (School of Mech. Eng., Nanjing Univ. of Sci. & Tech., Nanjing, 210094, People's Republic of China, smewms@yahoo.com), and Rolf Mueller (Shandong Univ., People's Republic of China)

To navigate in a natural environment, it is practical for an artificial agent to use vegetation as landmark. This is the way bats have done this for its long history. For an outdoor biosonar system based on digital signal processing, it is ideal to extract characteristic features directly from natural landmarks. Any physical structure can be modeled as a number of springs, masses, and dampers. Vegetation with trunks, stems, and foliage in open air always sways in a certain degree; it is possible to extract features from a series of characteristic natural frequencies with electrical biosonar system. In this work, a biosonar system is utilized to study features extracted from several plants for mobile agent navigation. Experimental results indicate that there is always enough air turbulence in the outdoor environment and these biosonar features are efficient in classifying plants. Besides using natural frequency features with the same wind condition, a laboring calibration should be made to improve this method robustness in different wind environments. [Work supported by NSFC.]

2:00

4pAB3. Spatial unmasking of pure tones in broadband noise by zebra finches (*Taeniopygia guttata*). Micheal Dent and Jarrod Cone (Dept. of Psych., Univ. at Buffalo, SUNY, Buffalo, NY 14260, mdent@buffalo.edu)

Detecting a signal embedded in noise is known to be enhanced by spatially separating the signal and noise in humans and other animals. This process is known as spatial unmasking and is a part of the larger phenomenon of the cocktail party problem. The exact mechanisms of unmasking are unknown, but binaural processes are thought to be at least partially involved. Most animals that exhibit unmasking are fairly adept at localizing pure tones in space. We wished to study spatial unmasking in an animal that is very poor at sound localization: the zebra finch. Zebra finches were trained using operant conditioning techniques and the psychophysical method of constant stimuli to peck keys for food reinforcement when they detected a tone embedded in a broadband noise masker. Thresholds were obtained for pure tones ranging from 500 Hz to 4000 Hz when the signal and the noise were emitted from the same speaker and when they were emitted from speaker locations separated by 180 deg. Zebra finches showed relatively little unmasking and there was large variation across subjects and frequencies, suggesting that the mechanisms underlying sound localization are related to those that result in spatial unmasking.

2:15

4pAB4. Categorization of budgerigar (*Melopsittacus undulatus*) warble elements. Hsiao-Wei Tu, Edward Smith, and Robert Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742)

The warble song of budgerigars is composed of a variety of elements without any obvious sequential order. Some of the elements also occur as single utterances. A previous study classified warble elements into 42 groups by visual inspection of spectrograms. However, the density of warble (about 140 elements/min, up to 30 min in duration) makes this method both laborious and time-consuming for analyzing a large amount of warble. Here three human raters took 860 elements from 3 birds and sorted the sonograms into 9 general groups with an inter-rater reliability of 89%. Next, these elements were used to train a neural network. This network learned to categorize a large number of warble elements efficiently with 84% reliability (compared to human raters). Further examination of other warble streams revealed that warble elements are not evenly distributed across these nine groups for the same bird, but the relative proportion of different elements in warble categories is similar across three budgerigars. Ongoing studies are examining whether birds vary the proportion of elements in different social contexts to better understand the biological function of this complex vocalization. [This work is supported in part by DC-00046 and DC-00198 to R.D.]

2:30

4pAB5. The effect of social and environmental manipulations on the vocal response rate of budgerigars (*Melopsittacus undulatus*). Peter Marvit (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, pmarvit@gmail.com), Zach Payne (Univ. of Maryland, College Park, MD 20742), Adam Ratner (Bates College, Lewiston, ME 04240), and Robert Dooling (Univ. of Maryland, College Park, MD 20742)

The budgerigar (parakeet) exhibits complex social interactions and relationships. This project investigates the budgerigar "contact call" that is used for both local and distance social cohesion. It is hypothesized that budgerigar vocal response rates will vary with the social "distance" of the caller (e.g., mates greater than conspecific strangers) and that overall rates will increase as the background acoustic environment becomes more complex and naturally rich. To test these hypotheses, vocal response rates of budgerigars were measured to four groups of recorded contact calls, within four acoustic environmental contexts. Three mated pairs were subjects. The results showed huge individual differences. Overall, males responded significantly more than females. There was a trend toward greater responses to mates than self, flockmates, or strangers. There were a large and significant response increases during a familiar flock background and significant decreases during silence. The response vocalizations themselves were analyzed for possible patterns to the different auditory contexts. The results are consistent with a hypothesis of social facilitation, although the individual differences pose a curious problem that a larger sample size may help elucidate. [Work supported by NIH/NIDCD 5R01DC000198 and 2P30DC004664.]

2:45

4pAB6. Transmission fidelity in rhesus monkey (*Macaca mulatta*) "coos" and "screams". Eric Tarkington, Lisa Heimbauer, and Michael Owren (Dept. of Psych., Georgia State Univ., P.O. Box 5010, Atlanta, GA 30302-5010, etarking@ooadvocate.com)

In animal communication, studies of environmental acoustics examine issues such as habitat-specific propagation effects on vocalizations and likely impact on psychologically significant aspects (e.g., localizability and identifiability). The current work compared transmission fidelity in harmonically structured "coos" versus noisy "screams" produced by female rhesus monkeys (*Macaca mulatta*). Calls were broadcast in a temperate, mixed forest using a loudspeaker (Genelec 1029A) positioned either 0.5 m or 1.0 m above the ground, and re-recorded using two microphones (Sennheiser MKH106T) and a digital audio deck (TAS-CAM DA-P1). The "near" microphone remained at a constant distance of 2.5 m from the loudspeaker, while the "far" microphone was positioned at 10, 20, 30, 40, or 50 m. Eight different coos and eight different screams were played four times each at every distance, with cross-correlation values calculated for each pair of near- and far-microphone recordings. Results included that correlations decreased monotonically with far-microphone distance and were higher for calls broadcast at 1.0 m than at 0.5 m. Furthermore, however, correlations were significantly higher for coos than for screams, consistent with other evidence that harmonically structured calls are better suited to functions such as long-distance transmission of caller identity than are noise-based vocalizations.

3:00

4pAB7. Ultrasonic characterization of insect wing reflectivity and wing-beat motion at 200 kHz. David Swanson (Appl. Res. Lab., Penn State Univ., State College, PA 16804), Tom Baker and Ryan terMeulen (Acoust. Program, Pennsylvania State Univ., State College, PA 16802)

A bistatic active ultrasonic sensor is used to detect and identify an insect attracted to a pheromone lure in the beam intersection. The ultrasound beam is a continuous waveform (CW) of approximately 200 kHz to avoid detection by the insect auditory system, which is very sensitive near 40 kHz and below to detect predatory bats. Both amplitude modulation (AM) and frequency modulation (FM) of the CW beam are examined. The reflection factor of the wing material is measured using ultrasonic pulses. The near-field interference nulls of the transmitting beam are modeled and measured to ensure that the lure is not placed in one of these zones. While good AM and FM signals are seen in the received waveform, the pivoting and flapping motions of the wings along with the insect movements create a complex pulse-train like waveform. This is seen to be inherently due to the narrow beam-width for the transducers and a lack of nonspecular scattering from the moving wings. The data show that a simple AM detector can be used to detect the presence of an insect at the lure and to measure the wing-beat frequency. [Work supported by USDA].

3:15

4pAB8. Numerical analysis of flapping sound generated by hovering insects. Yoshinobu Inada (Numerical Anal. Group, Aerosp. Res. and Development Directorate, Japan Aerosp. Exploration Agency (JAXA), 7-44-1, Jindaiji-Higashimachi, Chofu, Tokyo 182-8522, Japan, inada@chofu.jaxa.jp), Hikaru Aono (Univ. of Michigan, Ann Arbor, MI 48109), Hao Liu (Chiba Univ., Chiba 263-8522, Japan), and Takashi Aoyama (Japan Aerosp. Exploration Agency (JAXA), Chofu, Tokyo 182-8522, Japan)

Insect flapping sound is a consequence of flow disturbance generated by flapping wing motion. Spatial and temporal changes of pressures on the wing surface and vortex structures generated by the wing motion are considerable sources of the flapping sound. To analyze such mechanism in the sound generation, we have developed an integrated method combining computational fluid dynamics (CFD) techniques and acoustic analysis. Unsteady flows around a hovering insect are simulated by NS solution-based CFD analysis with a multiblocked, moving-overset grid technique. In the acoustic analysis, monopole and dipole sound sources generated by the wing motion are analyzed by using Ffowcs–Williams and Hawkings method. In this study, numerical analyses of the flapping sounds are carried out for three insects of hawkmoth, honeybee, and fruitfly. The CFD analysis provides a detailed picture of the flow fields around the hovering flyers where complicated vortex aspects and the induced pressure distributions on the wing surfaces show concrete evidence of the sound sources. The acoustic analysis further clarifies the characteristics of the flapping sounds induced by these sound sources, including the sound directivity and its spectrum distribution. These results effectively demonstrate the cause-and-effect relationship in sound generation and thus indicate the availability of this integrated CFD-acoustic method.

3:30

4pAB9. Creating a regionally focused online archive for natural sounds: The Western Soundscape Archive. Jeff Rice (J. Willard Marriott Library, Univ. of Utah, 295 South 1500 East, UT 84112, jeff.rice@utah.edu)

Variations in species distribution and abundance, general topography, as well as regional dialects among species ensure that no two places or environments will sound the same. Regionally focused environmental sound archives can play an important role in documenting these variations. This pa-

per will describe the recent creation of the Western Soundscape Archive (Westernsoundscape.org), an ongoing project led by the J. Willard Marriott Library at the University of Utah that seeks to build a representative and free online resource of animal and environmental sounds of 11 contiguous western United States. The archives primary includes the region's terrestrial vertebrates, as well as targeted ambient recordings. By gathering existing sound recordings, creating new recordings where appropriate, and employing innovative mapping technology, the archive hopes to document and preserve the soundscape of the West and to create a replicable model for other libraries in other regions. [Work supported by an IMLS National Leadership Grant.]

3:45

4pAB10. Relative saliency of envelope and fine structure cues in zebra finch song. Beth Vernaleo and Robert Dooling (Neurosci. and Cognit. Sci. Program, Dept. of Psych., Univ. of Maryland, College Park, MD 20742, bgoldman@umd.edu)

Birdsong provides a useful model for communication and vocal development, and zebra finch song in particular is attractive for its acoustical complexity and repetitive nature. Males sing one song for the purpose of mating display and territory defense, whereas females do not sing. In this study, we are particularly interested in which acoustic features of a male's song are most perceptually salient. Using a repeating background of a single song motif, zebra finches were trained to discriminate changes to two cues in song: increases in intersyllable interval duration (envelope cue) and reversals of single syllables within a song motif (fine structure cue). Results show that zebra finches are able to discriminate changes to fine structure of syllables much more easily than changes to the overall envelope of song, specifically intersyllable intervals. Further experiments have been done using a noise burst modeled song in which song syllables were replaced by frozen random noise bursts of the same duration. Results show that zebra finches are able to discriminate single noise burst reversals, suggesting that they are able to attend to and follow fine structure on a very small time scale. [Work supported by NIH/NIDCD 5R01DC000198 and 2P30DC004664.]

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 3, 1:30 TO 4:20 P.M.

Session 4pEA

Engineering Acoustics, ASA Committee on Standards and Biomedical Ultrasound/Bioresponse to Vibration: High Precision Acoustical Measurements

Victor Nedzelnitsky, Chair

Natl. Inst. of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899-8220

Invited Papers

1:30

4pEA1. A radiation force technique for acoustic power calibration of high intensity focused ultrasound transducers. Subha Maruvada (U.S. Food and Drug Administration, 10903 New Hampshire Ave., Bldg. WO62-2222, Silver Spring, MD 20993)

It is essential to know the acoustic power radiated by transducers used in high intensity focused ultrasound (HIFU) surgery devices for both safety and effectiveness considerations. The power radiated by medical ultrasound transducers usually is measured via radiation force balance (RFB) methods. However, for the high power focused fields encountered in HIFU applications, such measurements can be difficult due to the need for short measurement times to prevent transducer damage, heating of the RFB target, and bubble formation. To address these challenges, a procedure based on pulsed measurements was formulated. High output focused ultrasound transducers were characterized in terms of an effective duty factor, which was then used to calculate the power during the pulse at high drive levels. Two absorbing target designs were used, and both gave reliable results and displayed no damage and minimal temperature rise if placed near the HIFU transducer and away from the focus. The procedure was reproducible up to the maximum power generated of approximately 230 W, representative of the HIFU range, thus allowing the radiated power to be calibrated in terms of the peak-to-peak voltage applied to the transducer.

1:50

4pEA2. Reducing measurement errors: Taking the design of microphones to the limit. Johan Gramtorp and Erling Sandermann Olsen (Brüel & Kjær Sound and Vib. Measurements A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark)

It is impossible to make one microphone that covers all possible applications. In the real world, the properties of a microphone are the result of a number of carefully chosen compromises based on the anticipated measurement situation and the measurement equipment. The electroacoustic properties of condenser measurement microphones are rigidly bound to the physical properties of the microphone cartridge and the electrical properties of the preamplifier. This means that in the design of a measurement microphone the challenge is to find the right balance between the various physical properties of the microphone. However, recent development in simulation models, construction technology, and knowledge of material properties has opened for the design of microphones that has pushed the limits. A new class of general purpose measurement microphones has been born. In this paper, the physical properties that determine the electroacoustic properties of a condenser microphone are reviewed, and the various compromises that must be chosen are discussed. The possibilities and limitations in microphone design are illustrated with examples of the newest microphone designs from Brüel & Kjær.

2:10

4pEA3. Phase frequency responses of measurement microphones and their calibration. Erling Frederiksen (Danish Primary Lab. of Acoust. and Brüel Kjaer, Skodsborgvej 307, 2850 Naerum, Denmark)

For more than 20 years, comparison calibration of microphone phase responses has been a topic of interest. The interest grew up in the 1980s, when it became technically possible to design instruments for measurement of sound intensity and particle velocity. At that time, the calibration frequency range was limited by the first generation of comparison couplers, which typically worked up to 5 kHz. Since then, new applications and needs for wider frequency range or higher accuracy have occurred. Today phase response comparison calibration of half-inch and quarter-inch microphones can be made with couplers from below 20 Hz up to about 20 kHz. The uncertainty is estimated to about 0.010 deg between 20 and 500 Hz from where it increases to some tenths of a degree at 20 kHz. Microphone array applications require large numbers of microphones with essentially equal phase responses. This has increased the need for reference standards, whose absolute phase responses are known. The Danish Primary Laboratory of Acoustics calibrates such reference standards by the reciprocity method. This paper gives an overview of the spread of microphone phase responses, commonly required tolerances, phase calibration methods, and their uncertainties.

2:30

4pEA4. Acoustical impedance measurement. Charles King (Knowles Electronics, 1151 Maplewood Dr., Itasca, IL 60143)

Many components of specialty transducers have extremely tight acoustic impedance tolerances and become difficult to measure in a production environment. Comparing the device to the acoustical impedance of a cavity is one method of determining its acoustical impedance. A precise method of determining the acoustical impedance is described. The technique applies methods normally found in defining cavities for pressure reciprocity calibration to obtain a precision acoustical reference. Pressure reciprocity standards also define how environmental factors influence acoustical properties and these are included into the system to maintain the precision regardless of the environmental conditions or the frequency of operation. Further improvements are obtained by employing differential measurements throughout the measurement and calibration process to eliminate many common errors common to field calibration of microphones. Ultimately the measured data are curve fitted to an equivalent circuit model to extract the desired components from the measurement artifacts. This talk will discuss the method of defining the acoustical components and extracting the desired value from the equivalent circuit model.

2:50

4pEA5. Developments at the National Institute of Standards and Technology in pressure calibration of laboratory standard microphones. Victor Nedzelnitsky, Randall P. Wagner, and Steven E. Fick (Natl. Inst. of Standards and Technol., 100 Bureau Dr., Stop 8220, Gaithersburg, MD 20899-8220, victor.nedzelnitsky@nist.gov)

Improvements in apparatus and procedures for determining the pressure sensitivities of IEC Types LS1Pn and LS2aP laboratory standard microphones by the reciprocity method aim at reducing the effects of significant components of uncertainty. One such component involves the uncertainty with which the front cavity depths of the microphones are known. Therefore, various noncontact methods and procedures for measuring these depths using a depth-measuring microscope have been developed and investigated. Other significant uncertainty components involve the voltage measurements performed during the reciprocity method. Careful consideration of the dependence of measurement uncertainty on voltmeter operational mode, voltmeter range, signal voltage variation with frequency, and the voltages at which range switching occurs can significantly reduce the uncertainty components involving voltage measurements. Another component, involving the uncertainty with which the temperature of the gas in an acoustical coupler is known, can contribute significantly to the uncertainty with which the acoustic transfer impedance in the reciprocity method can be determined. Measurement of acoustical coupler temperature during calibration can significantly reduce this component. Examples of how improvements aimed at reducing these uncertainty components are being systematically incorporated in an evolving test bed are described.

3:10

4pEA6. Primary sinusoidal calibration of accelerometers. Lixue Wu and Peter Hanes (Inst. for Natl. Measurement Standards, Natl. Res. Council, Bldg. M-36, Ottawa, ON K1A 0R6, Canada, lixue.wu@nrc-cnrc.gc.ca)

Accurate vibration measurements require suitably calibrated transducers. A new system for primary sinusoidal calibration of the voltage sensitivity of accelerometers according to Method 1 (fringe-counting) of ISO 16063-11 has been implemented at the NRC Institute for National Measurement Standards. The system includes a new shaker that is capable of generating sufficient acceleration for calibration of accelerometers at frequencies as low as 2 Hz. The design, construction, and testing of the system will be described. Factors that influence the measurement uncertainty of the calibrations such as transverse motion, rocking motion, acceleration stability, distortion, and repeatability have been evaluated and will be discussed.

3:30

4pEA7. Microphone calibration. Gunnar Rasmussen (G.R.A.S. Sound Vib., Skovlytoften 33, 2840 Holte, Denmark, gras@gras.dk)

Microphone calibration accurate acoustical measurements are desirable for legal reasons and for research and development. The uncertainty in acoustic measurements may cause uncertainty in determining improvement in design and development of products. International standards may still be improved for calibration of the measurement technique at low frequencies and at high frequencies especially. The absolute calibration of measurement microphones is dominated by reciprocity calibration. This has been dominating in acoustic metrology for many years and rightly so for the middle frequency range 100–1000 Hz. Other methods are very useful at low

frequencies and at high frequencies with a good overlap to the reciprocity technique and may offer lower uncertainty especially for calibration in the field where the major part of acoustical measurements take place. Microphone comparison, piston phone calibration, scaling in microphone size at high frequencies, electrostatic actuator, and directionality are some of the techniques used in practical calibration work.

Contributed Paper

3:50

4pEA8. Frequency limitations of coupler-based calibrations of the pressure sensitivity of measurement microphones. Allan J. Zuckerwar (Analytical Services and Mater., 107 Res. Dr., Hampton, VA 23666) and Qamar A. Shams (NASA Langley Res. Ctr., Hampton, VA 23681)

Calibration of the pressure sensitivity of measurement microphones in a coupler requires by definition that the incident pressure be uniform over the surface of the microphone diaphragm. This requirement is readily satisfied at frequencies up to the cut in frequencies of the radial modes and, if the microphone is not azimuthally symmetric, the spin modes. Above the lowest natural frequency of either of these modes, the uniformity of the pressure distribution will be destroyed. This effect is alluded to in current standards, ANSI S1.15-2005/Part 2 and IEC 61094-5:2001, but the suggested countermeasure "to use more than one coupler with different dimensions" or to compare "calibrations performed in a variety of other jigs and couplers" does not resolve the problem. Specification of the maximum excitation frequency to preserve pressure uniformity in a coupler is presented for several laboratory standard and working standard microphones.

4:05

4pEA9. Demonstration of an impedance based method for sonar calibration and monitoring. Corey Bachand, David A. Brown, and Boris Aronov (BTech Acoust, LLC and Electro-Acoust. Res. Lab., ATMC/ECE, UMass Dartmouth, 151 Martine St., Fall River, MA 02723)

It is well known that impedance based methods can be used to monitor the performance degradation of sonar transducers. A new approach that involves precise measurements of the voltage (V), current (I) and phase (P), termed the *VIP*-impedance method, has been developed and implemented by the authors for characterizing individual transducer and array performance in sonar systems. Examples of predicted transducer sensitivities and beam patterns on both cylindrical and tonpiltz sonar arrays using the new method are presented and compared with experimental data.

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 10, 1:30 TO 3:20 P.M.

Session 4pED

Education in Acoustics and ASA Student Council: Project Listen Up

James M. Sabatier, Chair
Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, University, MS 38611

Chair's Introduction—1:30

Invited Papers

1:35

4pED1. A Helmholtz resonator experiment for the Listen Up project. Chad A. Greene, Theodore F. Argo, IV, and Preston S. Wilson (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

The Listen Up project seeks to develop an educational booklet and apparatus that will be used to foster interest in acoustics and teach basic acoustical concepts to middle school students. The packaged kit must be low cost; hence this experiment uses a 16 oz plastic water bottle as a Helmholtz resonator, a disposable syringe, and an inexpensive pitch pipe. Understanding the Helmholtz resonator is fundamental to many applications in acoustics. A simple algebraic model of the device, derived from a mechanical mass-spring analogy, relates the volume of air in the bottle to the resonance frequency of the system, and hence the pitch that is produced when one blows over the bottle opening. The volume of air inside the bottle can be easily and accurately controlled by adding water with a graduated syringe. The dependence of pitch on the volume of air in the bottle can be systematically demonstrated, and the validity of the model can be checked by comparing the sound produced to musical notes from the pitch pipe. This experiment yields a physical understanding of a common acoustical phenomenon and inspires further interest and study of more complex phenomena.

1:55

4pED2. Standing waves on a string and electromagnetic induction. Joseph Gladden, III (Dept. of Phys., Univ. of Mississippi, University, MS 38677, jgladden@phy.olemiss.edu)

Standing waves on a string are a popular classroom demonstration which helps students understand such fundamental acoustic concepts as wavelengths, interference, and resonances. The close connection between this demonstration and stringed musical instruments helps bridge the gap between the science classroom and other interests they may have. In this version of the demonstration, we replace the string with a copper wire and add a permanent magnet. This allows the vibrations in the string to be directly translated into an electrical signal which can be broadcast through a standard stereo amplifier and speakers similar to an electric guitar pickup. The

effects of string tension and length can then be explored. Alternatively, an ac can be driven through the wire creating a periodic driving force from the magnetic field of a precise frequency allowing standing waves to be established on the string. More advanced students can also examine the Fourier transform of the signal produced when the wire is struck and find peaks at the resonant frequencies.

2:15

4pED3. Tuning forks, resonators, and beats: Demonstrations for "Project Listen Up". Murray S. Korman (Dept. of Phys., U.S. Naval Acad., 572 C Holloway Rd., Annapolis, MD 21402, korman@usna.edu)

Gearing to an audience of all ages, a demonstration apparatus is presented including student versions of Hermann von Helmholtz's tuning fork sounder, circa 1859 (where an intermittent current in an electromagnetic coil drives the tuning fork, which is located near a resonator), and Alexander Graham Bell's tuning fork experiment, circa 1876 (where a variable resistance circuit due to a slight tuning fork contact in a conductive liquid causes a relay to vibrate at the same frequency as the fork). Using the popularity of a basic mechanical acoustical apparatus and briefly unfolding the histories behind Helmholtz combining several of his tuning fork sounders to generate vowel sounds and Bell's first telephone, students are then given careful explanations of the demonstrations. Next, the focus shifts to well developed basic experiments that students can perform in a laboratory setting with emphasis placed on presenting the details of a carefully thought out scientific approach. "Project Listen Up" (an initiative of the Education in Acoustics Committee) hopes to develop an educational apparatus for a broad range of learners as one of their goals. [Work supported by USNA.]

