

ACOUSTICAL NEWS-USA		1255
USA Meeting Calendar		1255
ACOUSTICAL NEWS-INTERNATIONAL		1261
Standards Meeting Calendar		1261
BOOK REVIEWS		1265
REVIEWS OF ACOUSTICAL PATENTS		1267
LETTERS TO THE EDITOR		
Comment on "The directionality of acoustic T-phase signals from small magnitude submarine earthquakes" [J. Acoust. Soc. Am. 119, 3669–3675 (2006)] (L)	Delwayne R. Bohnenstiehl	1293
Reply to "Comment on 'The directionality of acoustic T-phase signals from small magnitude submarine earthquakes' [J. Acoust. Soc. Am. 119, 3669–3675 (2006)]" (L)	N. Ross Chapman, R. Marrett	1297
A pseudo-inverse algorithm for simultaneous measurements using multiple acoustical sources (L)	Ning Xiang, Shu Li	1299
GENERAL LINEAR ACOUSTICS [20]		
Characteristics of penalty mode scattering by rigid spllices in lined ducts	WenPing Bi, Vincent Pagneux, Denis Lafarge, Yves Aurégan	1303
Finite element analysis of guided waves in fluid-filled corrugated pipes	Matthias Maess, Jan Herrmann, Lothar Gaul	1313
NONLINEAR ACOUSTICS [25]		
Radiation force and shear motions in inhomogeneous media	Lev Ostrovsky, Alexander Sutin, Yuri Il'inskii, Oleg Rudenko, Armen Sarvazyan	1324
Noncontact transportation in water using ultrasonic traveling waves	Shinfuku Nomura, Thomas J. Matula, Jun Satonobu, Lawrence A. Crum	1332
AEROACOUSTICS, ATMOSPHERIC SOUND [28]		
Experiments on sound generation in corrugated pipes with flow	Ulf R. Kristiansen, Geir A. Wiik	1337
Vortex sound due to a flexible boundary backed by a cavity in a low Mach number mean flow	S. K. Tang, R. C. K. Leung, R. M. C. So	1345
A computational method of evaluating noncompact sound based on vortex sound theory	T. Takaishi, M. Miyazawa, C. Kato	1353

CONTENTS—Continued from preceding page

UNDERWATER SOUND [30]

Multipath pulse shapes in shallow water: Theory and simulation	Chris H. Harrison, Peter L. Nielsen	1362
Propagation in an elastic wedge using the virtual source technique	Ahmad T. Abawi, Michael B. Porter	1374
Gray whale target strength measurements and the analysis of the backscattered response	Irena Lucifredi, Peter J. Stein	1383
Miniaturized mandrel-based fiber optic hydrophone	Gregory H. Ames, Jason M. Maguire	1392

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

Aberrations in materials with random inhomogeneities	J. A. Hernández, M. Clark, S. D. Sharples, M. G. Somekh, V. H. Lopez	1396
Reproducible cavitation activity in water-particle suspensions	Bram M. Borkent, Manish Arora, Claus-Dieter Ohl	1406
Parallel capillary-tube-based extension of thermoacoustic theory for random porous media	Heui-Seol Roh, Richard Rasset, Henry E. Bass	1413
Near-field ultrasound tomography	Serge Mensah, Emilie Franceschini	1423
Acoustic power calibration of high-intensity focused ultrasound transducers using a radiation force technique	Subha Maruvada, Gerald R. Harris, Bruce A. Herman, Randy L. King	1434

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Elastic waves excited by a plane source on the surface of a multilayered medium	Bixing Zhang	1440
Simulation of guided waves in complex piping geometries using the elastodynamic finite integration technique	Kevin E. Rudd, Kevin R. Leonard, Jill P. Bingham, Mark K. Hinders	1449

NOISE: ITS EFFECTS AND CONTROL [50]

A moving zone of quiet for narrowband noise in a one-dimensional duct using virtual sensing	Cornelis D. Petersen, Anthony C. Zander, Ben S. Cazzolato, Colin H. Hansen	1459
--	--	------

ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION [58]

Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution	Paul Dickens, John Smith, Joe Wolfe	1471
Calibration sphere for low-frequency parametric sonars	Kenneth G. Foote, David T. I. Francis, Philip R. Atkins	1482

ACOUSTIC SIGNAL PROCESSING [60]