Contributed Papers

2:35

4pED4. Introducing perceptual coding using a double-blind A/B testing demonstration. Scott Porter (The Graduate Program in Acoust., The Penn State Univ., 217 Appl. Sci. Bldg., State College, PA 16802, scott.porter@psu.edu)

Many of today's college students experience music in compressed formats on a regular basis. While most are familiar with the term "MP3," only a few know of the psychoacoustic principles and perceptual coding schemes which this and other compressed music formats rely upon. In fact, it is not uncommon that they do not realize the audible differences that exist between the CD-quality audio files and those compressed at common commercial bit rates. To illustrate this, the author used freely available software (WINABX and AUDACITY) to demonstrate the differences with double-blind A/B testing as part of an introductory course in acoustics offered at the Pennsylvania State University. The software will be briefly explained, the questionnaire used to acquire the students' responses will be shown, and an interactive demonstration will be included.

2:50

4pED5. Outdoor pulse echo speed of sound measurement for all ages. James M. Sabatier (NCPA, Univ. of Mississippi, University, MS 38655), Charles H. Sabatier (Fairfax County Public Schools, Alexandria, VA 22309), and Celeste S. Taylor (Putnam City Schools, Oklahoma City, OK 73162)

One of the major thrusts of ASA Vision 2010 is to develop education outreach for a spectrum of learners from kindergarten students to senior citizens. Along these lines we explore a simple speed of sound measurement that can be accomplished by this broad range of learners. Understanding slow speed of sound compared to the speed of light allows the learner to comprehend the lack of temporal/optical correlation between sound and position of airplanes as they move across the sky. Other examples include the late arrival of sounds from aerial salutes during firework displays and sounds of lightning during storms. Synchronizing the acoustic echo of a sound pulse from an outdoor wall with the pulsed source for several duty cycles allows sound speed measurements. Analysis of the time of arrival with distance is presented for various wall sizes and air temperatures. The successes of the learning exercise and results of the measurements when accomplished by the K-gray audience will be presented (see, for example, http://arts.ucsc.edu/EMS/Music/tech_background/TE-01/soundSpeed.html).

3:05

4pED6. Visualization of harmonic starting process for flute sound experiment. Yoshinori Takahashi (Dept. of Comput. Sci., Kogakuin Univ., Tokyo, Japan, yoshinori@ieee.org), Mikio Tohyama (Waseda Univ., Tokyo, Japan), and Kazunori Miyoshi, Prof. (Kogakuin Univ., Tokyo, Japan)

Signal analysis based on spectral accumulation has great potential for monitoring the condition of structures using random vibration records excited by natural forces. This article describes cumulative harmonic analysis (CHA) by introducing a spectral accumulation function into Berman and Fincham's conventional cumulative analysis, thus exploring a potential new area in cumulative analysis. CHA effectively visualizes the feedback process from the flow of air against an edge. In this work, we visualized the harmonic starting process for understanding of physical experiment of musical instrument study especially the flute.

Session 4pMU

Musical Acoustics, Speech Communication, and Signal Processing in Acoustics: Statistical Approaches for Analysis of Music and Speech Audio Signals

Paris Smaragdis, Cochair
Adobe Systems, Inc., 275 Grove St., Newton, MA 02466

George Tzanetakis, Cochair
Univ. of Victoria, Dept. of Computer Sci., P.O. Box 3055, Victoria, British Columbia, V8W 3P6 Canada

Chair's Introduction—1:30

Invited Papers

1:35

4pMU1. Audio analysis using sparse representations. Mark D. Plumbley, Samer A. Abdallah, Maria G. Jafari, and Andrew Nesbit (Dept. of Electron. Eng., Queen Mary Univ. of London, Mile End Rd., London E1 4NS, UK, mark.plumbley@elec.qmul.ac.uk)

The method of "sparse representations," based on the idea that observations should be represented by only a few items chosen from a large number of possible items, has emerged recently as an interesting approach to the analysis of images and audio. New theoretical advances and practical algorithms mean that the sparse representations approach is becoming a potentially powerful signal processing and analysis method. Some of the key concepts in sparse representations will be introduced, including algorithms to find sparse representations of data. An overview of some applications of sparse representations in audio will be described, including for automatic music transcription and audio source separation, and pointers will be given for possible future directions in this area. [This work has been supported by grants and studentships from the UK Engineering and Physical Sciences Research Council.]

1:55

4pMU2. Modeling local stationarity in speech wave forms. Daniel Rudoy, Prabahan Basu, and Patrick J. Wolfe (Statistics and Information Sci. Lab., Harvard Univ., Oxford St., Cambridge, MA 02138)

Typical speech wave forms are well modeled as slowly time-varying, or so-called locally stationary, stochastic processes. This talk outlines recent work in detecting and modeling locally stationary speech time series, and describes a new method of adaptive short-time Fourier analysis and reconstruction based on local measures of time-frequency concentration. While adaptive analysis measures have previously been proposed in order to overcome the limitations of fixed-resolution schemes, the scheme presented here derives from quantifiable and rigorous notions of local stationarity. This yields demonstrable robustness properties for the case of noisy speech, as well as improved mean-square error estimation properties and other quality improvements relative to standard estimation procedures in which time-frequency resolution is fixed. [Work supported in part by DARPA and NSF.]

2:15

4pMU3. Toward automatic music transcription from audio input. Shigeki Sagayama, Hirokazu Kameoka, and Haruto Takeda (Grad. School of Info. Sci. Tech., Univ. of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan, sagayama@hil.t.u-tokyo.ac.jp)

Transcribing audio input to obtain a music score has been a basic but yet hard problem in music signal and information processing. It can be paralleled with continuous speech recognition comprising of acoustic analysis, acoustic model, language model, and decoder. This presentation discusses functional modules for the music case, i.e., estimation of multiple fundamental frequencies, rhythm models, and chord modeling. For multi-F_0 estimation, a computational auditory scene analysis-motivated approach is taken to model human hearing where each acoustic object is modeled with Gaussian mixture both along frequency and time axis [Kameoka (2005)]. An extended EM-algorithm is applied to iterative estimation of fundamental frequencies and onset/offset timings of music notes contained in the given spectrogram of the audio input. As for rhythm estimation, HMM is used to model the sequence of note lengths where tempo is treated in the model as a varying hidden variable [Takeda (2006)]. Chord progression is also modeled with HMM [Kawakami (2000)] where transition probabilities between chords and emission probabilities of notes from the hypothesized chord have been trained with a music database. Related issues are also discussed including other approaches to multiple fundamental frequencies with specmurt analysis [Sagayama (2005)], non-negative matrix factorization [Raczynski (2007)], and timber modeling [Miyamoto (2007)].

2:35

4pMU4. Prior structures for non-negative matrix factorization based audio source separation. Tuomas Virtanen (Dept. of Signal Processing., Tampere Univ. of Technol., P.O. Box 553, FI-33101 Tampere, Finland, tuomas.virtanen@tut.fi) and Ali Taylan Cemgil (Univ. of Cambridge, Trumpington St., CB2 1PZ Cambridge, UK, atc27@cam.ac.uk)

A generative signal model corresponding to non-negative matrix factorization algorithms is presented, and the model is extended by using priors for the matrices to be estimated. In the analysis of audio signals, this approach allows using prior information about the sounds, while allowing adaptation of the model parameters to the exact characteristics of the observed signal. Specifically, conjugate priors lead to fast inference algorithms. Methods for learning the parameters of the prior distributions from training material are described. The presented approach is evaluated on various sound source separation tasks, including standard data sets of music and speech.

Contributed Papers

2:55

4pMU5. Neighborhood indicator based on tonality and arrangements in musical organization: Automatic system for selecting similar excerpts based on musical information. Yuichiro Yamakaji (Graduate School of Sci. and Technol., Ryukoku Univ., 1-5 Yokoya, Oe-cho, Seta, Otsu-shi, Shiga 520-2194, Japan) and Masanobu Miura (Ryukoku Univ., Japan)

Recently, various methods for searching similar musical excerpts have been studied. However, it has not yet been realized that a method of automatically selecting similar excerpts based on features was obtained from musical information in consideration of a music theory. Therefore, this study was to realize the method. Concretely, employed parameters are "chord trigram," modeling sequence of chords based on the *n*-gram model, "emotional parameter," corresponding to emotional expressions such as tempi, modes, and average of duration for melody, and "note density," corresponding to the occurrence frequency of notes. Excerpts dealt with here are grouped into several categories based on proposed parameters by the CART algorithm, and degree of similarity is calculated between excerpts in each category. Developed here is a system for automatically selecting similar excerpts based on features obtained from musical information, named "NITAMO," which stands for neighborhood indicator based on tonality and arrangements in musical organization. The new system deals with more than 5400 excerpts and employs the XF format, a commercially sold format as an extended form of MIDI. We confirmed that the proposed system was useful by comparing similar musical excerpts provided by proposed system with those by human. [Work supported by the HRC, Ryukoku Univ.]

3:10—3:25 Break

Invited Papers

3:25

4pMU6. Inferring missing spectral data. Paris Smaragdis (Adobe Systems Inc., 275 Grove St., Newton, MA 02466) and Bhiksha Raj (MERL, Cambridge, MA 02139)

In this talk we will present a methodology that allows us to infer missing portions of spectrograms. We will present an approach that constructs models of sounds by observing either examples or the existing portions of a spectrogram with missing data. Once the model is learned we can use it to reconstruct missing areas of a spectrogram with inaudible artifacts. This process is very useful when trying to correct errors from spectral editing or when dealing with processes that corrupt a signal in the time/frequency domain. We will show how this approach is appropriate for polyphonic audio signals and that it significantly outperforms approaches using generic statistical models which are ill suited for spectral data.

3:45

4pMU7. Discovery of temporal patterns in continuous nonrandom sound sequences. Rita Singh and Bhiksha Raj (Human Lang. Technologies Ctr. of Excellence, Johns Hopkins Univ., Stieff Bldg./810 Wyman Park Dr., Baltimore, MD 21211)

The problem addressed is that of automatically determining the minimal structures in structured sounds such as human speech. It is well established that human speech comprises consistent sound units (such as phonemes), a fact that is exploited by speech recognition systems, which only model the units, characterizing all speech as sequences of units. Currently, the units are typically manually defined, and their statistical models are assigned structures based on subjective judgments. In this work it is attempted to identify these units automatically through an analysis of data. The problem is treated as one of entropy minimization. The minimum entropy estimation principle assumes a structured universe and dictates the estimation of the most predictable model that the observations used to train the model will allow. For the current problem, this amounts to identifying a set of units such that every word can be represented by a minimum-perplexity network over them. Experiments demonstrate that this procedure identifies units with consistent structures that, although not always identical to phonemes, are nevertheless able to explain the data just as well as quantified by recognition accuracy. More interestingly, the algorithm is generic and may be employed to learn component units of other natural sounds besides speech.

4:05

4pMU8. AudioDB: Scalable approximate nearest-neighbor search with automatic radius-bounded indexing. Michael Casey (Dept. of Music, Dartmouth College, HB 6242 Hallgarten Hall, Hanover, NH 03755)

This talk describes new approximate nearest-neighbor methods employed in a scalable audio-feature database system called "AudioDB." This open-source system is designed to scale to storing and searching hundreds of millions of feature vectors on standard UNIX workstation platforms. A radius-bounded nearest-neighbor vector-sequence search algorithm, based on locality sensitive hashing (LSH), achieves sublinear retrieval times at this scale. The performance of the LSH-based algorithm depends critically on the choice of radius bound supplied—the wrong value impacts retrieval accuracy or retrieval time. An optimal radius estimator is derived by modeling the minimum value distribution of a random sample of a data set's pairwise distance distribution. When used with LSH this yields

accurate search results with retrieval times several orders of magnitude faster than exhaustive search methods and space-partitioning methods. The same statistical sampling method is used to perform retrieval tasks at successively higher levels of specificity on labeled or unlabeled audio collections. The result is a system that (a) unifies audio retrieval tasks across a range of specificities, using the statistical framework of background distance-distribution sampling and hypothesis testing (b) is as accurate as exhaustive search methods and (c) is three orders of magnitude faster than exhaustive search methods.

4:25

4pMU9. Combining prior-knowledge and grouping cues using a spectral clustering approach. George Tzanetakis (Dept. of Comput. Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada, gtzan@cs.uvic.ca) and Luis Gustavo MartinsI (INESC Porto, Portugal)

Learning happens at the boundary interactions beween prior knowledge and incoming data. The same interplay takes place when trying to analyze and separate complex mixtures of sound sources such as music. Many approaches to this problem can be broadly categorized as either model based or grouping based. Although it is known that our perceptual system utilizes both of these types of processing, building such systems computationally has been challenging. As a result most existing systems either rely on prior source models or are solely based on grouping cues. In this work it is argued that formulating this integration problem as clustering based on similarities between time-frequency atoms provides an expressive but discipined approach to building sound source characterization and separation systems and to evaluating their performance. After describing the main components of such an architecture, we describe a concrete realization that is based on spectral clustering of a sinusoidal representation. We show how this approach can be used to model both traditional grouping cues such as frequency and amplitude continuity as well as other types of information and prior knowledge such as stereo panning, onsets, harmonicity, and timbre models for specific integration will also be described.

Contributed Paper

4:45

4pMU10. A theoretical frequency band limitation for analog recordings. Michael Zucker (ARL-Penn State Acoust., Appl. Sci. Bldg., North Atherton St., State College, PA 16802, mxz174@psu.edu)

Proponents of analog recording techniques have described their superiority over digitally sampled signals in part by referencing a theoretically infinite frequency band response. This is a typical argument used in favoring audiophile grade phonograph records to digitally sampled music. The actual frequency response of any individual transducer cartridge or phonograph record is determined by empirical measurements and is often bounded by electrical noise considerations as well as the physical condition of the record being played. These measured and predicted frequency responses, however, still hold the contention that there is theoretically infinite frequency band content available on perfect analog recordings and that the actual limit on frequency content is simply a result of an imperfect implementation of recording and playback. In this study an equation analogous to the Nyquist sampling theorem is derived using the geometry of a phonograph pickup system and simple calculus to obtain an absolute theoretical frequency limit for analog recordings based on the size of the pickup and the speed at which the time signal is written and read. This equation may prove useful in designing analog recording systems specific to a desired frequency band response.

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 4, 2:30 TO 4:15 P.M.

Session 4pNS

Noise and Architectural Acoustics: Sound Levels and Acoustical Characteristics of Modular Classrooms

Paul D. Schomer, Chair
Schomer and Associates, 2117 Robert Dr., Champaign, IL 61821

Chair's Introduction—2:30

Invited Papers

2:35

4pNS1. Sound field amplification competes with noise control. David Lubman (DL Acoust., 14301 Middletown Ln., Westminster, CA 92683-4514)

Sound field amplifiers continue to be aggressively promoted for mainstream classrooms. Studies showing improved academic achievement with sound field systems mislead by not revealing that the subject classrooms were excessively noisy. Sound field is a band-aid solution for noisy classrooms. Surely, quiet classrooms and natural, unstrained voices improve academic achievement as well as or better than amplifiers. Soundfield is being promoted to school building officials as low cost substitutes for noise control. Marketers mislead school officials by suggesting that classrooms cannot be quieted affordably and that occupied classrooms are so noisy that amplification is indispensable. One might ask how students were classroom educated prior to the invention of electronic amplifiers! In a recent "Trojan horse" strategy, sound field systems are being purchased for classroom multimedia. Sound field reinforcement then becomes a "freebie." The problem is that all amplified systems work best in quiet rooms. Moreover, though soundfield amplification works well in special education classrooms for hearing impaired students. It is seldom needed in quiet mainstream classrooms. Moreover hearing impaired students in mainstream classrooms are better served by FM systems. Classrooms that depend on amplification are unavailable when the system becomes inoperative due to malfunction, missing parts, or power failures. Quiet classrooms are robust.

2:55

4pNS2. Progress update from S12 working group 46: Relocatable classrooms. Tom Hardiman (944 Glenwood Station, Ln. Ste. 204, Charlottesville, VA 22901, tom@modular.org)

The purpose of forming working group 46 was to develop an amendment to ANSI S12.60-2002 that specifically focuses on the challenges and unique circumstances surrounding the design, construction, use, and relocation of relocatable classrooms. S12/WG46 is currently drafting an addendum to S12.60-2002 Guideline for Classroom Acoustics. This presentation will provide an update on the progress being made by S12 working group 46.

3:15

4pNS3. Measurement of the ambient in relocateable modular classrooms. Paul Schomer, Wilbur Chang (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821), and Irvin Derks (Bard Manufacturing Co., Bryan, OH 43506-0607)

The ambient in a classroom consists of noises internal to the classroom (e.g., HVAC, lighting, IT equipment, refrigerators, and fish tank pumps) and noises external to the classroom (e.g., street traffic, aircraft, children in the playground, and adjacent classroom noise). Currently, ANSI S12.60, Classroom Acoustics, requires that the A-weighted ambient level, unoccupied, be less than 35 dB at the noisiest position in the classroom. A basic reasonable procedure to implement the required measurements has been developed and tested. The authors find that the procedures developed can be readily used by inexperience but technically oriented individuals. The procedure separates the measurement of internal and external noises into two measurements, typically at different positions. Because the two noises are variable, unrelated in position, and uncorrelated in time, it does not make sense to combine them in any fashion. Rather, in the procedure to be recommended by WG 68, the two measurements are to be reported and evaluated, separately, against the 35 dB criterion. Other notable changes are expansion of the measurement hours to include busy traffic periods (high environmental noise periods) prior to and after school and on Saturdays, and implementation of HVAC duty cycles to better estimate a typical A-weighted hourly equivalent level from HVAC equipment (the subject of a companion paper).

3:35

4pNS4. Outdoor to indoor A-weighted sound level reduction of typical modular classrooms and assessment of potential performance improvements based on the outdoor-indoor transmission class spectrum. Noral Stewart (Stewart Acoust. Consultants, 7406 L. Chapel Hill Rd., Raleigh, NC 27607, asamiami@sacnc.com)

Several different designs for modular classrooms from various parts of the United States were examined for their expected A-weighted outdoor to indoor sound level reduction based on the sound spectrum used for the ASTM E1332 Outdoor Indoor Transmission Class (OITC). Analysis was simplified by using an average absorption effect and the OITC rating of walls and roofs as estimated using the INSUL computer program. Results showed a significant variation among designs with the better designs providing a good foundation for further improvement.

3:55

4pNS5. Exploration of alternative means to evaluate exterior noise loss. Kenneth Good (Acoust. Privacy Enterprises, LLC, P.O. Box 252, Mount Joy, PA 17552)

The outdoor-indoor transmission class measurement can be problematic to execute due to many environmental and logistical issues. This leads to the following question: Can the measurement be made from the inside to the outside? This presentation will explore a technique to evaluate a building's ability to block exterior sound by placing the source signal inside and measurement device on outside.

Session 4pSAa

Structural Acoustics and Vibration: Aeroacoustic and Hydrodynamic Interactions with Structural Acoustics and Vibration

Jerry H. Ginsberg, Chair
Georgia Inst. of Tech., Sch. of Mechanical Eng., Atlanta, GA 30332-0405

Invited Papers

1:30

4pSAa1. Comparison of time-integration schemes for fluid-structure interaction. Aldo A. Ferri, Mohammed Kapacee, Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Tech., Atlanta, GA 30332-0405, al.ferri@me.gatech.edu), and Marilyn Smith (Georgia Inst. of Technol., Atlanta, GA 30332-0150)

Computational structural-dynamics codes invariably use finite element (FE) methods and relatively course meshes, whereas finite-difference methods with fine meshes are popular for modeling the fluid domain. Fluid-structure interaction (FSI) problems in which there is a mean flow feature can be addressed by simultaneously employing both techniques. Because of the dissimilarities of the two formulations, a time-domain solution is most readily obtained by allowing each code to march forward in time in a loosely coupled manner. A proper FSI implementation must address the physical and computational issues associated with compatibility of displacements and surface tractions at the boundary of the two domains. It is also necessary that one identifies a suitable numerical scheme to perform unsteady time-marching simulations of the coupled system. It is the latter issue that concerns this paper. Various techniques for accurate and stable time integration of loosely coupled systems are compared. A two-dimensional example is studied that consists of an inviscid compressible fluid and a thin elastic beam/plate. Small amplitude motion is assumed for the plate, so linear approximations are used in the FE model. The fluid domain, on the other hand, is modeled using the nonlinear Euler equations. [Work supported by NASA Contract No. NAS1-02117, Task Order 6101-GT.]

1:55

4pSAa2. Flight systems aeroelastic-acoustics simulation. Kajal K. Gupta (NASA Dryden Flight Res. Ctr., Edwards, CA 93523), Sangbum Choi (CSULA, Los Angeles, CA 90032), and Adem Ibrahim (Norfolk State Univ., Norfolk, VA 23504)

Many practical problems, such as flight vehicles, are often characterized by complex interactions among a number of primary disciplines as structures, fluids, controls, and propulsion, among others. In some critical flight regimes as transonic flow, the dynamics behavior of fluids tend to be highly complex, needing computational fluid dynamics (CFD)-based modeling, rather than linear aerodynamic methods, for accurate prediction of unsteady flow. Such unsteady pressure values, generated by the interaction of fluids and elastic structures, could then be conveniently used to compute acoustic wave frequencies and sound levels. For accurate simulation of complex engineering problems such as advanced aircraft, it is necessary to employ unstructured grids to model the fluid discipline, such being mostly the case for structure modeling. Effective CFD-based aeroelastic modeling, using finite element discretization of both fluids and structures, employing unstructured grids, have been efficiently utilized for simulation of Hyper-X and F-18 aircraft. In connection with a current ongoing project Stratospheric Observatory For Infrared Astronomy, evaluation of acoustic activities in the cavity of the modified Boeing 747SP aircraft, housing the telescope, became a crucial issue. The multidisciplinary code STARS7 was recently extended to include the acoustic effects.

Contributed Papers

2:20

4pSAa3. Turbulent boundary layer shear stress transmitted through a viscoelastic layer. E. Capone and William K. Bonness (Appl. Res. Lab., Pennsylvania State Univ., P.O. Box 30, State College, PA 16804)

Transfer functions are developed for the transmission of unsteady shear stress, generated by a turbulent boundary layer in water, through a viscoelastic layer backed by a rigid plate. Existing analytical models are used to estimate the unsteady wall pressure and shear stress from 10–1000 Hz for a flat plate boundary layer with zero pressure gradient. A new model is developed for the transmission of unsteady shear stress through the viscoelastic layer. The model is used to predict the unsteady pressure fluctuations, or flow noise (due to the unsteady shear stress), which would be seen by a finite size sensor embedded under the elastomer layer. The calculated unsteady pressure and shear stress levels are in good agreement with recent experimental measurements. The unsteady shear stress transfer functions are found to have a peak at the acoustic wavenumber.

2:35

4pSAa4. Removing unwanted signals from wall pressure and vibration measurements on a structure subjected to turbulent boundary layer excitation. William Bonness, David Jenkins, and Dean Capone (Appl. Res. Lab, Penn State Univ., State College, PA 16804)

Fluid-structure interaction experiments typically involve measurements of the excitation force (or pressure) and the corresponding vibration response to the excitation. In addition to desired flow information, measured turbulent boundary layer wall pressure data often include unwanted signals such as acoustic pressures and vibration induced pressures. Measured vibration data on a structure can also include unwanted electrical noise and vibration energy from adjacent structures. A noise removal technique is presented, which allows one to remove an unlimited number of unwanted correlated signals from a set of measured data. In its simplest form, this technique is related to the coherent output power. However, the more general technique provides an ability to remove multiple signals and to retain complex values (magnitude and phase). These advantages can yield signifi-

cantly greater information of the flow field and structure under investigation. This technique is demonstrated using measurements from an aluminum cylinder internally filled with water flowing at 20 ft/s.

2:50

4pSAa5. A theoretical model to simulate two-phase vibroacoustical frequency and damping effects in a pipe. Vincent Debut and Jose Antunes (Appl. Dynamic Lab., Inst. of Nuclear Technol., Sacavem, Portugal)

In order to avoid excessive flow-induced vibrations in industrial components operating in two-phase flow, the analysis and understanding of energy dissipation mechanisms are of prime importance. Several experimental studies since the classic work [L.N. Carlucci, "Damping and hydrodynamic mass of a cylinder in simulated two-phase flow," J. Mech. Des. **102**, 597–602 (1980)] revealed the strong dependence of two-phase damping on the characteristics of the two-phase flow, particularly the void fraction and fluid used. All these investigations agreed in the complexity of the dissipation processes involved. To tackle such an intricate problem, we consider in this paper the fluid-structure dynamics of a pipe filled with a bubbly liquid which interacts with two single-degree of freedom piston terminations. Several formulations for the coupled problem are stated and compared, the two-phase acoustics being based on a homogeneous mixture formulation [L. van Wijngaarden, "One-dimensional of liquids containing small gas bubbles," Annu. Rev. Fluid. Mech. **4**, 369–396 (1972)]. Then, extensive numerical computations are performed. Plots of the evolution of the modal frequencies and damping as a function of the mixture void-fraction are obtained, and the corresponding vibroacoustical transfer functions and numerical time-domain simulations are presented.

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 1, 3:15 TO 5:30 P.M.

Session 4pSAb

Structural Acoustics and Vibration: Vibration—Measurements and Analysis

Joseph M. Cuschieri, Cochair
Lockheed Martin Corp., 100 E. 17th St., Riviera Beach, FL 33404

Sean F. Wu, Cochair
Wayne State Univ.,. Dept. of Mechanical Eng., 5050 Anthony Wayne Dr., Detroit, MI 48202

Contributed Papers

3:15

4pSAb1. Performance comparison of viscoelastic materials typically used for acoustical damping of aircraft fuselage structures. Esen Cintosun, Noureddine Atalla (Groupe d'Acoustique de l'Univ. de Sherbrooke, Dept. of Mech. Eng., Univ. de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada), Tatjana Stecenko (MTI Polyfab Inc., Mississauga, ON L5T 2A4, Canada), and Maxime Bolduc (Univ. de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada)

Vibration tests were performed to compare typical viscoelastic damping materials attached to two distinct representative aircraft fuselage structures. The fuselage structures were a 19×20-in^2 Al ribbed panel and a 20×21-in^2 carbon composite ribbed panel. A shaker was used to excite each panel with and without viscoelastic material treatment. Each test structure along with the shaker was situated in an environmental chamber to collect measurements at low temperatures of -20, -30, and -40 °C. A laser vibrometer was used to collect velocity measurements at 15–26 random locations on the surface of the panels. The parameters that were compared included input mobility, damping loss factor (DLF), and space averaged squared velocity. The DLF values that were obtained using the decay rate, power input, and half-power bandwidth methods were presented as averages in 1/3 octave frequency bandwidths (from 100 to 2500 Hz). The space and frequency averaged squared velocity was used to quantitatively categorize the viscoelastic damping materials. The effect of viscoelastic damping material coverage was also evaluated by comparing measurements collected at 50% and 80% coverages. The results of this experimental study are being utilized in finite element modeling, optimization of viscoelastic material treatment, and development of alternatives to viscoelastic damping. [Work supported by the MTI Polyfab Inc.]

3:30

4pSAb2. Coupling of extensional and flexural displacement fields in straight bars. Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, jerry.ginsberg@me.gatech.edu)

Extensional and flexural displacements of straight elastic bars are conventionally treated as independent superposable effects. However, coupling of displacement fields due to boundary effects is an important feature in many realistic configurations. The present paper examines the vibration modes of a beam in which the aforementioned coupling is caused by a roller support that is mounted on an incline. A closed form solution of the coupled dislacement modes is derived. The transcendental characteristic equation is shown to depend solely on the incline angle for the end support, the reduced frequency formed from the bar wave speed, and the ratio of the cross-section's radius of gyration to the span length. The characteristic equation is solved numerically. The resulting eigenvalues are used to evaluate the coupled mode shapes, which leads to a quantitative assessment of coupling effects relative to the limiting cases where the incline is horizontal or vertical, for which the displacement fields truly are uncoupled. An interesting aspect of the analysis is its mathematical similarity to the application of Timoshenko theory to bars.

3:45

4pSAb3. Harmonic control of vibration and sound radiation of a plate using a virtual impedance approach. Nicolas Quaegebeur, Philippe Micheau, and Alain Berry (GAUS, Dept. de Gnie, Univ. de Sherbrooke, 2500 Blvd. de l'Universit, Sherbrooke, QC, Canada J1K 2R1

The present work focuses on the harmonic control of plate vibration and radiation using distributed collocated piezoceramic units. The objective is to add virtual impedance units on the plate and to optimize its complex value

for each frequency. In active damping (velocity feedback controllers), the added virtual impedance is equivalent to a resistive effect. In the present study, the added virtual impedance is chosen arbitrarily in the complex plane. Each added impedance unit is composed of a collocated PZT actuator and a PVDF sensor. This approach allows obtaining dual variables for harmonic control. In order to tune the PZT voltage at each frequency, the virtual impedance approach is implemented with a harmonic adaptive control. The stability is always ensured with a centralized processing, and for certain frequencies, a decentralized processing can be used. Preliminary theoretical study is carried out to assess the behavior of the smart panel for different added virtual impedances for a mechanical excitation by a shaker. Those numerical results are compared to experimental measurements performed in an anechoic chamber. Our proposed strategy compare favorably to common strategy of active damping to reduce harmonic sound radiation. Hence, a reduction of 20 dB can be achieved at certain frequencies.

4:00

4pSAb4. Frequency spectra of bilaminar prolate spheroidal shells. Sabih I. Hayek (Dept. of Eng. Sci. and Mech., Penn State Univ., Univ. Park, PA 16802) and Jeffrey E. Boisvert (NAVSEA Newport, Newport, RI 02841)

The kinetic and strain energy densities were derived for the vibration of a bilaminar prolate spheroidal shell of constant thickness. The bilaminar shell is composed of two bonded concentric prolate spheroidal layers of different material properties and thicknesses. The elastic strain energy density has seven independent kinematic variables: three displacements, two thickness shears, and two thickness stretches. Continuity of displacements is enforced at the interfacial reference surface. The shell has constant thickness $h=h1+h2$, where $h1$ and $h2$ denote the thicknesses of the respective layers. The reference surface eccentricity is defined by $1/a$, where a is its shape parameter. Using appropriate comparison functions in terms of Legendre polynomials that satisfy the boundary conditions for a closed spheroidal shell, the system is solved using the Galerkin method. Numerical results of the frequency spectra are presented for various ratios of $h1/h2$ and various material properties for the two layers. Initial results are presented for $a=100$ (a nearly spherical shape). [Work supported by the ONR/ASEE Summer Faculty Research Program.]

4:15

4pSAb5. Noncontact surface wave testing of pavements using microphones. Nils Ryden (Faculty of Eng., Eng. Geology, Lund Univ., Box 118, SE-22100 Lund, Sweden, nils.ryden@tg.lth.se), Michael J. S. Lowe, and Peter Cawley (Imperial College, London SW7 2AZ, UK)

Pavements are constructed using several layers of materials, and their durability depends mainly on the stiffness modulus and thickness of these strata. Surface wave testing is an effective tool to measure the stiffness and thickness of pavement layers. However, measurements are still based on spot testing with fixed receivers and source. The typical large size of pavements and the cost of closing down roads to make stationary testing makes these measurements impractical. We present experiments where a multichannel array of microphones and an automatic source are attached on a small trolley so that measurements can be taken almost continuously while moving. Measurements on asphalt or concrete pavement layers are based on supersonic leaky air-coupled surface waves. We also demonstrate that the same approach can be applied to softer granular pavement layers utilizing the seismic-to-acoustic coupling in poroelastic materials. Results show that microphones can be successfully used to measure correct surface wave dispersion curves even while moving along the surface. This opens up the possibility for faster on-the-fly surface wave testing of pavement layers since surface contact is no longer required. The theoretical background along with experimental results of the application to nondestructive testing of pavements will be presented.

4:30

4pSAb6. Determining source location from vibration "slowness" vectors. Byung-ghun Kim, Jangbom Chai (Dept. of Mech. Eng., Univ. of Ajou, Suwon 443749, Korea, z3hif@naver.com), and Richard H. Lyon (RH Lyon Corp., Belmont, MA 02478-2021)

The time for a wave to transit from one sensor to another is related to its "slowness" and is measured by the phase delay of the cross spectrum of the signals from the sensors. If the sensor separation is much less than a wavelength, this cross spectrum is related to the energy flow in the wave. An "L-shaped" array of three sensors allows one to determine the two vector components of slowness and the direction of the wave. If wave directions are computed at two or more such L-shaped arrays, then it may be possible to "triangulate" and fix a source location. It is possible of course that the directions may be such that there is no solution for a source location. Generally, the wave directions at the arrays are determined for a fixed set of frequencies. Then, at each frequency, a source solution is sought. At some frequencies, there may be no solution so the set of frequencies is diminished. Also, it is possible, even likely, that waves at different frequencies or wavelengths may have taken different paths from the source to the arrays. This paper reviews the results of applying the slowness vector computations to data taken on a model nuclear plant flow loop.

4:45

4pSAb7. Diffuse field interferometry for experimental Green's function estimation and damage detection. Adelaide Duroux, Karim Sabra (School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW., Atlanta, GA 30332-0405), James Ayers, and Massimo Ruzzene (Georgia Inst. of Technol., Atlanta, GA)

Structural health monitoring systems often rely on propagating elastic waves through complex structures, which can result in the formation of diffuse-fields over a long reverberation time. Recent theoretical and experimental studies have demonstrated that Green's functions between a pair of monitored points can be extracted from cross-correlation of the recorded diffuse-fields (e.g., scattered fields or ambient noise). Knowledge of the Green's functions between large numbers of points can be used to successfully identify and localize damage in complex structural components. In this work, Green's functions are first estimated experimentally from full-field measurements obtained with a scanning laser vibrometer. This provides the wealth of *a priori* information necessary to detect and localize "secondary" sources, such as damages, when only a limited number of sensors are actually mounted on the structure. The proposed approach relies on the detailed knowledge of the structural response, which is exclusively obtained through experimental measurements performed on the actual component under consideration.

5:00

4pSAb8. Flaw localization in a structure using model-based backpropagation. David Chambers, Sean Lehman, Lisle Hagler, Henry Hsieh, and Karl Fisher (Lawrence Livermore Natl. Lab., P.O. Box 808, L-154 Livermore, CA 94551, chambers2@llnl.gov)

Damage localization is an important part of structural health monitoring. In this talk we present the results of a method of localizing pointlike damage in a structure using changes in the vibrational response induced by the damage. The vibrational response is measured at discrete locations on a structure both before and after damage is induced. The difference in response is used to drive a numerical model of the undamaged structure (backpropagation). The damage location is marked by a peak in the calculated displacement field. Results are shown using laboratory measurements of a cylinder and two nested spherical half-shells. Numerical simulation is used to demonstrate performance in more complex structures. [Work supported by the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.]

5:15

4pSAb9. Singing sands, musical grains, and booming sand dunes. Tom Patitsas (Dept. of Phys., Laurentian Univ., 935 Ramsey Lake Rd., Sudbury, ON P3E 2C6, Canada)

The origin of the seismic and acoustic emissions from a bed of singing sands or musical grains, impacted by a pestle, is sought in a boundary layer, several millimeters thick, at the leading front end of the pestle. It is assumed that such a layer is the result of the fluidization of the grain asperities due to the high stress level at the pestle front end. Such a fluidization results in a very low modulus of rigidity and in a shear phase velocity about 1 m/s in the boundary layer. The frequencies of the shear modes of vibration, in such a layer, are compared with those determined experimentally. The frequencies depend weakly on the geometry of the leading pestle front. The same concept of the boundary layer can account for the emissions from plates of sand sliding on a dune surface and from grains shaken in a glass jar.

Session 4pSC

Speech Communication: Articulatory Measures and Signal Processing (Poster Session)

Stefan Frisch, Chair
Univ. of South Florida, Communication Sci. and Disorders, 4202 E. Fowler Ave., Tampa, FL 33620-8100

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbers papers will be at their posters from 3:00 p.m. to 4:30 p.m.

4pSC1. Speech enhancement for noise-robust speech recognition. Vikramjit Mitra and Carol Espy-Wilson (Inst. for Systems Res., Dept. of Elec. Comput. Eng., Univ. of Maryland, College Park, MD 20742, vmitra @umd.edu)

A modified phase opponency (MPO) model was combined with the aperiodic/periodic/pitch detector (APP) by Deshmukh *et al.* [JASA, **121**, 3886–3898 (2006)] for speech enhancement. The MPO-APP passes the speech-dominated regions as is but greatly attenuates the noise-dominated regions. Thus, the noise in nonspeech regions and in the regions between the formants are significantly reduced. Unlike many other techniques, the MPO-APP does not introduce musical noise. However, given that the noise in formant regions is passed along with the speech signal, the enhanced signal sometimes has a shadow effect. To get rid of this shadow effect, we recently developed a preprocessor for the MPO-APP. First, the signal to noise ratio of the signal is estimated. Second, we use an adaptive generalized spectral subtraction technique to reduce the noise in the speech dominant regions. The results show a significant improvement in the quality of the enhanced signal and, in addition, increased accuracy in automatic recognition of the enhanced signals over that obtained from the original signals or the signals enhanced without the preprocessor. [Research supported by NSF Grant IIS0703859.]

4pSC2. Variable resolution spectral/temporal features for automatic speech recognition. Stephen Zahorian and James Wu (Dept. of Elec. and Comput. Eng., Binghamton Univ., Binghamton, NY 13903, zahorian @binghamton.edu)

Many studies from speech science have shown that the mel frequency scale more closely matches speech perception than the linear frequency scale. Automatic speech recognition engineers have empirically demonstrated that the use of the mel scale results in more accurate speech recognition than that obtainable with features computed with respect to a linear frequency scale. The features most typically used for automatic speech recognition are mel frequency cepstral coefficients (MFCCs), along with delta and acceleration terms that represent the temporal evolution of MFCC over very short time intervals. However, the MFCC features do not encode the better temporal resolution that is possible at higher frequencies with low-frequency resolution. In this paper, a two-dimensional feature set is presented that incorporates good frequency resolution and low-time resolution at low-frequencies, and low frequency resolution and good time resolution at high frequencies. These features are computed from overlapping blocks with an effective length of approximately 100 ms at low frequencies and approximately 20 ms at high frequencies. Experimental results are given for phonetic recognition using the TIMIT database. The implications are that features that encode the temporal evolution of speech spectra are important for automatic speech recognition.

4pSC3. Modeling prosodic rhythm: Evidence from second language speech. Emily Nava, Louis Goldstein (Linguist. Dept., Univ. of Southern California, GFS 301, Los Angeles, CA 90089, nava@haskins.yale.edu), Hosung Nam, Michael Proctor, and Elliot Saltzman (Haskins Labs., New Haven, CT 06511)

The global prosodic structure of languages has been described using the typological dichotomy of stress-timed versus syllable-timed. Various indices have been successfully employed in literature for quantifying these classifications, one of which is the duration ratio between the total voiceless and total voiced stretches in the signal. It has been further shown that various language-specific characteristics, such as syllable-structure phonotactics and stress-sensitive lengthening and shortening, can contribute to this difference. To reveal the interaction of these components, acoustic data from running speech of L1 Spanish/L2 English and native English speakers were analyzed. Total ratio of voiceless-to-voiced durations discriminated L1 Spanish (lower) and L1 English (higher); L2 speakers showed ratios in between the two, with higher proficiency L2 speakers showing ratios closer to L1 English. A task dynamic application, a speech planning and production model [Nam et al., J. Acoust. Soc. Am. **115**, 2430 (2004)] was used to model the performance of the L1 and L2 speech of L2 speakers, allowing determination of which contributors to global rhythm are more readily acquired by L2 speakers. [Work supported by NIH and NSF.]

4pSC4. Comparison of a new frequency domain periodicity detector derived from a temporally stable power spectral representation and zero frequency filtering based periodicity detector. Hideki Kawahara, Hanae Itagaki, Ryuichi Nisimura, and Toshio Irino (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama 640-8510, Japan, kawahara @sys.wakayama-u.ac.jp)

Recently, an event based periodicity detection method was proposed using zero frequency filtering method [Yegnanarayana *et al.*, ITRW Aalborg 2008]. The authors also introduced a new periodicity detector based on TANDEM-STRAIGHT [Kawahara *et al.*, ICASSP2008 (2008)], a combination of temporally stable power spectrum estimation method for periodic signals and a spectral envelope recovery method based on consistent sampling [Unser, Proc. IEEE (2000)]. These methods provide a complementary set of information on excitation source signals of speech sounds. MATLAB implementation of these methods was evaluated using publicly available EGG databases and compared with popular existing methods. It was found that the zero frequency filtering based method runs eight times faster than real time even with MATLAB implementation and yielded comparable performance to popular methods. The TANDEM-STRAIGHT based methods also yielded comparable performance and additional information that is useful to represent diplophonia, for example. Detailed analysis examples and their application for TANDEM-STRAIGHT will also be discussed.

4pSC5. Vowel and speaker classification based on multivowel linear discriminant classfication. Al Yonovitz (Dept. of Communicative Sci. and Disord., Univ. of Montana, Missoula, MT 59812, al.yonovitz@umontana.edu), Ryan Anderson (Univ. of Montana, Missoula, MT 59812), and Sarah Van Orden (Univ. of Montana, Missoula, MT 59812)

Accurate and automated voice or speaker identification has been a major goal for those involved in forensic issues. In addition, voice and speaker identification has many applications in security and business. Numerous previous efforts to derive features for speaker classification have failed to achieve a sufficiently low-error rate. In this study a linear discriminant analysis was independently performed for vowel formants (*F*1, *F*2, *F*3) for

each of ten vowels. The standardized canonical discriminant coefficients (SCDCs) weight each of the formants toward the group separation. These SCDC values are then linearly combined to form a single scalar for each vowel. An analysis that considered the multivariate data vector composed of the ten vowel scalars was then used for classification. All analysis was accomplished with the classic archived data set of Peterson and Barney (1952) and an additional data set derived for this study. The classification was over 98% accurate in separating males (M), females (F), and children (C) into subgroups, greater than any other study has provided. The methodology was used in a manner which tested specific speaker identification.

4pSC6. Reconstruction of missing formants based on spectral power series expansion. Kesaaki Minemura (Dept. of Comput. Sci., Waseda Univ., 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan, minemura@shirai.cs.waseda.ac.jp), Satoru Goto, and Mikio Tohyama (Waseda Univ., Saitama 367-0035, Japan)

In this paper, we investigated spectrum estimation for vowels using power series expansion. In general, sound signals can not keep their characteristics when the spectral peaks and dips are missing. Consequently, the missing spectral peaks and dips are estimated by using power series expansion. First, vowels are divided into cycle by cycle. Second, the one-cycle vowels are converted into frequency by FFT. Third, damaged spectra were made for demonstrating spectrum estimation. Fourth, we estimated missing spectral peaks and dips using power series expansion. In our spectrum estimation, the differential coefficients for power series expansion were calculated from nearby spectrum of missing spectrum. In this experiment, the estimated spectrum was almost the same compared with the original spectrum in the case of approximation by power series expansion on the rank of 6. The spectrum is restored about the range of 180 Hz using nearby 10-Hz samples. Consequently, power series expansion can estimate spectral peaks and dips for recovering spectrum.

4pSC7. Multipitch tracking and speaker separation for two or more speakers. Srikanth Vishnubhotla and Carol Espy-Wilson (Dept. of Elec. Eng., Inst. for Systems Res., A V Williams Bldng., Univ. of Maryland, College Park, MD 20742, srikanth@umd.edu)

Accurate robust pitch tracking in multispeaker environments is an important issue in monaural speech separation. This research presents a spectrotemporal multipitch algorithm designed to detect voiced unvoiced regions in a mixture of multiple speakers, to identify the number of speakers in voiced regions, and to yield the pitch estimates of voiced speakers. Pitch estimation is based on identifying the minima of the multidimensional (MD) AMDF. The number of voiced speakers is determined by the temporal evolution of the MD AMDF. The algorithm has been developed and tested for the two-speaker case and is being extended to the three-speaker case. Evaluation on a framewise basis for the two-speaker case yielded an insertion error rate of 1.78%, a deletion error rate of 14.37%, and a substitution error rate of 3.26%, comparable to the state of the art. Pitch estimation errors resulted primarily from speaker domination. The extension to the three-speaker case shows promising preliminary results. The proposed algorithm can contribute to speaker separation by helping to identify spectrotemporal regions dominated by one of the speakers from those that have comparable contribution from all speakers. Appropriate strategies can then be used to separate speech in these different scenarios. [Research supported by NSF Grant BCS-0519256].

4pSC8. Experimental and numerical determination of the surface deformation of a synthetic model of the human vocal folds. Li-Jen Chen and Luc Mongeau (Dept. of Mech. Eng., McGill Univ., McDonald Eng. Bldg., 817 Sherbrooke St. West, Montreal, QC H3A 2K6, Canada, ljchen @purdue.edu)

A model of human vocal folds was fabricated using a three-component liquid platinum-catalyzed silicone solution. The size, idealized shape, and mechanical properties of the homogeneous synthetic model were selected based on the available data. The superior surface displacement of the synthetic model during self-oscillations was measured using the digital image correlation technique. A finite element model of the synthetic model was created to calculate the state of the deformable solid. Modal testing of the synthetic model was performed to establish the material properties and to verify boundary conditions in the simulation. The self-oscillation of the synthetic model was simulated by imposing a sinusoidal pressure loading over model surfaces, with frequency and amplitude determined from the direct measurement. From the simulation, the von Mises stress over the inferior surface was found to be around 2.2 kPa during the maximum orifice opening, which is around twice of that over the superior surface. So far, only the superior surface deformation data have been available because of technical limitations in clinical studies. The current study may provide additional information, such as the maximum amplitude and location of the peak stress, which is useful for diagnostic and treatment purposes. [Work supported by NIH.]

4pSC9. Recent improvements to the University of California, Los Angeles' voice synthesizer. Norma Antonanzas-Barroso, Jody Kreiman, and Bruce R. Gerratt (Div. Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, nab@ucla.edu)

A number of enhancements have recently been added to the UCLA voice synthesizer. Additions have been made to add functionality and to address several theoretical issues that arose during development and application. New functions include the ability to manipulate the source spectrum directly by changing the slope and/or amplitude of a user-defined group of harmonics or by manipulating individual harmonics directly. The synthesizer also allows users to add zeros to the vocal tract transfer function, improving the modeling of many pathological voices. A number of other enhancements improve the ability to create and systematically vary stimuli for perceptual experiments. Theoretical development has focused on the importance of pitch-synchronous Fourier analysis in modeling the voice source, particularly with respect to measuring and manipulating the amplitudes of $H1$ and $H2$. Issues surrounding spectral effects of upsampling, downsampling, and pulse stretching, which are needed for precise manipulation of $F0$, will also be discussed. This synthesizer is available with documentation as open source shareware at www.surgery.medsch.ucla.edu/glottalaffairs/, and copies will be available at the conference. [Work supported by NIH/NIDCD Grant No. DC01797.]

4pSC10. A survey of respiratory system behavior during pauses in spontaneous speech. Janet Slifka (MGH Voice Ctr., 11th Fl., One Bowdoin Square, Boston, MA 02114)

This work is part of an ongoing study to characterize respiratory system involvement during the generation of pauses in connected speech. At pauses, the speaker takes a breath or the speaker does not take a breath. Previously reported results for read speech observed that without-breath pauses were generated with a sharp movement toward net inspiratory effort followed by a sharp return toward increased net expiratory effort [J. Slifka, *In Dynamics of Speech Production and Perception* (IOS, 2006), pp. 45-58]. In the present work, 52 utterances of spontaneous speech from four speakers were analyzed. Behaviors in spontaneous speech include not only the activity similar to that observed for read utterances but also additional types of behaviors. Some without-breath pauses were observed to have regions of oscillation in net effort. Such behavior appears to be more common during longer pauses and during hesitations. Secondly, some without-breath pauses were generated with a reversed pattern—sharp movement toward increased net expiratory effort and back to lesser net expiratory effort. This pattern was much less common and appears to be associated with relatively short pauses in regions of hesitations. These results characterize the respiratory system as an active component in speech production. [Work supported by NIH-NIDCD DC007986]

4pSC11. Modulation of middle and long-latency cortical potentials during pitch-shifted auditory feedback perturbation. Roozbeh Behroozmand, Hanjun Liu, Laura Karvelis, and Charles R. Larson (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208)

It has been shown that applying perturbation to the real-time voice auditory feedback evokes compensatory responses that minimize the vocal output error. Auditory feedback-based control of the voice fundamental frequency involves the integration of multiple neural substrates that take part in the process of vocal output error detection and correction. Previous studies have shown that the cortical brain potentials are modulated during active vocalization versus passive listening to the self-generated voice in the absence of auditory feedback perturbation. To learn more about these central

mechanisms, the present study tested the effect of vocalization condition across different pitch-shifted auditory feedback magnitudes. Twenty English speaking young adults vocalized a vowel sound and received pitch-shifted feedback of +100, 200, or 500 cents while scalp potentials were recorded. Results show that the middle-latency P50 responses were significantly larger during the active vocalization task only at 200 cents frequency shifts. Long-latency N100 and P200 peaks were significantly larger during vocalization for 100 cents and 100 cents/200 cents, respectively. The overall conclusion is that the middle- and long-latency cortical responses are significantly greater during vocalization, but the difference between response peak magnitudes across vocalization and passive listening conditions diminishes with larger pitch frequency shifts in the voice auditory feedback.