Imaging with unfocused regions of focused ultrasound beams	Roger Zemp, Michael F. Insana	1491
Spectrum analysis of seismic surface waves and its applications in seismic landmine detection	Mubashir Alam, James H. McClellan, Raymond R. Scott, Jr.	1499
Separation of components from impulses in reassigned spectrograms	Sean A. Fulop, Kelly Fitz	1510
A Bayesian method to estimate the depth and the range of phonating sperm whales using a single hydrophone	Christophe Laplanche	1519
Toward bias minimization in acoustic feedback cancellation systems	Christos Boukis, Danilo P. Mandic, Anthony G. Constantinides	1529

CONTENTS—Continued from preceding page

A full vectorial contrast source inversion scheme for three-dimensional acoustic imaging of both compressibility and density profiles	Koen W. A. van Dongen, William M. D. Wright	1538
Statistically optimized near field acoustic holography using an array of pressure-velocity probes	Finn Jacobsen, Virginie Jaud	1550
Nearfield binaural synthesis and ambisonics	Dylan Menzies, Marwan Al-Akaidi	1559
PHYSIOLOGICAL ACOUSTICS [64]		
Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions	Christopher A. Shera, Arnold Tubis, Carrick L. Talmadge, Egbert de Boer, Paul F. Fahey, John J. Guinan, Jr.	1564
Examination of bone-conducted transmission from sound field excitation measured by thresholds, ear-canal sound pressure, and skull vibrations	Sabine Reinfeldt, Stefan Stenfelt, Tobias Good, Bo Håkansson	1576
Bias due to noise in otoacoustic emission measurements	Bradford C. Backus	1588
PSYCHOLOGICAL ACOUSTICS [66]		
Modeling binaural loudness	Brian C. J. Moore, Brian R. Glasberg	1604
Hearing loss from interrupted, intermittent, and time varying Gaussian noise exposures: The applicability of the equal energy hypothesis	Wei Qiu, Bob Davis, Roger P. Hamernik	1613
Across-frequency pitch discrimination interference between complex tones containing resolved harmonics	Christophe Micheyl, Andrew J. Oxenham	1621
Illusory percepts from auditory adaptation	Lucas C. Parra, Barak A. Pearlmutter	1632
Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration	Julie Arenberg Bierer	1642
Quantifying the effects of fast-acting compression on the envelope of speech	Michael A. Stone, Brian C. J. Moore	1654
SPEECH PRODUCTION [70]		
Word-internal versus word-peripheral consonantal duration patterns in three languages	Melissa A. Redford	1665
Noise estimation in voice signals using short-term cepstral analysis	Peter J. Murphy, Olatunji O. Akande	1679
SPEECH PERCEPTION [71]		
Effect of spectral frequency range and separation on the perception of asynchronous speech	Eric W. Healy, Sid P. Bacon	1691
Effects of stimulus variability and adult aging on adaptation to time-compressed speech	Julie D. Golomb, Jonathan E. Peelle, Arthur Wingfield	1701
Evaluation of companding-based spectral enhancement using simulated cochlear-implant processing	Andrew J. Oxenham, Andrea M. Simonson, Lorenzo Turicchia, Rahul Sarpeshkar	1709
SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]		
Sinusoidal modeling for nonstationary voiced speech based on a local vector transform	Masashi Ito, Masafumi Yano	1717

CONTENTS—Continued from preceding page

MUSIC AND MUSICAL INSTRUMENTS [75]

- Sound pressure level and spectral balance linearity and symmetry in the *mesa di voce* of female classical singers Sally Collyer, Pamela J. Davis, C. William Thorpe, Jean Callaghan 1728

BIOACOUSTICS [80]

- Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California Marie A. Roch, Melissa S. Soldevilla, Jessica C. Burtenshaw, E. Elizabeth Henderson, John A. Hildebrand 1737
- Echo-intensity compensation in echolocating bats (*Pipistrellus abramus*) during flight measured by a telemetry microphone Shizuko Hiryu, Tomotaka Hagino, Hiroshi Riquimaroux, Yoshiaki Watanabe 1749
- Nonlinear acoustics in the pant hoots of common chimpanzees (*Pan troglodytes*): Vocalizing at the edge Tobias Riede, Adam Clark Arcadi, Michael J. Owren 1758
- The perception of complex tones by a false killer whale (*Pseudorca crassidens*) Michelle M. L. Yuen, Paul E. Nachtigall, Marlee Breese, Stephanie A. Vlachos 1768
- Bottlenose dolphin (*Tursiops truncatus*) steady-state evoked responses to multiple simultaneous sinusoidal amplitude modulated tones James J. Finneran, Dorian S. Houser 1775
- Comodulation detection differences in the hooded crow (*Corvus corone cornix*), with direct comparison to human subjects Kenneth Kragh Jensen 1783
- Singular spectrum analysis applied to ultrasonic detection and imaging of brachytherapy seeds Jonathan Mamou, Ernest J. Feleppa 1790

ERRATA

- Erratum: "Constrained comparison of ocean waveguide reverberation theory and observations" [J. Acoust. Soc. Am. 120(4), 1922–1931 (2006)] Charles W. Holland 1802

JASA EXPRESS LETTERS

- Interaural time difference processing of broadband and narrow-band noise by inexperienced listeners William A. Yost, Raymond H. Dye, Stanley Sheft EL103
- Acoustic micro-Doppler radar for human gait imaging Zhaonian Zhang, Philippe O. Pouliquen, Allen Waxman, Andreas G. Andreou EL110
- Ultrasonic wave generation due to human footsteps on the ground Alexander Ekimov, James M. Sabatier EL114
- Acoustic measures of low-frequency noise in extended high-frequency audiometry Nicolas Schmuziger, Matthias Brechbuehl, Rudolf Probst EL120

- CUMULATIVE AUTHOR INDEX** 1805

- TECHNICAL PROGRAM SUMMARY** 1809

Interaural time difference processing of broadband and narrow-band noise by inexperienced listeners

William A. Yost, Raymond H. Dye, and Stanley Sheft

Parmlly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626
wyost@luc.edu, rdye@luc.edu, ssheft@luc.edu

Abstract: Recent functional magnetic resonance imaging (fMRI) data might be interpreted as being in disagreement with existing psychophysical data regarding the laterality of broadband noise stimuli presented with large interaural time differences (ITDs). This study investigated the possibility that lateral judgments made by inexperienced listeners who did not receive feedback might be different than those reported for experienced listeners, especially when the ITD is longer than that occurring in nature, and therefore data from inexperienced listeners presented unnaturally long ITDs for the first time might be more consistent with the possible interpretation of the fMRI results. The results from this study using inexperienced listeners were not basically different from those reported in the literature based on experienced listeners, suggesting a possible difference does exist between inferences drawn from fMRI data and human psychophysical results.

© 2007 Acoustical Society of America

PACS numbers: 43.66.Pn, 43.66.Qp [QJF]

Date Received: October 12, 2006 **Date Accepted:** December 19, 2006

1. Introduction

Aspects of the classical coincidence-detection network originally proposed by Jeffress (1948) for processing interaural time differences (ITD) have recently been challenged based on observations of ITD processing in the inferior colliculus (IC). McAlpine and colleagues (see McAlpine *et al.*, 2001) suggest that a Jeffress-style coincidence network does not appear to exist in the IC of many animals (e.g., small mammals) and they have proposed an optimal population-based scheme (Harper and McAlpine, 2004) for processing ITD by all animals including humans. Thompson *et al.* (2006) recently published fMRI data recorded from regions of the IC in humans that appear consistent with observations of ITD processing in the IC of small mammals but could be interpreted as being inconsistent with published human psychophysical data, which can be accounted for by a Jeffress-style model. Thompson *et al.* (2006) suggest that models other than a Jeffress-style coincidence network might be required to account for ITD processing in humans and many other animals.

Thompson *et al.* (2006) used conditions in which relatively broadband noise (400-Hz wide) was presented with a whole-waveform ITD that was less than ($500 \mu\text{s}$) or greater than ($1500 \mu\text{s}$) a half period ($\pi = 1000 \mu\text{s}$) of the noise's center frequency (CF = 500 Hz). For $\text{ITD} < \pi$ of CF the IC on the side of the head opposite the ear receiving the first sound is activated, while for $\text{ITD} > \pi$ of CF the opposite IC is activated. If activation of the IC in each hemisphere is important in determining the laterality of images produced by varying ITD, this might suggest that broadband signals with small ITDs would be lateralized on one side of the head and those with large ITDs (relative to π at CF) on the other side. In human psychophysical experiments using similar broadband stimuli (see Trahiotis and Stern, 1989) the data suggest that broadband signals are always lateralized on the side of the head of the ear receiving the leading sound even for very large ITDs. This observation can be accounted for by a Jeffress-style model designed to account for psychoacoustical data based on interaural cross correlation as the mathematical operation of a coincidence network and other assumptions (see Trahiotis *et al.*, 2001),

Table 1. Stimulus and condition values for the three CF groups.