4pSC12. Hemisphere lateralization underlying auditory feedback control of voice *F*0. Hanjun Liu, Michelle Meshman, and Charles Larson (Dept. Commun Sci. Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL, 60208)

Reflexive compensation for unexpected pitch changes in auditory feedback for stabilization of voice fundamental frequency (*F*0) has been demonstrated using the frequency perturbation technique. However, neural mechanisms of the pitch-shift reflex are still unclear. This study investigated hemisphere lateralization in the auditory processing of pitch changes by using event-related potentials. During vocalization of sustained vowels, subject's voice pitch feedback was randomly shifted upward or downward 100, 200, or 500 cents with 200 ms duration. Simultaneously, brain activity was recorded through surface electrodes located on the scalp bilaterally referenced to linked earlobes. The results showed larger brain potentials (peak difference between *P*200 and *N*100) in the *C*3 (left) than in the *C*4 (right) site, indicating left hemisphere dominance in the central motor area, while right hemisphere dominance was found in the lateral frontal area with larger potentials in the *F*8 (right) than in the *F*7 (left) site. Furthermore, these patterns were maintained regardless of stimulus magnitude and stimulus direction. This preliminary study suggests that, although both left and right hemispheres are involved in voice production, functional lateralization of the brain may exist for voice *F*0 control. Results also suggest that lateralization varies along the anterior-posterior dimension of the neural axis.

4pSC13. Measuring glottal area and flow in an excised canine larynx model using stereoscopic particle imaging velocimetry. Ephraim Gutmark (Univ. of Cincinnati, 799 Rhodes Hall, P.O. Box 210070, Cincinnati, OH 45221, ephraim.gutmark@uc.edu) and Shanmugam Murugappan (Univ. of Cincinnati Medical Ctr., Cincinnati, OH 45267-0528)

As opposed to the glottal area waveform, the opening and closing phases of glottal flow are asymmetric. Near the end of the closing phase, flow exiting the glottis has a rapid deceleration not seen during the opening phase; this "skewing" of the flow waveform is important for producing loudness and higher frequency harmonics. The major mechanism for the skewing of the flow waveform, relative to the area waveform, has been attributed to the nonlinear interaction between the vocal tract and the glottal flow. Previous work from this laboratory supports a hypothesis that both area and flow skewing are seen even without the vocal tract due to intraglottal flow separation vortices producing relative negative pressure. To help test this hypothesis in an animal model, a technique is necessary that can reliably measure glottal area and flow. In this work, we will show results for both glottal area and flow during phonation, in excised canine larynges, using stereoscopic particle imaging velocimetry and high speed visualization. The results support the hypothesis that intraglottal flow separation can be one mechanism for producing both flow and glottal skewing. [Work supported by NIDCD 5K08DC005421.]

4pSC14. Unsteady laryngeal airflow simulations: An analysis of the generated intraglottal vortical structures. Mihai Mihaescu (Dept. of Aerosp. Eng. and Eng. Mech., Univ. of Cincinnati, 310 Rhodes Hall, ML 0070, Cincinnati, OH 45221, Mihai.Mihaescu@uc.edu), Sid Khosla, and Ephraim Gutmark (Univ. of Cincinnati, Cincinnati, OH 45221)

Unsteady flow simulations in diffuserlike vocal-fold models proved that laryngeal airflow generates vortical structures in the intraglottal region. However, the features of these vortices and their influence on the vocal-fold motion were not clearly addressed. The present work characterizes the intraglottal vortical structures developed during the closing phase of the phonation cycle in order to analyze their possible influence on the voice quality. It is shown that intraglottal vortices are formed on the divergent slope of the glottis, just downstream of the separation point. The core of the vortical structures is characterized by important negative static pressure values (lower as compared with the surrounding pressure field). These vortices gain strength and increase in size as are convected downstream by the flow. The mechanism for which the intraglottal flow structures are becoming stronger is attributed to the entrained air from the supraglottal region. The instantaneous pressure loads on the divergent glottal slope are not uniform (in both time and space) and dependent on the vortical structures traveling nearby the glottis. The negative static pressures associated with the intraglottal vortical structures suggest that the closing phase during phonation may be accelerated by such vortices.

4pSC15. Articulatory timing in sentence production in young normal speakers and speakers with apraxia of speech: A speed history analysis. Kana Taguchi, Michiko Hashi (Dept. of Commun. Disord., Prefectural Univ. of Hiroshima, 1-1 Gakuen-cho Mihara, Hiroshima 723-0053, Japan p724015ue@pu-hiroshima.ac.jp), and Katharine Odell (Marshfield Clinic, Marshfield, WI 54449)

Apraxia of speech (AOS) is typically considered as characterized by kinematic timing problems among articulators; however, kinematic descriptions of such phenomenon are scarce. The study describes kinematic timing based on a speed history of the lower lip and the jaw, across speaking rates in young normal speakers and speakers with AOS. Kinematic data were acquired through the x-ray microbeam system, and data of young normal speakers were taken from the x-ray microbeam speech production database. A short sentence was the speech material/target. Acoustic correlates of major speed peaks of the jaw and the lower lip, as well as their temporal relationships with each other, were investigated. Of particular interest was how changes in speech rate affect such temporal relationships in normal speakers and speakers with AOS. The goal of this study was to establish a description method of inter- and intra-articulator timing for use in analyses of point-parametrized articulatory data in sentence-level materials.

4pSC16. Experimental study and theoretical simulation of stress relaxation behavior of vocal fold lamina propria tissue. Yu Zhang, Megan Keuler, and Jack Jiang (Dept. of Surgery, Div. of Otolaryngol. Head and Neck Surgery, Univ. of Wisconsin Med. School, Madison, WI 53792-7375)

The viscoelastic properties of vocal fold lamina propria tissue play an important role in tissue modeling, as well as in clinical studies of the effects of carcinoma, scarring, atrophy, or edema on dysphonia. Although the impact of these properties has been well studied, quantification of viscoelastic behaviors such as creep and stress-relaxation continue to be important in the research. The current study examined the stress-relaxation curves of ten canine vocal fold cover samples by stretching the tissue to 5%, 10%, 15%, and 20% of the sample reference length. The force on the lamina propria tissue was then recorded for 5 mins, and graphed so the stress-relaxation response could be seen. The curves obtained were what we hypothesized, with the lamina propria tissue exhibiting exponential relaxation initially, and very slow decay at the end of the trial. These results are consistent with our previous studies on stress-relaxation predicted by the biphasic theory of vocal fold tissue [Zhang *et al.*, J. Acoust. Soc. Am. **123**, 1627–1636 (2008)]. Finite element simulation has also been given to investigate the stress distributions within the vocal fold epithelium and lamina propria tissues.

4pSC17. Videokymographic analysis of irregular vocal fold vibration in laryngeal paralysis. Miwako Kimura, Niro Tayama (Otolaryngol. and Tracheo-esophagology, Int. Medical Ctr. of Japan, Tokyo, Japan and Univ. of Tokyo, Tokyo, Japan), and Roger W. Chan (Univ. of Texas Southwestern Medical Ctr., Dallas, TX)

Despite recent advances in high-speed digital imaging methods, laryngeal stroboscopic imaging is still commonly used for the clinical diagnosis and assessment of voice disorders. Yet standard videostroboscopy, designed for the examination of periodic or near-periodic vibration, is unable to provide detailed information on irregular vocal fold vibratory patterns. This study examined the vocal fold vibratory patterns of two patients with unilateral vocal fold paralysis with high-speed digital kymographic imaging. Laryngeal vibration was examined before and after the medialization procedure of arytenoid adduction, and relevant features were extracted from the

videokymographic images. Results demonstrated two distinct vibratory frequencies for the contralateral vocal folds in both subjects prior to the procedure (269 Hz vs 361 Hz, and 114 Hz vs 154 Hz), but a single vibratory frequency following medialization (361, and 154 Hz). The restoration of vibratory symmetry after arytenoid adduction indicated that this procedure could significantly improve the synchronization and entrainment between asymmetric vocal folds, in parallel with improvements in glottal competence.

4pSC18. Perceptual cues for consonant identification. Feipeng Li and Jont Allen (Beckman Inst., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave, Urbana, IL 61801)

This research quantitatively explores the perceptual cues of initial consonants by using psychoacoustical methods. Speech sounds are encoded by across-frequency temporal onsets called events. To determine the time-frequency importance function of the consonant sounds, speech stimuli (16 nonsense CVs from the LDC-2005S22 database) are high-pass or low-pass filtered and time-truncated before being presented to normal hearing listeners. Databases of speech perception under various signal to noise ratio (SNR) conditions are constructed to investigate the effect of noise on speech recognition. A visualization tool that simulates the auditory peripheral processing, called the AI-gram, is used for the analysis of the speech events under various SNR conditions. To verify the nature of the events, a special software has been developed to convert one sound into another, starting from real speech sounds, by removing primary noise robust cues. In pilot studies with a hearing-impaired subject, it is shown that feature boosting improves the robustness of select speech consonants to noise.

4pSC19. The effect of articulatory placement on acoustic characteristics of nasalization. Panying Rong and David P. Kuehn (Dept. of Speech and Hearing Sci., Univ. of Illinois at Urbana-Champaign, 901 S. Sixth St., Champaign, IL 61820, prong2@uiuc.edu)

Vowels adjacent to nasal consonants (/m/, /n/, /ŋ/ in English) tend to be nasalized. The effect of velopharyngeal opening on vowel nasalization has been studied by Feng *et al.* (1996) and Pruthi *et al.* (2007) with a transmission line model of the vocal tract based on MRI data of the area function. In order to find out the effect of articulatory placement on acoustic properties of nasalized vowels, the current study simulated the transfer function of coarticulated vowels in different nasal-vowel utterances. The results revealed different acoustic characteristics of the same vowel in various nasal contexts, which demonstrated the effect of articulatory placement on vowel nasalization and suggested the possibility for oral articulation to compensate for spectral changes caused by failure of velopharyngeal closure among hypernasal patients. The articulatory parameters for compensatory articulation were optimized by minimizing the spectral differences (first four formants) between the compensated nasalized vowel and its oral counterpart. The nasalized vowel was then synthesized with an articulatory synthesizer and evaluated for vowel quality and nasality in the following perceptual study.

4pSC20. The effects of coarticulatory timing and lexical frequency on vowel nasalization in English: An aerodynamic study. Jason B. Bishop (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095, j.bishop@ucla.edu)

Recently it has been suggested that a certain degree of variability in coarticulatory vowel nasalization is due to variation in the temporal alignment of nasal and oral gestures for N. In particular, the extent of vowel nasalization in VNC sequences is shown to be inversely related to the duration of the oral gesture for the nasal. Beddor [nasals and nasalization: the relation between segmental and coarticulatory timing, ICPhS (2007) present acoustic data which suggests this to be the case in English; in environments where nasals are shorter, such as VNC[−voice] versus VNC[+voice], anticipatory nasalization is longer. The present study examines vowel nasalization in such environments and attempts to corroborate the aforementioned acoustic findings with aerodynamic data. Additionally, other possible sources of variation in the extent of anticipatory nasalization are explored, namely, the effects of lexical frequency, which has been claimed to correlate negatively rather than positively with coarticulation [carborough (2004).]

4pSC21. Children form tense and lax vowel classes under aerodynamic constraints. Piers Messum (112 Warner Rd., London SE5 9HQ, UK, p.messum@gmail.com)

In English, lax vowels contrast with tense vowels as follows: (1) They are always "checked" by a following consonant, (2) they require only moderate displacement of the tongue from its resting position, and (3) they are "short." If these characteristics are arbitrary and independent–as currently believed–then it is a remarkable coincidence that each divides the vowel inventory into classes with the same membership. Alternatively, these characteristics emerge together under constraints in child speech that do not appear in the adult model. Aerodynamically, the pressures in child speech are higher, flows are similar, and airways are smaller. Mechanically, the respiratory drive that supports speech is pulsatile rather than smooth. This latter difference is heightened in stress-accent languages, such as English, where a child must reinforce pulses for greater loudness on stressed syllables. The constraints created by these factors require a child to check vowels made with open articulations and lengthen those made with close ones. The former behavior protects the subglottal pressure head. The latter is the indirect result of limiting airflow to avoid unwanted turbulent noise at the point of maximum oral constriction. Increasing laryngeal resistance to do this prolongs the time it takes to dissipate a pulse.

4pSC22. Interarticulator timing in a sentence production in young normal speakers and speakers with apraxia of speech. Mayuko Fujishita, Michiko Hashi (Dept. of Commun. Disord., Prefectural Univ. of Hiroshima, 1-1 Gakuen-cho, Mihara, Hiroshima 7230053 Japan, p724019fa@pu-hiroshima.ac.jp), and Katharine Odell (Marshfield Clinic, Marshfield, WI 54449)

Apraxia of speech (AOS) is typically considered as characterized by kinematic timing problems among articulators; however, kinematic descriptions of such phenomenon are scarce. The goal of this study was to describe articulatory timing between the lips and jaw and the lips and tongue in young normal and AOS speakers across three speaking rates, using x-ray microbeam data. Data of young normal speakers were taken from the x-ray microbeam speech production database. Lip protrusion and tongue elevation timings were derived from the kinematic and acoustic data in the production of "too" in the sentence "The other one is too big" and lip and jaw closing timings for /b/ and /w/ were derived from the kinematic data of the same sentence. The temporal relationships between these articulatory movements in each event were then examined with consideration of the effect of speaking rate. The results are discussed in relation to the effect of speaking rate on interarticulator timing and interspeaker variability in interarticulator timing.

4pSC23. Articulatory and acoustic measures of vowel frontness in a study of velar-vowel coarticulation. Sylvie M. Wodzinski, Gena Rizzitano, and Stefan A. Frisch (Dept. of Commun. Sci. and Disord., Univ. of South Florida, 4202 E Fowler Ave. PCD1017, Tampa, FL 33620, wodzinsk@mail.usf.edu)

Previous work has found a strong correlation between the frontness of closure location for velar stops (measured manually from ultrasound images by a trained expert) and the frontness of the following vowel (measured by F2). In this study, semiautomatic measures of tongue frontness for vowels were made from ultrasound images of the vowel articulation. However, it was found that the acoustic measure, F2, correlated more closely with the frontness of the preceding consonant than any of the ultrasound based articulatory measures of the vowel. Explanations for why an acoustic measure of coarticulation would be better than an articulatory measure will be discussed. It may be that the ultrasound based measures do not adequately capture retraction of the tongue root, which would influence F2 and presumably affect coarticulation with the consonant. Surprisingly, it was also found that the frontness measures for the consonant were more highly correlated with F2 than the frontness measures for the vowel, suggesting that the

tongue frontness measures may be indirectly affected by tongue height in some way. Overall, the manual measures of an expert appear to be superior to semiautomatically generated measures.

4pSC24. Semiautomatic measures of velar stop closure location using the EdgeTrak software. Sabrina J. McCormick, Stefan A. Frisch, and Sylvie M. Wodzinski (Dept. of Commun. Sci. and Disord., Univ. of South Florida, 4202 E Fowler Ave., PCD1017, Tampa, FL 33620, smccorm3@mail.usf.edu)

Previous work has found a strong correlation between the frontness of closure location for velar stops (measured manually from ultrasound images) and the frontness of the following vowel (measured by F2). In this study, semi automatic measures of tongue frontness during a velar closure were made. Tongue edge traces were made using the EDGETRAK software (Li *et al.*, 2005, Clinical Linguistics and Phonetics, 545–554). Frontness was then quantified from these traces using three different measures: Bressman's anteriority index (Bressman *et al.*, 2005, Clinical Linguistics and Phonetics, 573–588), a modified version of the anteriority index created for this study, and a measure of the center of mass of the tongue created for this study. When compared to the original manual measures, the modified anteriority index correlated most highly with both the manual measurement of the consonant closure location and also with F2 of the following vowel. The modified anteriority index uses an angle based weight in the anteriority calculation (as opposed to the arbitrary weights of Bressman's anteriority index). The center of mass was the worst performing measure, and it appeared to be overly sensitive to the extreme anterior and posterior portions of the tongue edge trace.

4pSC25. Tongue body movements in speech: Straight or curved paths? Anders Lfqvist (Haskins Labs., 300 George St., New Haven, CT 06511 and Dept. Logopedics, Phoniatrics, Audiol., Lund Univ., Lund, Sweden, lofquist@haskins.yale.edu)

This paper examines tongue body movements between two vowels with particular emphasis on the shape of the movement paths. Earlier work on tongue movements in speech has mostly focused on movements for consonants. The movements analyzed are from the first to the second vowel in a sequence of vowel-bilabial consonant-vowel. Native speakers of Japanese and Italian served as subjects. The linguistic material consisted of words with a long or short labial consonant. Recordings were made using a magnetometer. To assess the movement path, the movement magnitude was calculated in two ways as a straight line, the Euclidean distance, and as the actual path, obtained by summing the individual Euclidean distances between successive samples from movement onset to offset. The ratio between the path and the Euclidean distance is 1 when the movement is a straight line and greater than 1 when the path is curved. For most of the movements, the ratio was less than 1.1, thus suggesting that these movements are almost a straight line. There was no clear difference between movements during long and short labial consonants. [Work supported by NIH.]

4pSC26. Evidence for interaction between speech rhythm and gesture. Sam Tilsen (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720, tilsen@berkeley.edu)

Temporal patterns in speech occur on multiple timescales. Speech gestures generally occur on a fast timescale; previous work has found evidence for a *c*-center effect in complex syllable onsets (e.g. /spa/), whereby initiations of tongue blade and lip movements associated with /s/ and /p/ are equally displaced in opposite directions from initiation of tongue body movement associated with the vowel [C. Browman and L. Goldstein, Phonetica **45**, 140–155 (1988)]. Speech rhythms occupy a slower timescale; in metronome-driven phrase repetition tasks, rhythmic timing is more variable for higher-order target ratios of intervals between stressed syllables and phrases [F. Cummins and R. Port, J. Phonetics **26**, 145–171 (1998)]. An experiment was conducted to investigate how these gestural and rhythmic patterns interact. Gestural kinematics were recorded using electromagnetic articulometry during a repetition task with the phrase "take on a spa." Significantly greater variance in relative timing of tongue and lower lip movements associated with /s/ and /p/ was observed with the more difficult (less harmonic) target rhythms. Within-gesture effector synergies between jaw and tongue differed systematically across rhythmic conditions. These results demonstrate a substantial interaction between rhythmic and gestural systems. Observed variability patterns are simulated with a dynamical model of phase-coupled oscillators.

THURSDAY AFTERNOON, 13 NOVEMBER 2008 LEGENDS 12, 2:00 TO 3:45 P.M.

Session 4pSP

Signal Processing in Acoustics: Target Tracking and Beamforming

Charles F. Gaumond, Chair
Naval Research Lab., 4555 Overlook Ave., S.W., Washington, D.C. 20375-5350

Contributed Papers

2:00

4pSP1. Multitarget tracking using acoustic arrays. R. Daniel Costley, Jay E. Williams, Robert C. Clark, Matthew A. Gray, Zachary Williams, Gary Harrington (Miltec Res. and Technol., A Div. of Miltec Corp., A Ducommun Co., 9 Industrial Park Dr., Oxford, MS 38655), William G. Frazier (Independent Consultant, Oxford, MS 38655), and Jere Singleton (U.S. Army Space Missile Defense Command, Huntsville, AL)

A tracking system has been developed, which consists of three or more five-microphone arrays. In previous experiments, these arrays have been deployed and used to track broadband, high-speed (subsonic) airborne, acoustic sources. Recent developments to this system have been made so that multiple acoustic sources can be tracked simultaneously. A field test was conducted this past year, which demonstrated this capability by tracking two low-flying learjets. The tracking system consisted of six acoustic arrays spaced approximately 400 m apart. The bearing angles from each array to the acoustic sources were determined from the blind source separator (BSS) algorithm. The BSS utilizes a mathematical propagation model along with an optimization technique, which minimizes the difference between the model and the measured data to yield a bearing to the target. The bearing angles from two or more arrays are transmitted to a master node, which uses this information to estimate the position, speed, and heading of the targets using a tracking algorithm based on the extended Kalman filter algorithm. The field test will be described and the results will be presented and discussed.

2:15

4pSP2. A novel approach for the localization of sound sources. Piervincenzo Rizzo and Giacomo Bordoni (Dept. of Civil and Environ. Eng., Univ. of Pittsburgh, 949 Benedum Hall, 3700 O'Hara St., Pittsburgh, PA 15261, pir3@pitt.edu)

Targeting people or objects by passive acoustic sensors is of relevant interest in several military and civil applications, spanning from surveillance

and patrolling systems to teleconferencing and human-robot interaction. To date methods and patents focused solely on the use of beamforming algorithms to compute the time of arrival of sounds detected by using omnidirectional microphones that are sparsely deployed. This paper describes the preliminary results of a novel approach devoted to the localization of ground borne acoustic sources. It is demonstrated that an array made of at least three unidirectional microphones can be exploited to detect the position of the source. Pulse features extracted either in the time domain or in the frequency domain are used to identify the direction of the incoming sound. This information is then fed into a semianalytical algorithm devoted to the identification of the source location. The novelty of the method presented here consists of the use of unidirectional microphones rather than of omnidirectional microphones and of the ability to extract the sound direction by considering features such as the pulse amplitude rather than the pulse arrival time. It is believed that this method may pave the road toward a new generation of reduced size sound detectors and localizers.

2:30

4pSP3. Real-time noise source identification using field programmable gate array technology. Kurt Veggeberg (Natl. Instruments, 11500 N. Mopac C, Austin, TX 78759, kurt.veggeberg@ni.com)

Acoustic beamforming is usually used as an off-line analysis tool for noise source identification (NSI). The computational requirements of beamforming makes real-time processing difficult to achieve in conventional NSI measurement systems. This limitation prevents conventional NSI measurement systems for applications such as on-line monitoring. This paper presents the theory, design, and development of a real-time NSI measurement system based on reconfigurable input/output field programmable gate array (FPGA) technology. Some techniques used in applying beamforming analysis methods with an FPGA are proposed. Signal processing includes time filtering of input data and spatial beamforming with special treatment to obtain the best result. Pipelining method is used to make full use of the FPGA resources and reach high performance for real-time applications. Testing examples are given to demonstrate the application for NSI.

2:45

4pSP4. Effects of multiple contacts on cued beamforming in active-passive data fusion. T. W. Yudichak and Bryan A. Yocom (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, yudi@arlut.utexas.edu)

Active-passive data fusion seeks to combine the information measured on multiple sonar sensors to detect and track contacts more effectively than is possible with individual active or passive sensors. One way in which sonar signal processing can benefit from data fusion is cued beamforming, the allocation of beamforming resources based on the current estimate of the location of a contact. Cued adaptive beamforming (ABF) on passive arrays has been shown to provide more precise direction-of-arrival estimates than standard ABF in some circumstances involving a single contact of interest. At the same time, the performance of cued beamforming in directions away from the contact can be severely degraded relative to standard beamforming, resulting in the possible loss of detection of other contacts. This talk investigates the trade-offs between the gain in information on a single contact and the loss of information on other contacts when cued ABF is applied to passive arrays within a multisensor data fusion framework. The effectiveness of cuing different arrays with different contacts as well as alternating between standard and cued beamforming is examined and discussed. [Work supported by ONR.]

3:00—3:15 Break

3:15

4pSP5. Directional sources and beamforming. Christian Bouchard (SITE, Univ. of Ottawa, 800 King Edward, Ottawa, ON K1N 6N5, Canada and Inst. for Microstructural Sci., Natl. Res. Council of Canada, Ottawa, ON K1A 0R6, Canada, christian.bouchard@nrc-cnrc.gc.ca), David I. Havelock (Natl. Res. Council of Canada, Ottawa, ON K1A 0R6, Canada), and Martin Bouchard (Univ. of Ottawa, Ottawa, ON K1N 6N5, Canada)

Beamforming is done with an array of sensors to achieve a directional or spatially specific response. It relies on a model of the wave front (source model) arriving at the array to calculate the time delay, or frequency domain phase shift, that must be applied to the signal of each sensor so that they may be summed coherently. Beamforming may be used to improve signal to noise ratio, reduce reverberation, cancel interference, or estimate source location. In this talk the directionality of some real world sources that deviate from an ideal point source is discussed. Performance measures used to evaluate the directivity properties of a beamformer are reviewed. The validity of assuming a point source is examined and challenges for beamforming with nonpoint sources are discussed.

3:30

4pSP6. Robust direction-of-arrival estimation by understanding global acoustic scene. Mitsunori Mizumachi (Kyushu Inst. of Tech., 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 805-8550, Japan, mizumach@ecs.kyutech.ac.jp)

Direction-of-arrival (DOA) is an important clue in acoustic signal processing. It is, however, difficult to accurately estimate DOAs in the presence of acoustic interference such as background noises and reverberation. The author has proposed a robust DOA finder with an environmental noise model, which describes spectral characteristics of background noises. Knowledge on acoustic interferences helps to make DOA estimation robust, but is difficult to be estimated exactly. The method works well only using the dominant subband components, in which the target signal is distinguished compared with interferences, in some noisy conditions. The noise model is updated time by time based on the DOA estimates to cope with nonstationary noises. Better DOA estimation gives a more accurate noise model and vice versa. In contrast, even a little error starts to cause negative feedback in DOA estimation and finally it goes to the fatal error. This paper proposes to introduce a switching scheme, which evaluates the reliabilities of DOA estimates in a stochastic manner and judge whether the noise model should be updated or not into the previously proposed DOA estimator. It is confirmed that the DOA estimator with the proposed scheme can improve its noise robustness even under nonstationary noises.

Session 4pUWa

Underwater Acoustics: Boundary Scattering

Marcia J. Isakson, Chair
Univ. of Texas at Austin, Applied Research Lab., 10000 Burnet Rd., Austin, TX 78758

Contributed Papers

1:30

4pUWa1. Imaging surface roughness with continuous-wave ultrasound reflectometry. Farid G. Mitri, James F. Greenleaf, and Mostafa Fatemi (Mayo Clinic, Dept. of Physio. and Biomedical Eng., Ultrasound Res. Lab., 200 First St., SW, Rochester, MN 55905)

Measurement of surface roughness irregularities is an important indicator of product quality for many nondestructive testing industries. Many techniques exist; however, because of their qualitative, time-consuming and direct-contact modes, it is of some importance to work out new experimental methods and efficient tools for quantitative estimation of surface roughness. Continuous-wave ultrasound reflectometry (CWUR) is presented here as a novel nondestructive modality for imaging and measuring surface roughness in a noncontact mode. In CWUR, voltage variations due to phase shifts in the reflected ultrasound waves are recorded and processed to form an image of surface roughness. An acrylic test block with surface irregularities ranging from 4.22 to 19.05 μm as measured by a coordinate measuring machine (CMM) is scanned by an ultrasound transducer having a diameter of 45 mm, a focal distance of 70 mm, and a central frequency of 3 MHz. It is shown that CWUR technique gives very good agreement with the results obtained through CMM inasmuch as the maximum average percent error is around 11.5%.

1:45

4pUWa2. The effects of roughness on the frequency dependence of specular scattering. Marcia Isakson, R. Abe Yarbrough, and Nicholas Chotiros (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713, misakson@arlut.utexas.edu)

Specular scattering data were collected at grazing angles from 7 to 77 deg grazing and frequencies from 5 to 50 kHz at the experimental validation of acoustic modeling techniques (EVA) sea test near Isola in October 2006. High resolution microbathymetry data taken using a laser line scan system were also collected. The amplitude and phase of the specular scattered data exhibit a distinctive frequency dependence. This frequency dependence may be due to Bragg scattering from small interface ripples on the ocean bottom indicated by peaks in the measured interface roughness power spectrum. This study simulates the scattering using realizations of interface roughness in a finite element model to determine if the frequency dependence is due solely to scattering or if other mechanisms such as layering are important factors. Results will be compared to analytic approximations such as the Kirchhoff approximation as well as exact integral equations. [Work sponsored by the ONR, Ocean Acoustics.]

2:00

4pUWa3. Finite element modeling of acoustic scattering from rough interfaces. R. Abe Yarbrough and Marcia J. Isakson (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

The finite element method is applied to the problem of acoustic scattering from rough surfaces. This method has the advantage that the surface can have almost any form whereas analytic approximation methods generally require constrained surface roughness parameters. Scattering from a rough pressure release boundary and from a rough interface between two fluid media is considered in two dimensions. The problem domain is truncated with perfectly matched layers. These greatly decrease the number of degrees of freedom in the finite element problem while minimizing unwanted reflections of outbound energy back into the physical domain. The Helmholtz–Kirchhoff integral is used to extend the solution to points outside the modeled domain. Using finite elements as well as an exact integral equation method, solutions are obtained for multiple realizations of the rough scattering surface. The scattering strengths of energy reflected back into the water and of energy penetrating the sea floor in the case of bottom scattering are calculated over a range of scattering angles. The two exact numerical methods show good agreement. Results are compared with those of analytic approximation methods. [Work sponsored by ONR, Ocean Acoustics.]

2:15

4pUWa4. Scattering from an ocean bottom layer using steady-state and transient radiative transfer. Jorge E. Quijano and Lisa M. Zurk (Northwest Electromagnetics and Acoust. Res. Lab., Dept. of Elec. and Comput. Eng., Portland State Univ., 1900 SW 4th Ave., Ste. 160, Portland, OR 97201, zurkl@pdx.edu)

Research on volume scattering from a layered ocean bottom containing random media is usually conducted through methods based on the wave equation. Radiative transfer (RT) theory is an alternative formulation based on the conservation of energy and empirical laws that has the potential of being computationally more efficient and yet can provide considerable insight into the scattering phenomena. The RT theory has been successfully used in electromagnetic remote sensing and characterization of materials using ultrasound. In this work, the RT equation is applied to ocean acoustics to obtain the volume scattering from a layer of sand with flat boundaries overlying a limestone half space. Steady state results are presented to discuss the main processes that contribute to volume scattering and the transformation of energy between shear and longitudinal waves at the boundaries of the layer and at the scatterers. The transient solution of the RT equation as a function of frequency is outlined by introducing an incident pulse of infinitesimal duration that approximates an impulse. The time domain response for a broadband pulse of finite duration is then obtained by convolution with the computed impulse response, which would allow comparison with experimental chirp sonar data. [Research sponsored by the Office of Naval Research, Grant No. N000140510886.]

2:30

4pUWa5. Characterization of the near scattered acoustic vector field. Robert Barton, III and Kevin Smith (Sensors Sonar Dept., Naval Undersea Warfare Ctr., Newport, RI 02841, robert.barton@navy.mil)

In this study, we investigate the properties of the scattered acoustic vector fields generated by simple geometric objects, including the infinite rigid plate, disk, and sphere. Analytical solutions are derived from acoustic target strength scattering models in the near-field region. Of particular interest is the understanding of the characteristics of energy flow of the scattered acoustic vector field in the near- to far-field transition region. We utilize the time and space separable instantaneous active and reactive acoustic intensities to investigate the relative phase properties of the scattered field. Numerical results are presented for the near region scattered acoustic vector field of simple objects in both two and three dimensions.

2:45—2:55 Break

2:55

4pUWa6. An initial look at a continuing shallow water vector sensor acoustic ambient intensity study. David Deveau (Naval Undersea Warfare Ctr., Detachment AUTEC, PSC 1012 Box 701, FPO, AA 34058, david.deveau@autec.navy.mil) and Anthony Lyons (Penn State Univ., University Park, PA 16802)

Deployed in June 2008 off the Coast of Andros Island, Bahamas, seven compact Wilcoxon TV-001 vector sensors have been placed in 15 m of water with the goal of characterizing the ambient noise acoustic intensity with respect to environmental conditions. The study's objective is to better understand how intensity fields describe the shallow water noise environment. The system is designed to gather raw ambient acoustic noise pressure and particle acceleration every hour for a period of one year. The environmental database includes weather and wave information for the local area that can be correlated to the acoustic data. This report will present insight into the overall array development and deployment as well as an initial correlation of intensity fields to the recognized environmental stimulus. Results will be from individual sensors and these same sensors combined into vertical and horizontal arrays utilizing classic linear beamform processing. In addition to gathering data on acoustic ambient noise, data from local passing rain storms, marine mammals, fixed sources, and moving platforms will also be collected and analyzed.

3:10

4pUWa7. Statistics of synthetic aperture sonar image resolution degradation. Shawn F. Johnson, Anthony P. Lyons (Penn State Graduate Program in Acoust., Appl. Res. Lab., 117 Appl. Sci. Bldg., State College, PA 16804), and Douglas A. Abraham (CausaSci LLC, Arlington, VA 22205)

Synthetic aperture sonar (SAS) image statistics can often be characterized by a probability density function with a heavier tail than the expected Rayleigh distribution. The *K*-distribution shape parameter can be used as a metric of non-Rayleighness, with physical ties to both seafloor properties and sonar parameters. Recent results have shown that increasing the resolution cell size, or degrading the resolution of the image, produces images with statistics tending toward Rayleigh (i.e., a higher *K*-distribution shape parameter). In general, a doubling of the resolution cell area results in a doubling of the *K*-distribution shape parameter. A caveat to this generalization is the orientation of the sonar system to any features that may exist on the seafloor (i.e., sand ripples). In such a situation, image statistics may continue to be significantly non-Rayleigh for certain orientations in spite of resolution degradation. SAS images of seafloors with various bottom types and feature orientations collected with an AUV by the Naval Surface Warfare Center Panama City Division have been postprocessed to analyze changes in image statistics as the image resolution is degraded. Results of the image degradation on statistics will be discussed. [Work supported by ONR Grant Nos. N00014-04-1-0013 and 1N00014-06-1-0245, and Code 32.]

3:25

4pUWa8. Bistatic specular reflection by a rigid cone. Philip Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

It is possible to gain some insight into the high-frequency scattering of sound by objects in water by considering the presence or absence of rays constructed from geometric considerations. This presentation concerns the evolution of rays reflected from the sides of a vertical rigid cone for the case of illumination by a plane wave at an arbitrary grazing angle. The grazing angle with respect to a horizontal plane is usually taken to be small. The direction of the reflected rays depends on where the incident ray contacts the cone as specified by an azimuthal angle viewed from above. Define the meridional plane as that plane which contains the incident wave vector and the axis of the cone. Incident rays offset from the meridional plane can be reflected with relatively small vertical components in their wave vector. This analysis has implications for how bistatic hydrophones may be deployed to detect specular glints from cone-shaped objects. There may also be implications for understanding the high-frequency scattering by conical seamounts in deep water. [Research supported by ONR.]

3:40

4pUWa9. Excitation of low-frequency modes of solid cylinders by evanescent and ordinary propagating waves. Aubrey Espana and Phillip L. Marston (Dept. of Phys. and Astronomy, Pullman, WA 99164-2814, aespana81@msn.com)

When using sound to detect objects buried beneath the seafloor, often the acoustic source has a large horizontal stand-off distance. In such situations there is evidence that the incident acoustic wave in sand can have a significant evanescent component. In prior work, organ-pipe modes of water-filled shells were excited in a laboratory simulation, the most significant result being the "double spatial decay rate" effect [Marston *et al.*, J. Acoust. Soc. Am. **122**, 3034 (2007)]. To further understand the coupling by evanescent waves into low-frequency modes of cylinders, backscattering by small solid aluminum cylinders was studied with ordinary-wave illumination. Free-field experiments identified resonances worthy of investigation in an evanescent wave experiment. Several of the features identified have been reproduced in FEM-COMSOL calculations by Williams of APL-UW. For both ordinary and evanescent waves, strong modes were often excited when the cylinder was highly tilted. A further understanding of the modal features was gained by evaluating the temporal variance of the response. Furthermore, the aluminum cylinders also showed an enhanced spatial decay rate when compared to that of the evanescent soundfield. [Work supported by ONR.]

Session 4pUWb

Underwater Acoustics: Forward Scattering and Reverberation

Kevin D. LePage, Chair
NATO Undersea Research Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy

Contributed Papers

4:00

4pUWb1. Rapid simulation of sonar time series. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia (SP), Italy)

Rapid simulation of element level time series for bistatic sonars requires quality approximations in order to yield accurate results in real time. Here rapid physics-based time series approximations based on closed form expressions for reverberation intensity developed by Harrison [J. Acoust. Soc. Am. **114**, 2744 (2003)] are developed for boundary reverberation and target scattering in shallow water waveguides. The simulations include Doppler and waveguide angle and time dispersion effects, and since the simulations are at the element level, they can be processed like real data. Simulations are compared to reverberation predictions obtained with more accurate but slower codes showing good agreement.

4:15

4pUWb2. Higher moment estimation for reverberation simulation. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia (SP), Italy)

In recent work the second, third, and fourth spatial moments for exponentially distributed roughness with either Gaussian or von Karman spatial spectra have been presented by the author. In this talk results for the second, third, and fourth moments of reverberation pressure for perturbation theory scattering from surfaces with von Karman spectra are obtained using numerical quadrature. Results are compared to closed form expressions for the second and fourth moments obtained for Gaussian correlation functions. Methods to extend these results to the class of chi-square distributed roughness heights of arbitrary order, which encompasses both exponential and Gaussian height distributions, are also presented.

4:30

4pUWb3. Examination of loss mechanisms for a rough bottom Pekeris waveguide using a two-way coupled mode model. Steven Stotts, Robert Koch, and David Knobles (Appl. Res. Labs., The Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78752)

A benchmark scattered-field solution for a two-dimensional rough-bottom Pekeris ocean waveguide has been generated from a two-way coupled-mode model with normal modes evaluated in the complex horizontal wave-number plane. The received level time series coupled-mode solution for a range monostatic source/receiver geometry with a 250 Hz center frequency, 20 Hz band, Gaussian pulse was shown previously to be consistent with a single scatter Born approximation solution [Stotts *et al.*, J. Acoust. Soc. Am. **122**, 3075 (2007)]. A new result for a 1 kHz center frequency, 60 Hz band, pulse deviates significantly from the Born approximation solution. Alternative approaches, obtained with the Born approximation but including the additional propagation loss for Kirchhoff scattering [E. I. Thorsos, J. Acoust. Soc. Am. **83**, 78–92 (1988)] with and without second order small slope approximation effects are examined. The solutions at 1 kHz with these alternative methods are compared to the coupled-mode solution. The forward scattered fields propagated over finite length roughness segments of varying lengths are analyzed to understand the role of mode-mode interactions in the forward propagation loss mechanisms and for comparison with the Kirchhoff scattering loss description.

4:45

4pUWb4. Time-domain solutions for Rayleigh and Stoneley waves using the single-scattering parabolic equation method. Adam M. Metzler, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180), Michael D. Collins (Naval Res. Lab., Washington, DC 20375), Robert A. Zingarelli, and Stanley A. Chin-Bing (Naval Res. Lab., Stennis Space Ctr., MS 39529)

The parabolic equation method implemented with the single-scattering correction accurately handles range-dependent environments in elastic layered media. Interfaces between elastic media may be treated efficiently by subdividing into a series of two or more single-scattering problems [Küsel *et al.*, J. Acoust. Soc. Am. **121**, 808–813 (2007)]. In addition to environmental waveguide parameters, the procedure uses several computational parameters. The impacts of the number of interfacial scattering problems, an iteration scheme convergence parameter, and the number of iterations for convergence are shown on the accuracy and efficiency of the method. In particular, selection criteria for these parameters are developed. Fourier transforms and syntheses generate time-domain solutions for seismic applications of interest. Examples for model waveguides show features of Rayleigh and Stoneley wave propagation, and comparisons with solutions from other methods are shown. [Work supported by the ONR.]

5:00

4pUWb5. Monte Carlo simulation of rough surface scattering using a modified pseudospectral time-domain method. Yonghoon Ha, Keunwha Lee, and Woojae Seong (Dept. of Ocean Eng., Seoul Nat'l Univ., San 56-1, Sillim-dong, Seoul 151-744, Korea)

A pseudospectral time-domain (PSTD) method using a surface flattening transformation and image method is applied to a randomly rough pressure-release surface scattering problem. Above, the PSTD method efficiently avoids the Gibbs phenomenon appearing in the conventional PSTD method when treating irregular boundaries with large impedance contrast. The pressure-release surface is generated by Pierson–Moskowitz spectrum and Monte Carlo method is used to calculate the statistical properties of the rough surface. The Monte Carlo results for elemental scattering area will be investigated by comparing with Chapman–Harris formula.

5:15

4pUWb6. Three-dimensional spatial coherence measurements: Vertical, longitudinal/transverse horizontal coherence. Lin Wan, Ji-Xun Zhou, Peter Rogers (School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, lin.wan@gatech.edu), and David Knobles (The Univ. of Texas at Austin, Austin, TX 78713)

In the Shallow Water '06 experiment, two L-shape arrays were deployed. Two horizontal components of these arrays were laid on the sea bottom. One was exactly perpendicular to the direction of sound propagation. The other was exactly parallel to the direction of sound propagation. This configuration offered an opportunity to directly measure the vertical, longitudinal horizontal, and transverse horizontal coherence. The results of spatial coherence were averaged over different pairs of hydrophones and over a frequency bandwidth of 100 Hz. The vertical coherence showed receiver depth dependence. When the source and the receivers were located below the thermocline, both the vertical and longitudinal hori-

zontal coherence lengths (in units of wavelength) increased with increasing range and frequency. The longitudinal horizontal coherence length was much larger than the vertical coherence length. These results were similar to the predictions by Smith's model [J. Acoust. Soc. Am. **60**, 305–310 (1976)] with a frequency dependent bottom reflection loss [Zhou, Chin. Phys. **1**, 494–504 (1981)]. The transverse horizontal coherence length/wavelength decreased with frequency. When the source depth was within the thermocline, the transverse horizontal coherence lengths exhibited weak range dependence in the 100–300 Hz range and was larger than 40 wavelengths. [Work supported by ONR.]

5:30

4pUWb7. Model for scattered field from a vertically extended target in a range-dependent ocean waveguide. Elizabeth Küsel and Purnima Ratilal (Dept. Elelectrical and Comput. Eng., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115, kusele@alum.rpi.edu)

A model for the scattered field from a vertically extended target in a range-dependent ocean waveguide is developed by application of Green's theorem. The model is implemented with the parabolic equation (PE) method for a vertically extended pressure-release cylinder, such as the BBN target. The local scattered field at the surface of the target is approximated by the sum of Hankel functions. The total scattered field is calculated using Kirchhoff's integral approximation to Green's theorem. Results of calculations for a Pekeris waveguide show good agreement with the Ingenito normal-mode-based model for target scattering in a waveguide. A limitation with the Ingenito scattering model is that the target has to be located within an isospeed layer. The PE-based scattering model developed here, on the other hand, is able to take into account sound speed variations along the vertical extent of the target. We show that using the Ingenito model in a layered waveguide with the vertically extended target can lead to inaccuracies in estimating the scattered field. Illustrative examples are also provided for range-dependent environments.

5:45

4pUWb8. Atlantic herring low-frequency target strength and abundance estimation: Ocean acoustic waveguide remote sensing (OAWRS) 2006 Experiment in the Gulf of Maine. Zheng Gong, Mark Andrews, Daniel Cocuzzo, Saumitro Dasgupta, Purnima Ratilal (Dept. of Elec. and Comput. Eng., Northeastern Univ., Boston, MA 02115), Srinivasan Jagnnathan, Deanelle Symonds, Ioannis Bertsatos, Tianrun Chen, Nicholas Makris (MIT, Cambridge, MA 02139), Redwood Nero (Naval Res. Lab., Stennis Space Ctr., MS 39529), Hector Pena, Ruben Patel, Olav Rune Godoe (Inst. of Marine Res., Nordnes, N-5817 Bergen, Norway), and J. Michael Jech (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543)

The mean low-frequency target strength (TS) of spawning Atlantic herring populations in the Gulf of Maine is estimated from the experimental data acquired during September–October 2006 near the northern flank of Georges Bank. A low-frequency OAWRS system with an instantaneous imaging diameter of 100 km was deployed to provide spatially unaliased imaging of fish populations over wide areas. The OAWRS system's scattering strength measurements are calibrated with areal fish population density estimates obtained from concurrent localized line-transect measurements with several conventional fish finding sonars (CFFSs). Trawl sampling at selected locations enables the identification of the imaged species. The mean TS estimates of herring individuals exhibits significant variation over OAWRS operating frequency range, in accordance with the results from a resonant scattering model for swimbladder-bearing fish. The neutral buoyancy depth of herring and the species composition in the imaged population is inferred by comparing the measured TS with those derived from the model. Our analysis indicates that the herring population has a neutral buoyancy depth of between 70 and 90 m and is therefore negatively buoyant between 120 and 180 m water depth at which it is commonly found. The herring populations instantaneously imaged with OAWRS often exceeds 200×10^6, of which over 150×10^6 individuals can be organized into a large shoal.

Session 5aAA

Architectural Acoustics, Noise, and Speech Communication: Classroom Acoustics in Honor of Michael Nixon

David Lubman, Cochair
DL Acoustics, 14301 Middletown Ln., Westminster, CA 92633-3908

Louis C. Sutherland, Cochair
27803 Longhill Dr., Rancho Palos Verdes, CA 90275-3908

Chair's Introduction—8:30

Invited Papers

8:35

5aAA1. Classroom acoustics: A first step toward education for all. Karen L. Anderson (4736 Tony Sound Ln., Tallahassee, FL 32309, karenlanderson@earthlink.net)

Education is primarily provided through the medium of verbal instruction. With ever greater emphasis on test scores and teacher accountability, it is important to recognize the effects of excessive background noise and reverberation on student learning. Nixon was a leader in the movement to raise awareness of classroom acoustic effects and to achieve national written standards. He crossed professional lines to become an active member of the Educational Audiology Association, providing information and advice to almost 1000 audiologists who work for U.S. school districts. Nixon was instrumental in not only raising awareness, but raised the bar for what education audiologists should know about classroom acoustics. This paper will provide information on the listening challenges of learners with hearing loss (the most common birth defect), most of whom are educated in typically noisy classrooms. Acousticians are invited to learn more about the synergistic effects of excessive noise and reverberation on speech perception of this growing number of children and what the hearing industry is doing to address these issues. Together, the voices of the acoustics and hearing industries are needed to champion the case for listening, learning, and a better future footsteps left by Nixon.

8:55

5aAA2. Next steps toward improving classroom acoustics for all. Peggy Nelson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, peggynelson@umn.edu)

For the past ten years, the ASA has been very influential in the improvement of classroom acoustics, especially with the adoption of ANSI S12.60-2002. In fact, we have been a part of a worldwide effort to improve the learning environments for children, thanks in large part to the efforts of Nixon. Even after the adoption of the standard, we must still be involved in local and state policy decisions involving classroom acoustics, especially in urban districts with older school buildings, diverse students, and declining enrollment. We can and should form partnerships with personnel in local school districts to evaluate and fix acoustical problems in schools. Recently we worked with a local urban school district to evaluate schools and determine the best investment for improving acoustics. In some schools, sustained noise levels were well over 65 dBA, largely because of poor quality doors and windows that allowed in high levels of external noise from hallways and outside. Solutions included: damping ventilation vibration, adding door seals, and judicious use of amplification systems when higher-signal levels were needed. In most cases, reducing noise problems at the source was the best investment of district funds.