Conditions	200-Hz CF group	400-Hz CF group	800-Hz CF group
ITDs (μ s)	0, 417,834,1688, 3333, 4176	0, 417,834,1688,2085	0,209,417,834,1034
ITD μ s for π (π)	2500	1250	625
Bandwidths (Hz)	195–205 (10), 187.5–212.5(25), 162.5–237.5 (75), 112.5–287.5 (175), 12.5–387.5 (375)	395–405 (10), 387.5–412.5 (25), 362.5–437.5 (75), 287.5–512.5 (225), 62.5–737.5 (675)	795–805 (10) 775–825 (50) 725–875 (150) 462.5–1137.5 (675) 125–1475 (1350)
No. of trials/condition/listener	20; 40; 20	20; 40; 20	20; 40; 20
L-R; MOVE; P-D tasks			
No. of subjects	19	22	21

but appears inconsistent with the functional interpretation of the IC recordings of Thompson *et al.* (2006) as explained above. Jeffress-style cross-correlation models and IC recordings and their functional interpretation are compatible with the psychophysical data showing that narrow-band stimuli with $\text{ITD} < \pi$ are located toward the leading ear, but for $\text{ITD} > \pi$ the sounds are lateralized on the opposite side of the head from the leading ear (see Trahiotis and Stern, 1989; Yost, 1981). Thus, for narrow-band noise, the interpretation of the recordings from the IC might be consistent with the psychophysical data, but there is a possible lack of consistency for broadband noise.

The large ITDs used in the Thompson *et al.* (2006) study and in many psychophysical experiments do not occur in nature as they are considerably larger than the ITD based on the diameter of a human head and the speed of sound (between 650 and 800 μ s; see Kuhn, 1987). It has been shown that ITD processing can be variable in that the lateral location of images based on ITD (e.g., Yost, 1981) and ITD discrimination thresholds (McFadden *et al.*, 1973) exhibit large individual differences, and for some conditions it takes practice before asymptotic ITD performance is reached (see Wright and Fitzgerald, 2001). In addition, one of the arguments concerning ITD processing in the IC is that it appears limited to ITDs near π at CF, that is, usually relatively small ITDs. Sounds presented with ITDs over headphones can vary in the perceived lateral location and in the diffuseness of the lateralized image (see Blauret, 1997). It is possible that lateral images produced with large ITDs are more diffuse than those produced with smaller ITDs and, as a consequence, untrained listeners might have more difficulty than experienced listeners in assigning a lateral location to images produced with large ITDs. Finally, the subjects in the fMRI experiments had limited experience with long ITDs. This paper investigates the possibility that untrained/inexperienced listeners performing without feedback might indicate a different pattern of results for unnaturally large ITDs than that typically obtained for trained listeners who are provided feedback. That is, data from untrained listeners experiencing unnaturally long ITDs for the first time might be more consistent with the interpretation of IC results from Thompson *et al.* (2006) than the data from studies involving highly trained listeners who receive feedback of one sort or another regarding ITD processing, especially regarding long ITDs.

2. Methods

Four groups of 19–25 undergraduate students were used. Each group performed a different set of conditions for about an hour without receiving any training or feedback, and several conditions were randomized within a block of trials. These procedures helped ensure that listeners were not trained to respond with any particular bias or pattern.

Three groups (200-, 400-, 800-Hz CF groups) participated in three tasks: left-right lateralization task-L-R, lateral movement task-MOVE, and diffuse/punctate judgment task-D-P. ITD and bandwidth were randomized within each block of trials (see Table 1 for conditions). In

Table 2. Stimulus and condition values for the MIXED group.