9:15

5aAA3. A green pathway to classroom acoustics: A comparison of classroom acoustic standards. Daniel Bruck (BRC Acoust. Technol. Consulting, 1741 First Ave. S., Seattle, WA 98134, danb@brcacoustics.com) and Alexis Kurtz (Arup, New York, NY 10013)

Towne seems to have started the modern classroom acoustics movement, and Nixon was one of its earliest volunteers. Nixon contributed to an important success of that movement, the development of ANSI standard for classroom acoustics, S12.60-2002. Some of Nixon's collaborations with Towne are remembered. Nixon's efforts are also honored by showing important ways in which the standard influences contemporary practice of school building. The U.S. Green Building Council has adopted elements of the ANSI standard in its evolving LEED for Schools. Why not all? The acoustical requirements of LEED for Schools are compared with S12.60-2002. Ongoing challenges of integrating good acoustics into school buildings in the context of sustainable design are discussed. The authors discuss the issues encountered in development of the LEED standard when acoustics vies with competing needs. The importance of involvement by members of the classroom acoustics community to further Nixon's work is also discussed.

9:35

5aAA4. Issues in the sound absorption treatment in classrooms. Louis Sutherland (Consultant in Acoust., 27803 Longhill Dr., Rancho Palos Verdes, CA 90275) and David Lubman (DL Acoustics, 14301 Middletown Ln., Westminster, CA 92683)

Classrooms require adequate sound absorption to establish appropriate environments for listening and learning. Sound absorbing materials improve the classroom learning environment by reducing reverberation time and by reducing background noise levels. ANSI standard S12.60-2002 provides simple methods for determining the amount of sound absorption needed to limit reverberation time. One of those methods is briefly reviewed. Throughout his career, Nixon's forte was the sound absorption treatment of rooms. He and his colleague, the late Robin (Buzz) Towne, were pioneers—sounding the alarm that, with hardly anyones notice, poor classroom acoustics had become a widespread problem adversely impacting scholastic achievement. The two were also instrumental in motivating the Acoustical Society of America to lead the change toward a solution. The Society is indebted to Nixon both for his pioneering effort to awaken acousticians and others to the need for good classroom acoustics and for his diligent participation in the subsequent effort to develop an ANSI standard for that purpose.

9:55

5aAA5. Classroom acoustics: Moving toward needed regulation. Lois Thibault (U.S. Access Board, 1331 F St. NW, Ste. 1000, Washington, DC 20004)

This year, the U.S. Access Board voted to pursue regulation of classroom acoustics. The initiative born just a decade ago from Mike Nixon's advocacy for adequate listening conditions in schools can now move forward toward enforceability with a new ADA standard based on ANSI/ASA S12.60-2002, the voluntary standard recently reballotted by ASA. Today, Connecticut requires its new schools to meet the S12.60 standard; the LEED sustainability credential rewards good acoustics in its school evaluations, and California's High-Performing Schools movement recognizes acoustical performance. Armed with data Nixon collected and shared and connected by the classroom acoustics LISTSERV he established, stakeholders have been able to influence scores of jurisdictions to require new school designs to address classroom acoustic performance. Nixon's work in improving public understanding of the need for good classroom acoustics has made today's regulatory agenda possible: he provided both the tools and the energy used to build an effective constituency for quiet classrooms. The community of support Nixon encouraged among parents, educators, audiologists, acousticians, and advocates is a necessary underpinning for progress toward regulation. This paper will document the history and anticipate the future of classroom acoustics regulation, focusing on building and access code process and provisions.

10:15—10:30 Break

Contributed Papers

10:30

5aAA6. A survey of unoccupied and occupied acoustic conditions in existing modular classrooms. Norman H. Philipp and Lily M. Wang (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, nphilipp@mail.unomaha.edu)

Modular classrooms are widely in use across the United States, and an addendum on their acoustic performance is being prepared to supplement the classroom acoustics guidelines given in ANSI S12.60-2002. This paper presents measurements made in a large number of existing modular classroom units in Omaha, NE, both in unoccupied and occupied conditions. In the unoccupied classrooms, the gathered data include (a) reverberation times; (b) background noise levels in heating, cooling, and ventilation modes; and (c) airborne sound attenuation across all four exterior walls. In the occupied classrooms, sound levels were logged in both the room interior and exterior throughout an entire school day. Correlations have been made between the interior and exterior data to determine how often the interior noise levels could be attributed to exterior sources. Summaries of the results will be provided in an effort to benchmark the current conditions of modular classroom constructions.

10:45

5aAA7. Comparing classroom acoustics in green and nongreen schools. Pamela Harght and Robert C. Coffeen (School of Architecture and Urban Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

Minimal data are available on existing conditions regarding the acoustical environment in public schools throughout the U.S. Furthermore, rarely is there communication between those involved in the design/build process and the professionals who utilize these spaces on a daily basis. The research presented for this paper was conducted in two parts. First, an electronic survey was issued to teachers of public elementary, middle, and high schools in all 50 states in green and nongreen classrooms. This survey gathered results on questions on the present condition of the teachers' classrooms in terms of background noise levels, noise from adjacent spaces and the outdoors, speech intelligibility, and the level of importance of acoustics for a teacher with regard to other indoor environmental concerns, i.e., thermal comfort, daylighting, and layout of classroom. The second part of this research involved testing of classrooms, both green and nongreen. The testing complied with ANSI S12.60-2002 for testing procedures for reverberation time, background noise levels, transmission loss, and speech interference level. Finally, the results from both parts were compared for nongreen and green classrooms regarding acoustics and classroom design for potential future applications in the school and architecture communities.

11:00

5aAA8. Graphical representation of acoustic data. Michael Ermann and John Samuel Victor (School of Architecture + Design, Virginia Tech, 201 Cowgill Hall, Blacksburg, VA 24061-0205, mermann@vt.edu)

This line of research aims to graphically represent acoustic data for clear comparisons. With a particular focus on low-frequency sound absorption, noise reduction coefficient, and impact insulation class, acoustic data are presented visually and grouped. Trends emerge when looking at some data, and in other cases similar assemblies may have dissimilar acoustic values.

11:15

5aAA9. Gymnasium room acoustics. Comparisons of different metrics, criteria, measurements, and calculation methods (including acoustical modeling software). Joseph F. Bridger and Steven S. Stulgin (Stewart Acoust. Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607)

Gyms are nondiffuse large boxes and have unusual room acoustics challenges related to safety, durability, and volume. Classic room acoustics analysis methods, such as that of Fitzroy and Sabine, measurements, as well as acoustical modeling software results are compared. Speech intelligility results from the room acoustics software are compared against traditional reverberation time criteria. What has been learned about criteria and calculation methods is shared.

FRIDAY MORNING, 14 NOVEMBER 2008 LEGENDS 4, 9:00 TO 10:45 A.M.

Session 5aNS

Noise: Topics in Noise—Active Noise, Product Noise, and Community Noise

Erica Ryherd, Chair
Georgia Institute of Technology, Mechanical Eng., 771 Ferst Dr., Atlanta, GA 30332-0405

Contributed Papers

9:00

5aNS1. Noise levels in computer data centers: Potential occupational noise hazard. Matthew Nobile (IBM Hudson Valley Acoust. Lab., M/S P226, Bldg. 704, Boardman Rd. Site, 2455 South Rd., Poughkeepsie, NY 12601)

The noise levels in modern high-density data centers are encroaching on occupational noise limits, such as those set by OSHA laws in the United States or EC Directives in Europe. This is surprising to many because the high-tech computer data center is not usually thought of as a workplace that can cause hearing damage. Over the past decade two trends have contributed to this. First, high-end servers are being packaged more and more densely into a single rack, where the cooling fans or blowers now have to operate at higher speeds and airflow volumes to properly cool them. Secondly, more and more of these systems are being installed on data center floors, with racks often butted one against the other "as far as the eye can see." An associated problem is as follows: How to predict ahead of time what the sound pressure levels might be in a data center? This involves solving the following emission-to-immission problem: How to translate the emission sound power levels of individual racks into resulting immission sound pressure levels in the room? This paper will present an initial assessment of the noise exposures that exist in today's data centers and present some initial modeling results.

9:15

5aNS2. Lowering the effect of low-frequency noise generated from boiler exhaust stacks. Kevin Richardson and Byron Davis (Vibro-Acoust. Consultants, 490 Post St., Ste. 1427, San Francisco, CA 94102, kevin@va-consult.com)

Low-frequency noise from exhaust boiler stacks at a semiconductor plant was measured and the low-frequency noise spectrum was predicted using an analytical impedance model. Nearby residents complained that noise was rattling household items, such as windows and light fixtures. The semiconductor plant speculated that noise from two boiler exhaust stacks was the source of the problem. Sound pressure level measurements conducted at the plant revealed strong tones in the infrasound and low-audible frequency range. The boiler exhaust stack noise source was theoretically characterized by deriving the mechanical impedance of the boiler exhaust stack and the radiation impedance of the top of the boiler exhaust stack as a simple unbaffled source. The analytical model predicted the resonant frequencies of the boiler exhaust stacks and these resonances closely matched the strong low-frequency tones measured. Using this model, possible changes in controlling parameters of common exhaust boiler stack design are described that could decrease the overall noise radiating from the exhaust stacks. As one example, the plant appreciably decreased the sound power contribution from the upper exhaust stack resonances by lowering the overall output of the exhaust boilers. This change lowered the overall noise levels without the use of silencers.

9:30

5aNS3. Active control of diffuse sound fields using generalized energy density. Buye Xu and Scott Sommerfeldt (Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84602, buye.xu@gmail.com)

In enclosures, the total acoustic energy density (ED) has been shown both theoretically and experimentally to be more spatially uniform than the squared pressure. The generalized energy density (GED) has even more uniform properties than the acoustic energy density. In active noise control applications, the standard approach taken is to minimize the squared pressure response in the field. However, the use of ED as the minimization quantity has been demonstrated to yield improved performance in low modal density acoustic fields, often resulting in improved global attenuation. For diffuse acoustic fields (high modal density), local "quiet zones" can be achieved, with the volume of this quiet zone typically being characterized as a sphere with a diameter of about one-tenth of a wavelength when the squared pressure is minimized. It has been found that this performance can also be improved through the use of GED. By controlling GED instead of squared pressure, one cannot only increase the size of the quiet zone but also decrease the acoustic power added into the system. Results will be shown to illustrate the improved performance.

9:45

5aNS4. Active noise control for a short duct. Ho-Wuk Kim and Sang-Kwon Lee (Dept. of Mech. Eng., Inha Univ., Incheon, Korea)

Finite impulse response (FIR) filter for an adaptive filter algorithm is mostly used for an active noise control system. However, an FIR filter needs to be larger in size of the filter length than of its infinite impulse response (IIR) filter. Therefore, the control system using the FIR adaptive filter has slow calculation time. In the active noise control system of the short duct, the reference signal can be affected by the output signal, so IIR filter for the ARMA system can be more suitable for the active noise control of the short duct than FIR filter for the MA system. In this paper, the recursive LMS filter, which is the adaptive IIR filter, is applied for the active noise control inside the short duct. For faster convergence and more accurate control, a variable step size algorithm is introduced for this recursive LMS filter (R-VSSLMS filter). Using this algorithm and considering the secondary path, the filtered-u R-VSSLMS is conducted successfully on the real experiment in the short duct. The performance of the active control using the filtered-u R-VSSLMS filter is compared with the performance of the active control using a filtered-x LMS filter.

10:00

5aNS5. Analysis of commonly witnessed vehicle accident sounds *in situ*. William Neale and Toby Terpstra (Kineticorp, 44 Cook St., Ste. 510, Denver, CO 80206)

Research by Harber and Harber demonstrates the inaccuracy human beings have when recounting their experience witnessing a vehicle accident. However while witness statements can be unreliable, what a witness might have seen or heard is nonetheless important and cannot simply be disregarded. A witnesss recollection of audible sounds such as tire screech-

5a FRI. AM

ing, vehicle acceleration, or impacts between objects can profoundly affect how an accident sequence is interpreted by experts in vehicular accident reconstruction who may rely on witness statements when there is a dearth of physical evidence that properly defines the sequence of events. This paper provides a means for assessing the validity of what witnesses hear by recreating, recording, and measuring commonly heard accident sounds in real world conditions. Sounds, such as engine noise, tire screeching, and impacts, are recreated using various vehicles and in various environments. These sounds are recorded, observed, and analyzed to provide an understanding of how these sounds might be experienced by a witness. Among some of the testing variables are vehicle speed, roadway temperature, vehicle type, roadway surface, and background environment. These factors are evaluated in how they affect sound quality and clarity, frequency spectrum, sound level, and the sound's reverberation and directivity.

10:15

5aNS6. Statistical learning approach applied to road surface classification. Joel Paulo (DEETC, ISEL-Tecnical Inst. of Lisbon, 1959-007 Lisbon, Portugal, jpaulo@deetc.isel.ipl.pt) and José Bento Coelho (CAPS, Instituto Superior Tecnico, TU Lisbon, Lisboa, Portugal)

Measures aiming environmental noise abatement usually consider acoustic barriers alongside the road. However, the cost associated with these measures is usually considerably high and its performance in urban areas is reduced. The problem of the visual impact is another issue affecting the communities. Nowadays, road planners have started to consider silent surfaces as an alternative. These types of surfaces are constituted basically by changing the texture and/or porosity of the mixtures. In some conditions, noise level abatements up to 15 dB can be achieved. Therefore, a considerable variety of different surfaces are available. The main goal of this research is to identify and classify different types of road pavements by analyzing the noise profile, using the close proximity method. Feature extraction and selection is one of the first procedures on a classifier algorithm. Moreover, the accuracy of the results is strongly dependent on the right choice of the selected feature vector. Standard classifiers are being tested in order to establish guidelines for future developments of this research. In situations of net road surveillance, searching for inhomogeneities on the surface and the presentation of the results in a geographic map, showing the locations of the surface types and the noise levels, can improve the accuracy of the noise mapping models.

10:30

5aNS7. Investigation and analysis of urban noise for sustainability. Martha G. Orozco-Medina (Inst. de Med. Amb. y Com. Hum. Dep. de Cs. Amb., CUCBA, Univ. de Guadalajara, Km 15.5 Carretera a Nogales, Las Agujas Zapopan, Jalisco, Mexico.), Arturo Figueroa-Montao, and Javier Garca-Velazco (Univ. de Guadalajara, Jalisco, Mexico)

The concept of urban sustainability involves complex issues such as civil services, social participation, resilience, productivity, health, and development into a multilevel approach of environmental, social, and productivity sectors. Viewing environmental noise as part of urban dynamics is essential. However it is almost absent or underestimated in most current environmental, social, and health policies in the developing world. Efforts to investigate noise pollution toward analysis, regulation, fulfillment, inspection, and fines contribute in a positive way to improve acoustic quality within urban communities. Therefore the ultimate goal for decision makers is to consider noise as a key issue when establishing or discussing management policies of urban areas and to transfer the approach to lower levels with preventive and educational measures, rather than restrictive or prohibited, in order to attain sustainability of urban communities.

FRIDAY MORNING, 14 NOVEMBER 2008 LEGENDS 10, 8:30 TO 11:30 A.M.

Session 5aPA

Physical Acoustics: Outdoor Sound Propagation

Claus Hetzer, Chair

Univ. of Mississippi, Natl. Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

8:30

5aPA1. Applying well-known diffraction models to the sound field in the shadow zone of an isolated building. W. C. Kirkpatrick Alberts, II, John M. Noble, and Mark A. Coleman (U.S. Army Res. Lab., Attn. AMSRD-ARL-CI-ES, 2800 Powder Mill Rd., Adelphi, MD 20783)

An isolated building, a fundamental case of urban acoustics, has recently been the subject of an experimental effort to characterize the building's influence on propagating sound [Alberts and Noble, J. Acoust. Soc. Am. **121**, 3064 (2007)]. As a precursor to modeling efforts involving finite-difference time-domain simulations, two common diffraction models, the frequency domain model of Pierce and Medwin's extension of Biot and Tolstoy's time-domain model, have been utilized to estimate the sound pressure at various locations in the building's acoustic shadow. The discussion will include the models used and comparisons between calculated and measured data, demonstrating reasonable agreement between measured and calculated spectra in limited cases at some frequencies.

8:45

5aPA2. Finite-difference time domain simulations of outdoor sound propagation around manmade structures. Sandra L. Collier, W. C. Kirkpatrick Alberts, II, Leelinda Parker, and John M. Noble (U.S. Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783)

We develop a finite-difference time domain (FDTD) model of the acoustic propagation around outdoor man-made structures that includes the effects of atmospheric turbulence and porous ground surfaces. The wind flow around the building is determined with a Navier–Stokes approach, whereas, the basic acoustic propagation model is based on the coupled first-order partial differential equations for linear acoustic propagation in a dynamic environment developed in by [Collier *et al.*, Proceedings of the 2005 MSS BAMS]. Special numerical techniques are required to model the sound interaction at the man-made structure, in particular, for buildings with complicated geometries. For two-dimensional propagation, this numerical model can be run on a standard desktop computer. However, for three-dimensional

propagation, high performance computers are needed. Here we present results from the numerical simulations and compare them to recently collected data [Alberts and Noble, J. Acoust. Soc. Am. **121**, 3064 (2007)].

9:00

5aPA3. Predicted and experimental attenuation of obstructed projectile shock waves. James Perea and Brad Libbey (Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Fort Belvoir, VA 22060)

Acoustic sniper localization algorithms are well established for open field detection; however, these algorithms are less accurate in urban environments due to reverberation and diffraction. Research is being performed to understand the effects of obstructions on shock wave propagation. Barrier attenuation calculations that use Fresnel number to characterize obscurity will be compared to measured attenuation of shock wave frequency components. An artificial building was set up and shock waves from bullets were recorded by microphones with varying degrees of obscurity. The experimental setup allowed for a range of Fresnel numbers between 5 and 11 resulting in attenuation up to 30 dB. Attenuation increased as obscurity and Fresnel number increased up to a threshold. Beyond this point the magnitude of attenuation decreases in both theory and experiment. These data support the relationship between attenuation and microphone obscurity and may provide a means to estimate the amplitude and *N*-wave slope of the shock wave prior to obstruction.

9:15

5aPA4. Numerical simulation of sonic boom propagation through atmospheric turbulence. Andrew Piacsek (Dept. Phys., Central Washington Univ., 400 E. University Way, Ellensburg, WA 98926-7422, piacsek@cwu.edu), Lance Locey, and Victor Sparrow (Penn State Univ., State College, PA 16802)

To better understand the mechanisms by which atmospheric turbulence alters sonic boom rise times and peak overpressures, numerical calculations of sonic boom propagation through atmospheric turbulence have been performed using the NPE time domain model. Turbulence is incorporated into the model as a perturbation of the ambient sound speed that has a random spatial distribution determined by a von Karman energy spectrum. Each simulation employs a new realization of the turbulent field. Output from the finite difference solution includes high resolution movies showing the evolution of a full two dimensional wave field as it propagates from the upper region of the turbulent boundary layer to the ground. The evolution of wave forms corresponding to specific locations along the wave front is also obtained. Results are presented that illustrate how the magnitude, complexity, and spatial and temporal variabilities of the sonic boom wave field depend on turbulence spectrum parameters that represent atmospheric conditions. [V. Sparrow supported by NASA through Wyle Laboratories and by the FAA/NASA/Transport Canada PARTNER Center of Excellence.]

9:30

5aPA5. Acoustic tomography of the atmosphere at the Boulder Atmospheric Observatory. Vladimir E. Ostashev, Alfred J. Bedard (NOAA/Earth System Res. Lab., Boulder, CO 80305), Sergey N. Vecherin, and D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755)

An array for acoustic tomography of the atmosphere has been built at the NOAA Boulder Atmospheric Observatory. In this paper, a short description of the array and some acoustic tomography results are presented. The array consists of three speaker and five microphone towers located along the perimeter of a square with a side length of 80 m. The towers are 9.1 m high. The speakers and microphones can be located at different (multiple) levels on the towers to do three-dimensional tomography. The transducers are connected via cables with the central command and data acquisition computer. The array enables measurements of travel times of sound propagation between different pairs of speakers and microphones. The measurements are done repeatedly within a short time interval so that the information about the temporal change in the travel times can be employed in tomographic reconstruction. Then, these travel times are used as input data in a time-dependent stochastic inversion for reconstruction of temperature and wind velocity fields. Examples of the reconstructed turbulence fields are presented and discussed. [Work supported by the Army Research Office.]

9:45

5aPA6. Comparison of several acoustic models with measured data for a pure tone sound source. Bruce Ikelheimer, Micah Downing, and Michael James (13 1/2 W. Walnut St., Asheville, NC 28801)

A unique data set of outdoor acoustical measurements has been collected in a study area of over 800 sq km. Over 100 pure tone, high-level sources are uniformly distributed within this area. These sources were sounded simultaneously for 4 min on two consecutive days and were recorded by 24 sound level meters also distributed within the study area. In addition to the sound data, surface meteorological data were collected at the sound level meters site along with local area weather service data. The sound level meters collected third-octave band 1-s Leqs, while the weather data collected included wind speed, direction, temperature, humidity, and atmospheric pressure every 5 s. The terrain in the study area is hilly with varying degrees of vegetation and some large bodies of water. This data set provides a unique opportunity to compare various outdoor sound propagation models with measured data for propagation distances from 375 m up to several km. This presentation will compare the measured results with model calculations including simple and complex ray-tracing methods and a parabolic equation method.

10:00—10:15 Break

10:15

5aPA7. Infrasound studies of hurricanes. Claus Hetzer, Carrick Talmadge, Roger Waxler, and Kenneth Gilbert (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, claus@olemiss.edu)

Hurricanes, whether in the Gulf of Mexico or the open ocean, are known to produce infrasound in the microbarom band (around 0.2 Hz). Infrasound technology is an excellent choice for monitoring hurricanes because portable, sensitive, and relatively low-cost infrasound arrays can be deployed quickly in areas out of danger from storm winds and waves while still monitoring storm-related signals continuously. The National Center for Physical Acoustics is developing a Gulf Coast hurricane monitoring system involving permanent and portable infrasound arrays. This discussion will focus the results of infrasound studies of the 2008 hurricane season, along with background information on the history and theory of hurricane microbaroms. The latest results of theoretical calculations about the microbarom source and possible effects of storm winds on bearings will also be presented.

10:30

5aPA8. Frequency response, acoustic impedance and background noise of microbarometers. Damien Ponceau (CEA/DASE, Bruyres le Chtel 91297, Arpajon Cedex, France), Benoit Alcoverro (CEA/DEV, BP2 33114, Le Barp, France), and Serge Olivier (CEA/DASE, Bruyres le Chtel 91297, Arpajon Cedex, France)

Frequency response and background noise of microbarometers are derived from their principle of operation. Lumped element models are proposed to describe each part of these absolute infrasound sensors and are used to derive analytical expressions for frequency response, acoustic impedance, and background noise of microbarometers. A set of experimental methods to estimate all elements is discussed in this paper. Some of them need a specific instrumentation developed in our laboratories. Others are very simple as they require only common equipment and can be applied on field to calibrate these sensors. All these methods have been applied to a set of MARTEC MB2005 microbarometers. Results from the theory and measurements are compared in this paper.

10:45

5aPA9. Characteristics of porous-hose infrasonic wind-noise-reducing filters. Claus Hetzer, Jin So, Carrick Talmadge, Richard Raspet, Jeremy Webster (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, claus@olemiss.edu), and Douglas Shields (MilTec Res. and Technol., Oxford, MS 38655)

Spatial wind-noise-reducing filter technologies are widely used in infrasound recording because wind is the primary source of noise in this frequency band. While permanent infrasound arrays often make use of large solid-pipe rosette spatial filters, portable or temporary arrays must use other materials, the most popular of which is porous garden hose. While undeniably effective at reducing wind noise, porous hoses have distinct frequency-dependent effects on signal amplitudes and by changing the phase response of the microphone can have deleterious effects on bearing accuracy. In this presentation the effects of the use of porous-hose wind-noise-reducing filters are discussed, both in terms of the effectiveness of noise reduction and the effects on signal characteristics.

11:00

5aPA10. Subsurface windscreen for the measurement of outdoor infrasound. Qamar A. Shams, Cecil G. Burkett, Toby Comeaux (NASA Langley Res. Ctr., 4 Langley Blvd., M.S. 238, Hampton, VA 23681), Allan J. Zuckerwar (Analytical Services and Mater., Hampton, VA 23666), and George R. Weistroffer (Virginia Commonwealth Univ., Richmond, VA 23284)

A windscreen has been developed that features two advantages favorable for the measurement of outdoor infrasound. First, the subsurface location, with the top of the windscreen flush with the ground surface, minimizes the mean velocity of the impinging wind. Secondly, the windscreen material (closed cell polyurethane foam) has a sufficiently low acoustic impedance (222 times that of air) and wall thickness (0.0127 m) to provide a transmission coefficient of nearly unity over the infrasonic frequency range (0–20 Hz). The windscreen, a tightly sealed box having internal dimensions of $0.3048\times0.3048\times0.3556$ m^3, contains the microphone, preamplifier, and a cable feed thru to an external power supply. Provision is made for rain drainage and seismic isolation. A three-element array, configured as an equilateral triangle with 30.48 m spacing and operating continuously in the field, periodically receives highly coherent signals attributed to emissions from atmospheric turbulence. The time delays between infrasonic signals received at the microphones permit determination of the bearing and elevation of the source, which correlate well with locations of pilot reports within a 320 km radius about the array. The test results are interpreted to yield spectral information on infrasonic emissions from clear air turbulence.

11:15

5aPA11. High-order parallel discontinuous Galerkin method for time-domain acoustic simulations. Timo Lähivaara (Dept. of Phys., Univ. of Kuopio, P.O. Box 1627, Kuopio, FI-70211, Finland, timo.lahivaara@uku.fi), Tomi Huttunen (Kuava Ltd., Kuopio, FI-70210, Finland), and Simo-Pekka Simonaho (Univ. of Kuopio, Kuopio, FI-70211)

The modeling of acoustic waves in the time-domain poses a significant challenge in scientific computing. A promising candidate for solving the three dimensional wave equation is the discontinuous Galerkin (DG) method. Advantages of the DG method are the easy parallelization and a special matrix structure which can reduce the overall time and the computer memory needed for solving the problem. In this study, a high order parallel DG method is investigated. The DG solver is implemented using the C++ programming language. Communication between processors of the parallel computer is performed using the message passing interface. In the solver, the polynomial degree of the basis functions is chosen individually for each element of the computation mesh (up to ninth order polynomials can be used). The unbounded problem is truncated using the perfectly matched layers. The method is evaluated with numerical simulations that are performed on a personal computer cluster.

FRIDAY MORNING, 14 NOVEMBER 2008 LEGENDS 7, 9:00 A.M. TO 12:00 NOON

Session 5aSC

Speech Communication: Second Language Perception and Production (Poster Session)

Kanae Nishi, Chair

Boys Town Natl. Research Hospital, 555 N. 30th St., Omaha, NE 68131

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbers papers will be at their posters from 10:30 a.m. to 12:00 noon.

5aSC1. Foreign-accented speech in noise. Jonathan Dalby (Dept. of Audiol. and Speech Sci, Indiana-Purdue Fort Wayne, 2101 E. Coliseum, Ft. Wayne, IN 46805 and Commun. Disord. Technol., Inc, Bloomington, IN 47404, dalbyj@ipfw.edu)

Studies of the perception of foreign-accented English speech have shown that native-speaking listeners are able to adapt quite quickly and effectively to the phonetic and phonological characteristics of non-native speech [Bradlow and Bent, 2003, XVth ICPS Proceedings] and can do so even when the accent is unfamiliar [Clark and Garrett, J. Acoust. Soc. Am. **116** (2004)]. Such studies raise the question of how much cognitive effort underlies this adaptation as well as questions about the robustness of the adaptation under less-than-optimal listening conditions. Rogers *et al.* (2004) have shown that the intelligibility of connected speech, even that from highly proficient non-native speakers, was degraded more than that of native speech when presented in noise. This study attempts to replicate that noise effect for isolated words. Two Spanish-speaking adults with differing oral English proficiency and two American English speakers recorded 50 CVC words from each of three lists that have been shown to be of equivalent perceptual difficulty (NU lists 1, 2, and 4). Words were presented with no noise and at -5 and -10 dB SNR to native listeners. Early results suggest that

word identification scores are similar to those found for sentence comprehension. CDT, Inc. markets a speech intelligibility training software. [Work supported by Purdue Research Foundation.]

5aSC2. Quantifying the contribution of contextual information in speech perception in noise for native and non-native listeners. Kanae Nishi, Jessica Lewis, Judy Kopun, and Patricia G. Stelmachowicz (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, nishik@boystown.org)

When speech is degraded, listeners tend to rely on contextual information, if available. The goal of the current study was to evaluate the use of lexical, syntactic, and semantic context for native and non-native listeners. Using an adaptive tracking method, listeners' reliance on contextual information was assessed as signal-to-noise ratios for 70%-correct speech perception performance (SNR70) for single words and short grammatically and semantically appropriate sentences. Adult and child native speakers of American English and adult Spanish speakers learning English as a non-native language served as listeners. For all groups, SNR70 was significantly higher for words than sentences. No group difference was found for sentence SNR70, but significantly higher word SNR70 was observed for the non-native group than the two native groups, indicating greater benefit of context for non-native listeners. When non-native listeners' SNR70 measures were subjected to discriminant analysis to determine similarity to native groups, three subgroups were identified: (1) nativelike for both words and sentences, (2) nativelike only for sentences, and (3) not nativelike for either words or sentences. Possible application of this method to determine candidacy for auditory training to improve speech perception in noise will be discussed. [Work supported by the NIH.]

5aSC3. Effects of speechreading on understanding mainstream American English by second-language listeners of English. Yori Kanekama and David Downs (Dept. of Commun. Sci. and Disord., Wichita State Univ., 1845 Fairmount St., Wichita, KS 67260-0075, yxkanekama@wichita.edu)

Cross-language speech perception research has shown that, at least across some languages, first-language speakers of English show more of a McGurk effect for English nonsense syllables than second-language speakers of English. This suggests less auditory-visual integration and auditory-visual incompatibility when listening to nonsense syllables of a different language. Researchers, however, have not studied the role of auditory-visual integration and auditory-visual incompatibility on cross-language perception of ongoing everyday speech. The purpose of this study was to measure the effects of speechreading on understanding mainstream American English (MAE) by second-language listeners of English. Specifically, participants were 30 first-language MAE speakers and 30 first-language Indian speakers recruited from the same graduate school engineering program. They participated in two experiments in which they listened to Central Institute for the Deaf Everyday Speech Sentences under auditory-only, visual-only, and auditory-visual conditions at different signal-to noise ratios. Visual enhancement and auditory enhancement scores were computed and compared between language groups. Results have (1) theoretical implications for understanding the role of auditory-visual integration and auditory-visual incompatibility on cross-language perception of connected speech, and (2) have practical implications for understanding whether speechreading helps or hinders listening to a second language in noisy environments.

5aSC4. The amount of information needed for listeners to detect a foreign accent. Hanyong Park, Kenneth de Jong (Dept. of Linguist., Indiana Univ., 404 Memorial Hall, 1021 E. 3rd St., Bloomington, IN 47405, hanypark@indiana.edu), and Isabelle Darcy (Indiana Univ., Bloomington, IN 47405)

This study examined how much information is needed for listeners to detect a foreign accent. Two factors were considered regarding the amount of information: stimulus length and *L*1 phonotactics. Four Korean-English bilinguals and two native speakers of American English produced different lengths, but still short stimuli: the vowel /ɑ/, monosyllabic and disyllabic English words. The monosyllabic corpus, in particular, included the stimuli having both natural (i.e., CV) and unnatural syllable structures (i.e., CCV, CVC, and CCVC) as well as various segments in terms of Korean phonotatics. After being presented with a stimulus, eight native listeners were asked to judge whether the speaker of the stimulus was a native or a non-native speakers of American English. The examination of d' value indicates that all the listeners detected a foreign accent from hearing the monosyllabic and disyllabic stimuli, but only some listeners did from hearing the vowel /ɑ/. Furthermore, the listeners detected a foreign accent more often from the stimuli having the coda segment. Lastly, d' values for the stimuli having /l/ were lower than those for other stimuli. All these suggest that listeners may not need much information to detect a foreign accent, which, in turn, is closely related to *L*1 phonotactics.

5aSC5. Patterns of cross subject correlation in second language learning: Skill structure and feature-level grouping in production and perceptual learning. Kenneth de Jong, Yen-Chen Hao, and Hanyong Park (Dept. of Linguist., Indiana Univ., 322 Memorial Hall, Bloomington, IN 47405)

This study examines the degree to which achievement of accuracy in distinguishing contrasts in one set of segments tends to correlate with achievement of accuracy in a related set of segments across listeners. Forty Korean listeners identified anterior obstruents as produced by four American English speakers before, after, and between instances of the vowel /a/. Accuracy rates for fricatives and stops, differing in voicing and in point of articulation, correlated across the listeners, indicating that some listeners were better at manner distinctions as a class. Similarly, accuracy in different voicing contrasts also correlated across listeners, though only when consonants were in the same prosodic location. Voicing accuracy did not correlate with manner accuracy, indicating that particular listeners were specifically good at particular featural contrasts. The productions of 20 Korean listeners of the same consonants in the same prosodic locations were identified by 10 American listeners. Even though the same range of accuracies was found for production as perception, accuracy for various segments typically did not correlate across talkers. These results suggest that, while perceptual learning tends to generalize across segments along featural lines, production learning is more specific to particular segments.

5aSC6. On the relationship between the perception and learning of Hindi voicing and place contrasts by native speakers of American English. James D. Harnsberger (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611)

Many non-native speech sounds are challenging to perceive and, ultimately, to acquire. Predicting specific learning outcomes from perceptual data has been hampered by (1) problems in quantifying acoustic-phonetic similarity between non-native and native sounds, (2) calculating the perceptual weighting of acoustic cues in the native language, and (3) the high-variability commonly observed in the perceptual assimilation of non-native contrasts to multiple native categories. This variability may reflect long-term persistent patterns in learning or it may represent only a brief early stage prior to fossilization. To examine these two possible accounts, the perceptual assimilation by 15 American English listeners of 7 voicing and place contrasts produced by 6 Hindi speakers was examined before and after training in a paired-associate word learning task. Training utilized tokens from four of the six talkers used in perceptual assimilation tasks. The purpose of the study was to determine whether or not assimilation patterns could be greatly modified and simplified by very limited experience in acquiring the contrasts via word learning. The results showed that limited laboratory training significantly reduced talker and token variability in perceptual assimilation and allowed for non-native contrasts to be categorized more cleanly in terms of their predicted difficulty in learning.

5aSC7. Acquiring novel perceptual categories in a third language: Bengali-English bilinguals' perception and learning of Malayalam consonants. Divya V. Gogoi (Program in Linguist., Univ. of Florida, Gainesville, FL 32611)

The present study examines the acquisition of novel non-native speech contrasts by adult bilingual speakers of Bengali and English. One of the underlying issues in this study is the role that phonetic features may play in the development of new phonetic categories. For instance, features utilized in native contrasts may generalize in the learning of novel non-native contrasts, even if they play a limited or no role in the initial perception of these non-native contrasts [Polka (1992); Harnsberger (1998)]. To explore this feature generalization hypothesis, a high variability consonant identification training

paradigm was used with ten Bengali-English bilinguals who learned to identify 4 Malayalam place of articulation contrasts (dental versus retroflex nasals, dental versus retroflex lateral approximants, palatoalveolar versus retroflex voiceless fricatives, and alveolar tap versus retroflex approximant). The bilingual Bengali-English listeners were selected for their extensive experience with the relevant place features (e.g., dental, retroflex, palatoalveolar, and alveolar), though Bengali and English lack any direct correspondents to the Malayalam contrasts. Bilingual performance was analyzed in terms of both patterns of perceptual assimilation as well as rate of acquisition of the novel non-native contrasts.

5aSC8. Tracking non-native acquisition of the Spanish tap-trill distinction: Cross-modal priming differences between native and non-native Spanish speakers. Wendy Herd and Joan Sereno (Linguist. Dept., Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044, wenherd@ku.edu)

English-speaking learners of Spanish often fail to achieve nativelike pronunciation of the tap-trill distinction in words like *caro* "expensive" and *carro* "car." The trill proves difficult because it is neither a phoneme nor an allophone in English. Although the tap exists as an allophone of /t/ and /d/ in American English, learners of Spanish must learn to process it as a phoneme rather than an allophone. Similarly, English learners of Spanish have difficulty acquiring the spirantization of voiced stops, i.e., /d/ spirantizes in *codo* "elbow," which occurs in the same environment as flapping. This study uses a cross-modal priming task to investigate whether L2 Spanish learners are able to process intervocalic tap, trill, /d/, and /t/ in the same way as L1 Spanish speakers. Using a cross-modal priming paradigm, eight English-speaking learners of Spanish were compared to eight native Spanish speakers. Results show that auditorily presented words with intervocalic taps resulted in faster response times for words like *cada* for learners of Spanish, but the same auditory stimuli resulted in faster response times for identically matching targets like *cara* for native Spanish speakers. These results suggest that cross-modal priming can be used to track L2 acquisition.

5aSC9. The English l sound produced by Korean students. Byunggon Yang (English Education Dept., Pusan Natl. Univ., 30 Changjundong Keumjunggu, Pusan 609-735, South Korea)

This study examined the lateral l sound produced by 16 Korean students in order to tap a possibility of using acoustical and perceptual criteria to distinguish lateral variants and eventually to assess student's English pronunciation skills. The subjects read a short story in a quiet office at normal speed. Those words with the lateral sound in onset or coda positions and before a vowel of the following word were analyzed using PRAAT. The following results are shown. First, the majority of the subjects produced the clear l regardless of the contexts. Some students produced the sound as the Korean flap or the English glide /r/. A few missing cases were also seen. Second, the dark l was mostly produced by the subjects of English majors in coda position with a few cases before a vowel in a phrase. Visual displays from the computer analysis were helpful in determining lateral variants but sometimes personal listening to the given sound after temporal manipulation would be necessary in the cases of fast and weak productions of the target words. Further studies would be desirable to compare native productions of the lateral sound with those of non-native speakers.

5aSC10. Effects of childhood exposure to a second language on the production of voice onset time and closure duration. Tetsuo Harada (Dept. of Education, Waseda Univ., 1-6-1 Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan, tharada@waseda.jp)

This study compares the production of voice onset time (VOT) and closure duration for singletons and geminates in Japanese by English-speaking university students who were exposed to Japanese in an immersion program in childhood and those regular university students who had no previous exposure to Japanese. 20 informants enrolled in a third-year Japanese were asked to repeat several target words including initial /p, t, k/ for VOT, and medials /p, t, k/ and /pp, tt, kk/ for singletons and geminates in a sentence frame. Both VOT of the initial stops and closure duration of the medial stops were measured. The results show that the immersion graduates' VOT values in Japanese were shorter (i.e., more Japanese-like) than those of the learners who had had no exposure to Japanese in childhood ($p<0.05$). For the production of closure duration, although the learners of Japanese without childhood L2 experience did not distinguish singletons from geminates, the immersion graduates did ($p<0.005$). The findings may suggest that long-term benefits of L2 experience in childhood in the naturalistic setting, which were found in the work of Knightly *et al.* (2003), may also apply to the instructional setting like immersion education.

5aSC11. A computer graphic three-dimensional tongue and lip movement synchronized with English fricatives for Japanese learners. Toshiko Isei-Jaakkola (Dept. of English Lang. and Culture, Chubu Univ., 1200 Matsumoto, Kasugai, Aichi 487-8501, Japan, tiseij@isc.chubu.ac.jp), Shigeki Suzuki (Tokyo Univ. of Social Welfare, Nagoya, Aichi 460 0002, Japan), Shigeo Morishima (Waseda Univ., Shinjuku, Tokyo 169-8555, Japan), and Keikichi Hirose (Univ. of Tokyo, Bunkyo, Tokyo 113-8656, Japan)

Some English fricatives are difficult specifically for Japanese learners of English (JL2) to produce. Simultaneous articulation of the lip and teeth (as in lebiodentals), or the tongue and teeth (as in dentals), or protruding lips (as in postalveolars) do not exist in standard Japanese. To understand partially or completely, invisible articulatory movements are unavoidable for JL2 in order to produce these fricatives properly. Thus, as an aid for the basic pronunciation training, a visualized automatic lip and tongue movement program synchronized with these fricatives was developed, utilizing three-dimensional computer graphic technologies. In this program, not only the lips, teeth, and tongue but also the other necessary speech organs were made to be half transparent. Consequently, it enables the learner to listen to and repeatedly model the target segmental sound while looking at these articulatory movements vividly from all kinds of angles. This system changes the phonetic practice situation from two-dimensional paper-based learning/teaching methods with audio-visual tools into revolutionary user-friendly method in phonetic class and outside the classroom. In addition, this system is applicable to any language sounds in the near future. [Work supported by JSPS and Chubu University Grant (A).]

5aSC12. Adapting second language phonemic perception training to common instructional situations: Initial results. Thomas R. Sawallis and Michael W. Townley (English Dept., Univ. of Alabama, Tuscaloosa, AL 35487, tsawalli@bama.ua.edu)

Although current L2 pedagogy de-emphasizes phoneme-level pronunciation training, laboratory experiments demonstrate benefits from training non-natives in perception of difficult target-language phonemic contrasts. Specifically, evidence shows that learners' perceptual performance improves (Jamieson Morosan, 1986; Flege, 1995), improvements generalize to new talkers and words (Lively, Logan, Pisoni, 1993), perceptual training triggers production improvements (i.e., without production training, Bradlow *et al.*, 1997), and both perceptual (Lively *et al.*, 1994) and production improvements (Bradlow *et al.*, 1999) are maintained over several months. These laboratory studies typically used intensive protocols, with long sessions, several days per week, for 2–3 weeks. We have adapted such protocols for use in common L2 instructional situations, using short sessions spread over a longer study period, and have begun training Japanese and other Asian students on the English /l-r/ contrast using this new regimen. This paper reports on initial results and some comparisons with studies using the earlier more intense protocols.

5aSC13. A nonsense syllable confusion matrix task in Korean-English bilingual children. Seok-Youn Yoon (Dept. of Speech Hearing Sci., Univ. of Ilinois, 405 Mathews, Urbana, IL 61801, yoon5@uiuc.edu), Cynthia J. Johnson, and Jont B. Allen (Univ. of Illinois, Urbana, IL 61801)

Compared to extensive studies in bilinguals in adults and very young children, very few studies have investigated speech perception in school-aged Korean-English bilingual (KEB) children. Given that English (L2) and Korean (L1) have different contrast systems among consonants and vowels and children have different speech perception processing than adults, school-aged KEB children are expected to have more complicated patterns of L2 perception. The main goal of the present study was to discern which set of L2 sounds are salient enough not to experience interference from L1. Ten KEB children at age 8–13 years were asked to identify 30 nonsense syllables of L2 sounds. The results showed that dissimilar consonants to Korean (i.e., /f,θ/) are most confusing to perceive than similar consonants with familiar contrasts (/p,k,b,g/) next to similar consonants with unfamiliar con-

trasts (/ʧ,ʤ,s,z/). More interestingly, perception performance in vowels showed the opposite pattern: a similar vowel (/ɛ/) was less correctly identified than unfamiliar vowels (/I,æ/). The present study suggests that L2 consonants, which cannot be mapped to any of the L1 sounds, are the most challenging to perceive. Nonetheless, groups of sounds that are patterned together cannot simply be explained with the L1 influence. For example, /f,θ/ are most confusing to KEB as well as English-monolingual listeners (Phatak & Allen, 2007); yet confusion of /b/ with /θ/ is unique only to KEBs.

5aSC14. An epenthetic vowel between consonantal sequences in perception and production by Japanese. Mieko Sperbeck and Winifred Strange (Dept. of Linguist., City Univ. of New York-Graduate School and Univ. Ctr., 365 Fifth Ave., New York, NY 10016, msperbeck@gc.cuny.edu)

Native Japanese have a strong tendency to epenthesize a vowel when producing consonantal sequences in English. This study investigated how Japanese learners of English perceive and produce word-initial CCV versus CəCV contrasts (e.g., sport versus support). Two types of tasks were employed: a categorial ABX task for perception and a delayed imitation task for production. Nonsense words of the form /C1C2ani/ (e.g., spani) and /C1əC2ani/ (e.g., sepani) served as the stimuli where C1C2 combinations were /s-p/, /s-t/, and /s-k/. In the ABX task, participants heard three short sentences that contained the target words and answered whether the third target word was the same as the first or the second one. In the delayed imitation task, participants heard the production of a native speaker in a carrier sentence (e.g., say sepani now) twice and produced the target word in isolation and then in the carrier sentence. Japanese participants made both perception and production errors. Interestingly, the majority of errors in the production task were vowel deletion for the CəCani contexts rather than vowel epenthesis for the CCani contexts. The relationship between perception and production among *L*2 learners as well as influence of *L*1/*L*2 phonotactic differences will be discussed.

5aSC15. Production and perception of English vowel categories by native Korean speakers. Ga Yeon Son (Univ. of Pennsylvania, 619 Williams Hall, 255 S. 36 St., Philadelphia, PA 19104, gson@babel.ling.upenn.edu)

This study deals with the production and perception of a second language (*L*2) by non-native speakers assessing phonetic convergence in their production and perceptual divergence. Production and perception of nine American English monophthongs and eight Korean monophthongs by the experienced and inexperienced Korean speakers were acoustically analyzed. The experiments consist of the production test, identification tests, and discrimination test. The experienced group showed relatively successful production of English vowels that have dissimilar acoustic properties with Korean vowels, while the inexperienced group showed complete phonetic interlingual identification. However, in perception, both groups successfully scored in the identification and discrimination tests, not showing perceptual identification between English and Korean. It suggests that two language systems mutually influence in one native phonetic space, and linguistic experience with *L*2 brings the reorganization of the phonological system, adding new phonemic categories for dissimilar *L*2 phones. The biased results in production and perception tests demonstrate that production of *L*2 is strongly related to the native phonemic categories, but perception is not. Therefore, the findings suggest that two different phonological systems are incorporated into the same phonetic space and linguistic experience with *L*2 boosts this process.

5aSC16. The influence of signal-to-noise ratio and listener's native language on vowel intelligibility. Tessa Bent and Diane Kewley-Port (Dept. Speech and Hearing, Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47401, tbent@indiana.edu)

In a previous study, intelligibility rankings of American English (AE) talkers were different for native and non-native listeners [Bent, *et al.*, J. Acoust. Soc. Am. **122**, 3016 (2007)]. To equate overall performance, the native and non-native listeners were tested at two different signal-to-noise ratios (SNRs). Therefore, the differences in intelligibility rankings between the two listener groups may have been affected by the listener's native language and differences in SNR. The current study explored this issue by testing across-talker differences in AE vowel intelligibility for AE listeners and Korean listeners learning AE as a second language. Listeners heard recordings of ten AE vowels in /bVd/ context produced by ten talkers under three SNRs: −8, −5, and −3 dB. The words were presented for identification in a 10-alternative forced choice task. Results showed that when AE and Korean listeners' were tested under the same SNR, the across-talker intelligibility scores were highly correlated for the two listener groups. However, Korean listeners were less accurate at vowel identification compared to the AE listeners. Furthermore, Korean listeners' intelligibility rankings of the talkers were more variable across SNRs whereas the AE listeners were more consistent. (Work supported by NIH DC02229 and T32-DC00012).

5aSC17. Exposure effects on production and perception development for the learners of English as a second language. Seokhan Kang (Dept. of Linguist. and Philosophy, MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Exposure effects for L2 production and perception development were examined. Two types of intervowel VCV stop voicing with different accents (iambus and trochee) were presented to ten English native speakers, ten English-learning Korean university students who had residence experience in America, and ten English-learning Korean university students who had no residence experience in America. Results showed that L2 patterns of Korean students' production and perception with residence experience in America were in between the English native speakers and the Korean students without residence experience. However, the asymmetrical development patterns between the production and perception for the two Korean groups could be found. In the production tests, the prosodic factor deeply influences on the asymmetrical development. The difference is in the trochee pattern rather than in the iambic pattern. It means that residence experience improves prosody knowledge in their production. Also this effect has some influence on both the temporal and fundamental frequency features but not on the formant features. The acoustic difference could be seen in the perception tests, too. We could not find any difference in the identification test of the word-medial iambic environment. Rather the difference could be found in the cue influence test in the trochaic environment.

5aSC18. Perceptual cues to English lexical stress: Comparison between native speakers and Mandarin second language learners. Yuwen Lai and Joan Sereno (Dept. of Linguist., Univ. of Kansas, 1541 Lilac Ln., Blake Hall, Lawrence, KS 66044)

The present study investigates the role of *F*0, duration, and vowel quality in English lexical stress perception by second language learners with a tonal L1. Mandarin speakers learning English as a second language (advanced learners, n=25; beginning learners n=25) were compared to a control group of native English speakers (n=25). Resynthesized disyllabic nonwords (dada) were presented in a stress localization task. Simultaneously manipulating five different *F*0 ratios (ratio of maximum fundamental frequency values of the first and second vowels) and five different duration ratios (ratio of duration values of the first and second vowels), stimuli were synthesized based on acoustic measurements from a previous study [Y. Lai and J. Sereno, J. Acoust. Soc. Am. **121**, 3071 (2007)]. Both full-vowel and reduced-vowel stimuli were used. The results indicated that full vowels most often attracted stress across all three listener groups. More interesting, beginning second language listeners relied mainly on duration cues to determine the stressed syllables and advanced listeners focused more on *F*0 cues, while native listeners made use of both cues. The findings will be discussed in terms of the similarities and differences between the prosodic systems of Mandarin and English.

5aSC19. Intonational interference on the perception and production of Mandarin Chinese tones by English-speaking second language learners. Yen-Chen Hao (Dept. of Linguist., Indiana Univ., 2451 E. 10th St. Bloomington, IN 47408, yehao@indiana.edu)

Most previous studies on *L*1 transfer focus on segments. The present study investigated the effect of *L*1 intonation on English speakers' acquisition of Mandarin Chinese tones. Four intermediate learners read three sentence types: declaratives, yes/no questions, and a list sequence, both in English and Chinese. Their productions showed Tone 3 (low) to Tone 2 (rising), and Tone 4 (falling) to Tone 1 (high) errors, as expected if productions have a (English) terminal rise imposed on them. Nonfinal items in reading a list exhibited similar errors. In the perception task, subjects listened to Chinese productions of the different sentence types and identified

the tone and whether the sentence is a statement or a question. When subjects identified a sentence as a question, they tended to judge the final tone to be Tone 2 (rising) or Tone 1 (high). Similarly, with final Tone 2, subjects usually called the sentence a question; and with final Tone 4, subjects often judged it as a declarative. English-speaking learners associate the rising tone with questions and falling tone with declaratives, even though they know the final word can carry any tone in Chinese. This study quantitatively documents an effect of intonation on the learning of lexical tones.

5aSC20. Identification of Mandarin coarticulated tones by inexperienced and experienced English learners of Mandarin. Yunjuan He and Ratree Wayland (Program in Linguist., Univ. of Florida, Box 115454, Gainesville, FL 32611-5454)

Two groups of native English speakers, relatively inexperienced (IE) ($N=14$) with 4 months of Mandarin study and relatively more experienced (EE) ($N=14$) with 12 months of study, were asked to identify coarticulated Mandarin lexical tones in disyllabic words. The results show that the EE group was better at identifying Mandarin tones than the IE group. Interestingly both groups were more accurate at identifying tones of the second (i.e., last) syllables than tones of the first syllables. Two types of errors were found in both groups: tonal direction misperception and tonal height misperception. EE committed fewer tonal direction errors than IE. However, EE still made considerable amount of tonal height errors. These results suggest the following: (1) The ability to perceive coarticulated tones improves with learning experience. (2) Due perhaps to a recency effect, final tones are remembered better than initial tones, and final syllable tones are misperceived less frequently than initial syllable tones. (3) The ability to identify tonal direction may improve faster than the ability to identify tonal height among English speakers and thus while tonal direction errors decrease with experience, misperception of tonal height remains even with increasing learning experience.

5aSC21. The perception of English lexical stress by native Thai speakers. Jirapat Jangjamras (Linguist. Program, Univ. of Florida, 4131 Turlington Hall, Gainesville, FL 32611-5454, jirapat@ufl.edu)

This study investigates the influence of *L*1 stress patterns and stress cues on the perception of *L*2 lexical stress. Thai and English were examined due to their prosodic differences such as the stress pattern and native perceptual correlates of stress. Thai, being a tone language, has a fixed stress pattern and employs the duration contrast as the primary stress cue. In contrast, English has a variable stress pattern and employs a combination of acoustic cues such as *F*0, duration, intensity contrast, and vowel reduction. Perceptual difficulties by listeners from a fixed-stress background have been reported in *L*2 stress discrimination and identification tasks. In this study, two groups of listeners, 35 Thai and 10 American English (AE), identified the stress location of disyllabic English nonwords produced by a trained phonetician using the above stress cues except vowel reduction. The preliminary results show similar mean scores between the two groups with a bias toward initial stress in misidentified tokens by AE listeners. Further analysis will report the reaction time analysis and the interaction between the syllabic structure and stress location.

FRIDAY MORNING, 14 NOVEMBER 2008 LEGENDS 12, 9:00 TO 11:15 A.M.

Session 5aSP

Signal Processing in Acoustics: Detection, Estimation, and Classification of Acoustic Signals

Angelo J. Campanella, Chair
Campanella Associates, 3201 Ridgewood Dr., Columbus, OH 43026-2453

Contributed Papers

9:00

5aSP1. Ocean surface degradation of shallow water acoustic communication. Geoffrey Edelmann (U.S. Naval Res. Lab., 4555 Overlook Ave. SW, Code 7145, Washington, DC 20375, geoffrey.edelmann@nrl.navy.mil), Shaun Anderson (Georgia Inst. of Technol., Atlanta, GA 30332-0405), and Paul Gendron (U.S. Naval Res. Lab., Washington, DC 20375)

The accurate prediction of underwater acoustic communication system performance at high frequencies is essential to system risk mitigation prior to deployment. The deterioration of acoustic coherence by ocean surface dynamics will be predicted via three dimensional (3-D) ray tracing with the inclusion of interaction with a realistic dynamic ocean surface. Acoustic communication performance limits at high frequencies for most fixed platform systems are dominated by multipath interaction with the spectrally rich heaving ocean. The high-fidelity Wave-watch III surface spectra model is used here with a wave-number integration approach for surface time series modeling. A fully 3-D ray tracing model is used to model the propagation of sound under a dynamic surface boundary at the time and spatial scales that dominate coherence degrading effects for underwater acoustic communications. The proposed method will serve the rapid development of operating regimes for high-frequency underwater acoustic communication systems. Proposed applications include distributed underwater sensor networks and persistent deployable low-cost systems that meet a wide range of underwater sensing needs. This work seeks to solve problems in deploying advanced communication systems by predicting the degradation of signals due to ocean variability and automatically selecting appropriate signaling rates, schemes, and constants. [This work was supported by ONR.]

9:15

5aSP2. Acoustic monitoring of severe weather in the Northeast Pacific Ocean. Jeffrey Nystuen (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105, nystuen@apl.washington.edu) and Svein Vagle (Fisheries and Oceans Canada, Sidney, BC V8L 4B2, Canada)

Wind and rainfall are the principal physical processes responsible for the production of high-frequency (1–50 kHz) ambient sound in the ocean. The primary source of the sound is the resonant ringing of individual bubbles created during wave breaking and raindrop splashes. Larger bubbles (>300 μm diameter) quickly return to the surface, while smaller bubbles can be mixed downward at several meters. During severe weather a layer of smaller ambient bubble forms and effectively absorbs higher-frequency (>10 kHz) sound. These processes are revealed in a two-year record of ambient sound recorded from a subsurface mooring at 50N, 145W in the NE Pacific Ocean as part of the Canadian SOLAS program. The passive acoustic signal of wind, rain, and ambient bubble clouds are compared to the subsurface mooring data, including data from an upward looking 200 kHz active sonar and a 300 kHz ADCP. The acoustic signatures of light, moderate, and heavy rainfall are superimposed on the signature of high wind, demonstrating rainfall detection even in the presence of high wind. [Work supported by ONR, Fisheries and Oceans Canada, and the Canadian CFCAS NSERC.]

9:30

5aSP3. Model based target depth estimation using received signal statistics. Colin Jemmott (Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804, cwj112@psu.edu), R. Lee Culver, Brett Bissinger, and Nirmal Bose (Penn State Univ., State College, PA 16804)

The overall goal of our research is to develop model based sonar signal processing techniques that utilize predictions of the acoustic field while being robust to environmental variability and uncertainty. Specifically, we are estimating the depth of a target broadcasting a tonal signal in shallow water using a single passive hydrophone. Three different techniques, naïve Bayes, histogram filter, and estimator correlator, are examined. Each is applied to event S5 of the SWellEX-96 measurement that took place in shallow water off the coast of California. The data consist of sinusoidal signals transmitted simultaneously from moving sources at two different depths and received at a bottomed horizontal array. Received signal amplitudes are modeled using random access memory for the two sources at different depths, and these modeling results are used in the design of the target depth estimation algorithms. Thus, the ability of each of the techniques to estimate target depth depends on an accurate representation of the uncertainty in the acoustic field due to propagation effects. Despite each technique requiring different assumptions, general conclusions about their effectiveness to solving this problem can be drawn. [Work supported by ONR Undersea Signal Processing, Code 321US.]

9:45

5aSP4. Bat-inspired distance measurement using phase information. Said Assous, Peter Jackson, Clare Hopper, David Gunn, John Rees, and Mike Lovell (Ultrasound Res. Lab., Dept. of Geology, Univ. of Leicester, LE1 7RH, UK)

This paper shows the use of phase measurement to estimate the distance to a target. Inspiration for this work comes from the observation that bats have been shown to have exceptional resolution with regard to target detection when searching during flight. Au and Simmons ["Echolocation in dolphins and bats," Phys. Today **45(7)**, 40–45 (2007)] concluded bats with a center frequency of about 80 kHz (i.e., 4-mm wavelength), and 40-kHz bandwidth can have a resolution of distance in air approaching 20 μm. For this frequency, we see that the resolution achieved by the bat is about 200 times better than $\lambda/2$ (i.e., 2 mm at this frequency), which is usually used as a guide for resolution for analog systems. Moreover, Au and Simmons show, using time-frequency analysis, that there are essentially two frequencies present at any particular time within a single bat pulse. Considering this use of two frequencies we may infer a distance. A new bat-inspired algorithm is presented. This is based on phase measurement and, when applied to underwater acoustics, shows that a resolution of 1/50 of the wavelength can be achieved in practice.

10:00—10:15 Break

10:15

5aSP5. Contrast models and classification of active sonar signals. Charles F. Gaumond, Colin W. Jemmott, and Derek Brock (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

New features and algorithms are derived from human perception to improve the automation of active sonar signal classification. Data from human subject research on the perceived similarity of sound pairs can be assembled in the form of a dissimilarity matrix. A dissimilarity matrix can be analyzed in many ways, two of which are multidimensional scaling (MDS) and additive clustering (AC). Different underlying perception models can be inferred from each analysis technique: a feature vector space for MDS and a contrast model for AC. The relevance of each technique to the dissimilarity data is inferred from the quantification of model complexity to the measure of fitting the original data [D. J. Navarro and M. D. Lee, "Common and distinctive features in stimulus similarity," Psychon. Bull. Rev. **11**, pp. 961–974 (2004)]. An experimental dissimilarity matrix is shown and analyzed using MDS and AC. Potential features for each case are shown from singular value decomposition of various signal representations. Signal processing implications of real valued features in the contrast model and alternative implementation are discussed. [The authors gratefully acknowledge conversations with and results from Dr. Jason Summers. Work supported by the Office of Naval Research.]

10:30

5aSP6. Development and analysis of chirplet signal decomposition for detection, estimation, and classification in sonar and nondestructive evaluation (NDE) applications. Yufeng Lu (Dept. of Elec. and Comput. Eng., Bradley Univ., 1501 West Bradley Ave., Peoria, IL 61625, ylu2@bradley.edu), Erdal Oruklu, and Jafar Saniie (Illinois Inst. of Tech., Chicago, IL 60616)

In sonar and ultrasonic nondestructive evaluation applications, backscattered signal contains information pertaining to size, shape, and orientation of reflectors. The echoes from target reflectors are often overlapping and masked in the presence of high clutter. Therefore, it is a challenging problem for target localization and object recognition. In this paper, a chirplet model for an adaptive signal decomposition and parameter estimation algorithm is investigated. The proposed chirplet method is capable of representing a broad range of echo shapes, including narrow-band, broad-band, symmetric, skewed, and nondispersive/dispersive. Chirplet transform is used not only as a means for time-frequency representation but also to estimate the echo parameters, including amplitude, time-of-arrival, center frequency, bandwidth, phase, and chirp rate. These parameters embody the dispersion effect, frequency dependent absorption, and scattering in complex and inhomogeneous environments. Experimental data, including ultrasonic backscattered signal and bat chirp signal, have been analyzed to evaluate the algorithm. It has been shown that the adaptive algorithm performs robustly, yields accurate echo estimations, and results in considerable SNR enhancements. Furthermore, system-on-a-chip implementation based on FPGA has been developed to demonstrate the feasibility of real-time applications. This type of study addresses a broad range of applications, including pattern recognition, deconvolution, target sizing, and characterization.

10:45

5aSP7. An image processing based neural network method of wave form classification. Jeffrey Vipperman and Brian Bucci (648 Benedum Hall, 3700 O'Hara St., Pittsburgh, PA 15261, jsv@pitt.edu)

In an effort to identify military impulse noise, as it relates to civilian damage and disturbance claims, several metric based approaches utilizing artificial neural networks and Bayesian classifiers have proven successful in addressing this issue. However, in the course of research, it became apparent that the noise sources to be classified, namely, various types of military impulse noise, wind noise, and aircraft noise, could be easily identified by a minimally trained observer by way of a simple visual inspection of the wave form. Additionally, since the noise classification algorithm is desired to be implemented on DSP boards with possibly limited computational resources, it may prove beneficial to avoid the computation of metrics which involve complex mathematical operations. Borrowing from proven artificial neural network techniques already proven in the field of optical character recognition, this proposed noise classifier views a captured wave form as a number of points located on a spatial grid. The density of points within each grid sector is then used as input to an artificial neural network. The resulting classifiers performed with accuracies up to 0.997 on testing data. [This research was supported by the U.S. Department of Defense, through the Strategic Environmental Research and Development Program (SERDP).]

11:00

5aSP8. A data fusion and multiple ping method for improving the resolution of low-power acoustic and seismic sensing. Alexander Apartsin and Nathan Intrator (School of Comput. Sci., Tel-Aviv Univ., Ramat Aviv 69978, Israel)

Biological research has shown that some mammals use low-power seismic pulses for underground communications and navigation. For instance, the blind mole rat bangs its head against the walls of its underground tunnel to produce seismic signals. It has been demonstrated that this animal builds a three-dimensional map of the environment by collecting and analyzing

returned responses. This behavior enables the mole rat to avoid obstacles and voids while it digs its tunnels. Being able to use low-power signals for accurate remote sensing under low signal to noise ratio (15–25 dB) conditions has some important applications (e.g., underground installment detection). A practically used method for coping with noise is based on repeated transmission of identical pulses and averaging the peak value of autocorrelation function from multiple measurements. Our research suggests that by carefully selecting a pulse family and a fusion method of multiple measurements, one can significantly improve sensing resolution. In particular, our simulation results show that employing robust statistics fusion method and by producing sonar pings with a shifting phase, the average error rate of travel time estimates can be reduced by about 30%.

FRIDAY MORNING, 14 NOVEMBER 2008 LEGENDS 8, 9:00 TO 11:15 A.M.

Session 5aUW

Underwater Acoustics: Deep Water Propagation

Michael G. Brown, Cochair
Univ. of Miami, RSMAS-AMP, 4600 Rickenbacker Cswy., Miami, FL 33149-1098

Wendy Saintval, Cochair
Rensselaer Polytechnic Inst., Math Sci., 110 8th St., Troy, NY 12180

Contributed Papers

9:00

5aUW1. Mode coupling induced by random ocean sound speed structure: Mean intensity and revisiting the evolution equations for the cross-mode coherences. John Colosi (Dept. of Oceanogr., Naval Postgrad. School, Monterey, CA 93943) and Andrey Morozov (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The seminal papers of Dozier and Tappert (1978), Dozier (1983), and Creamer (1996) have analyzed mode coupling induced by a random ocean sound speed structure to derive transport equations for the range evolution of mean mode energy. A major assumption in this work is that the cross-mode coherences rapidly decay with range and can thus be ignored. Theoretical predictions of mode energy have been compared to a direct Monte Carlo simulation, yielding favorable results [Dozier and Tappert (1978) and Creamer (1996)], so the general accuracy of the transport equations cannot be questioned. However, if the observable of interest is the mean acoustic intensity, one will quickly realize that the cross-mode coherences do not decay rapidly with range and thus need to be estimated for accurate acoustic predictions; an example from deep ocean propagation will be provided. This paper presents a theoretical framework for computing cross-mode coherences and demonstrates why the coherences are not important for the evolution of mode energy.

9:15

5aUW2. Perth-Bermuda (1960) revisited global scale acoustic propagation. Kevin D. Heaney and James J. Murray (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039)

A recent paper in Geophysical Research Letter [Dushaw, Geophys. Res. Lett. **35**, L08601 (2008)] commented on the Perth-Bermuda Sound Propagation Experiment (1960): An Adiabatic Mode Interpretation JASA 95, November 1991, by Kevin D. Heaney, William S. Kuperman, and B. E. McDonald. Dushaw raises questions about the validity of the approach and conclusions of Heaney *et al.* In particular, the validity of the adiabatic mode assumption for long-range propagation in the presence of internal wave scattering and seafloor interaction is challenged. Dushaw finds no other explanation for the observed double pulse arrival in the original 1960 experiment. The approach and conclusions of Heaney *et al.* will be presented. The range of validity of the adiabatic mode approximation in global scale acoustics will be addressed. The original computations of Heaney *et al.* were for a single mode at 10 Hz due to the computational power available at the time. Broadband results (10–50 Hz) using multiple modes summed coherently will be presented.

9:30

5aUW3. Resonant forward scattering of sound in deep ocean environments. Irina Rypina (PO Dept., WHOI, Clark Bldg., Woods Hole, MA 02543), Ilya Udovydchenkov (WHOI, Woods Hole, MA, 02543), and Michael Brown (Univ. of Miami, Miami, FL, 33149)

Resonant forward scattering of sound in deep ocean environments occurs when a ray cycle (double loop) distance is equal to the wavelength of a spectral component of the sound speed perturbation field or, more generally, when these quantities are rationally related. In the presence of a sound speed perturbation field with a broad spectrum (e.g., an internal-wave-induced perturbation field), many resonances are excited and these can be expected to overlap everywhere. In spite of this complexity, there are theoretical reasons that lead to the expectation that forward scattering, in general, and mode coupling, in particular, are largely controlled by resonant scattering. Relevant theory will be briefly described and numerical simulations (ray, mode, and PE) will be presented to illustrate important concepts. [Work supported by ONR.]

9:45

5aUW4. Low mode-number scattering of transient signals in deep ocean. Ilya Udovydchenkov (AOPE Dept., WHOI, 98 Water St., Woods Hole, MA 02543, ilya@whoi.edu), Michael Brown (Univ. of Miami, Miami, FL 33149), and Timothy Duda (WHOI, Woods Hole, MA 02543)

A modal description of the propagation of transient signals in deep ocean environments is considered. The environment is assumed to consist of a range-independent background sound channel with a single sound speed minimum on which a structured range-dependent perturbation, due, for example, to internal waves, is superimposed. Mode coupling in such environments is predominantly local in mode number, leading, to an excellent approximation, to the diffusive spreading of energy in mode number with increasing range. Because mode number is non-negative, however, the simple diffusive spreading of energy in mode number requires modification for low mode numbers (corresponding to the near axial sound energy). Simulations of distributions of energy in (low) mode number and time are presented and compared to estimates based on the theoretical work of Virovlyansky *et al.* [J. Acoust. Soc. Am. **121**, 2542–2552 (2007)]. The results presented are relevant to the analysis of measurements made during the recent long-range acoustic propagation experiment. [Work supported by ONR.]

10:00—10:15 Break

10:15

5aUW5. Vertical profiling of ambient noise with Deep Sound. David Barclay, Fernando Simonet, and Michael Buckingham (Marine Physical Lab, Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, dbarclay@mpl.ucsd.edu)

Deep Sound is a free-falling high-bandwidth acoustic recording system designed to profile ambient noise from the surface to depths of 9 km. The recording platform is autonomous and descends under gravity to its preprogramed maximum depth, where a burn-wire releases weight, permitting the system to return to the surface under its own buoyancy. Two hydrophones are mounted at half meter vertical spacing allowing the noise spectrum and vertical coherence (directionality) to be obtained over four decades of frequency (10 Hz–100 kHz). The acoustic recordings are made continuously as the instrument descends and ascends along with measurements of sound speed and depth. The system's low power and large data storage capabilities allow round trips to the deepest trenches of the ocean. Alternative modes of operation include (1) synthetic aperture signal detection and (2) residence on the seabed with return to the surface at a later time. Deep Sound's design and acoustic characteristics will be described and data from initial shallow water deployments around San Diego will be reported. [Work supported by the Office of Naval Research.]

10:30

5aUW6. Deep shadow zone arrivals on the ocean bottom. Ralph Stephen (WHOI, 360 Woods Hole Rd., Woods Hole, MA 02543), Matthew Dzieciuch, Peter Worcester (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238), Rex Andrew, Bruce Howe, James Mercer (Univ. of Washington, Seattle, WA 98105-6698), and John Colosi (Naval Postgrad. School, Monterey, CA 93943)

To gain insight into the physical mechanisms responsible for deep shadow zone arrivals in long range ocean acoustic propagation we have characterized some key features of the replica correlated signals observed on an ocean bottom seismometer (OBS) deployed on the 2004 long-range ocean acoustic propagation experiment. We compare the vertical seismometer data from an OBS at 4973-m depth beneath the deep vertical line array (DVLA) with the deepest hydrophone on the DVLA at 4270 m depth. The results of the preliminary analysis show (1) that the vertical geophone channel has more arrivals than the deep DVLA hydrophone channels, (2) that the first arrivals on the OBS geophone and DVLA deep hydrophones correspond to energy in the first deep arriving path predicted by the parabolic equation solution, (3) that the later arrivals which are much larger in amplitude are not explained by the parabolic equation method, and (4) that it is the later arrivals that contribute to the long range (up to 3200 km) shadow zone "receptions." Possible mechanisms for the later arrivals and some consequences of the observation are discussed. [This work was supported by the Office of Naval Research.]

10:45

5aUW7. Modeling the three-dimensional field of a seismic airgun array and comparison to 2003 measured data. Arslan M. Tashmukhambetov (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, atashmuk@uno.edu), Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

The full three-dimensional field of a seismic airgun array is modeled using an enhanced parabolic equation run on a parallel cluster computer system, which is part of the Louisiana Optical Network Initiative network cluster. Source signatures are calculated using GUNDALF and NUCLEUS softwares. The calculated results are compared to available measurements collected in Green Canyon in the northern Gulf of Mexico in 2003. Three-dimensional maps showing angular variation (both emission and azimuthal angles) and range dependence are generated, which show peak pressures, sound exposure levels, total shot energy spectra, and one-third octave band analyses. [Research supported by the International Association of Oil and Gas Producers and the International Association of Geophysical Contractors.]

11:00

5aUW8. The source characterization study 2007: A detailed three dimensional acoustic field measurement of a seismic airgun array. Arslan M. Tashmukhambetov, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, atashmuk@uno.edu), Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA), Joal J. Newcomb (NAVOCEANO, Stennis Space Ctr., MS), James M. Stephens, and Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS)

In September 2007 the Littoral Acoustic Demonstration Center (LADC) collected acoustic and related data from three moored arrays and ship-deployed hydrophones spanning the full water column to measure the three-dimensional acoustic field of a seismic airgun array. The seismic source vessel shot a series of lines to give a detailed angle and range information concerning the field. The data were collected in the western Gulf of Mexico between the East Break and Alamos Canyon regions. Peak pressures, sound exposure levels, total shot energy spectra, and one-third octave band analyses are measures used to characterize the field. Three dimensional maps of these quantities are generated to show dependence on emission and azimuthal angles and range. Both the direct and indirect fields are characterized. Moveout analysis is done to delineate arrivals and to detect ducted and interface waves. [Research supported by the International Association of Oil and Gas Producers.]