Conditions	MIXED group				
ITDs (μ s)	0, 208, 833, 3333.3				
CFs (Hz)	200, 400, 800				
IPD (deg)	CF ITD (μ s)=	0	208	833	3333.3
	200 Hz	0	15	60	240
	400 Hz	0	30	120	480
(IPD $re: 2\pi$)					(120)
	800 Hz	0	60	240	960
(IPD $re: 4\pi$)					(240)
Bandwidths (Hz)	10-Hz wide and 400-Hz wide, see Table 1				
No. of trials/condition/listener,	50				
L-R task					
No. of subjects	25				

the L-R task, listeners indicated whether the noise appeared on the left or right side of the head. In the MOVE task, the first interval of each trial contained a stimulus with an ITD followed 200 ms later in the second interval with a stimulus with a larger ITD (except in one condition when the larger ITD was followed by a smaller ITD). In all cases the listener indicated if the sound image appeared to move left or right between the first and second interval. In the P-D task listeners were to indicate if the image was punctate or diffuse. At the beginning of each block of trials, the listeners were presented samples of diotic noises as examples of a punctate image and interaurally uncorrelated noises as examples of a diffuse image, with the concept of a diffuse image explained before data collection. No feedback was provided for any task.

The fourth group (MIXED) performed a left-right lateralization task (L-R task) but with ITD, bandwidth, and CF (200, 400, and 800 Hz) all mixed within a block of trials (see Table 2 for conditions). Only two bandwidths were used (10 and 400 Hz, based on observations from the other three groups that performance appeared constant with fixed rather than proportional bandwidth). Again, no feedback was provided.

All listeners were also administered a screening test in which a 12 000-Hz low-pass noise was presented eight times monotically to either the left or right ear and they indicated if the sound was left or right. The data from listeners with reported hearing loss, recent significant exposure to loud sounds, or who did not perform with 87.5% or better accuracy in the screening task were not used (Tables 1 and 2 indicate the number of listeners out of 25 whose data were used). Every block began with a 1-min presentation of a 12 000-Hz low-pass noise and listeners were asked to adjust their headphones so that an image appeared in the middle of their heads, and to not move their headphones during a block of trials once a centered image was obtained.

All noises were generated in the frequency domain with Rayleigh distributed amplitudes and uniform and randomly determined phases, transformed to the time domain using a 32 768-point inverse FFT, presented at 32 768 points per second via a GINA sound card, and then low-pass filtered at 12 000 Hz. All stimuli were 200 ms in duration, shaped with 20-ms \cos^2 rise-fall times, and presented at an overall level of 65 dB SPL. All ITDs were waveform ITDs in that the right ear was always presented the signal first. The sounds were presented over Sennheiser headphones (HD 280 Pro) to listeners seated in a sound-proof room.

3. Results

Figures 1 and 2 indicate the data for the L-R and MOVE tasks for the 400-Hz CF group, while Fig. 3 shows the data from the MIXED group. None of the data from the P-D task are shown in that the only variable that affected average punctate/diffuse judgments for all three CF groups was noise bandwidth; the broader the bandwidth the higher the proportion of “diffuse” judgments. ITD had no influence on the average punctate/diffuse judgments for all bandwidths and

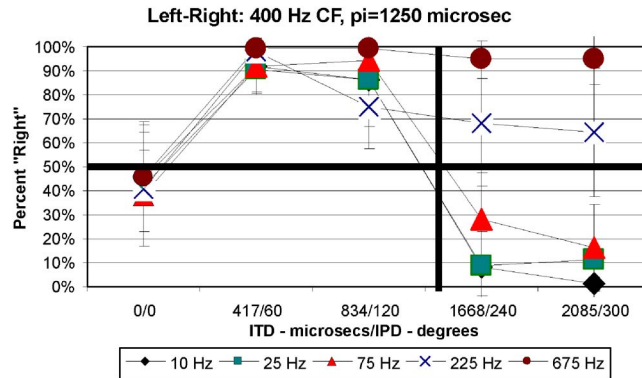


Fig. 1. (Color online) Average data (with standard error of the mean) for the 400-Hz CF group for the L-R task indicating percent “Right” judgments as a function of ITD (μs)/IPD (relative to π at CF). Different curves represent different bandwidth conditions indicated in the legend. The thick vertical line in the approximate middle of the figure represents an ITD at π of CF.

CFs. The P-D protocol may be insensitive to evaluating the diffuseness of images produced with different ITDs. The data from the 200- and 800-Hz CF groups are very similar to those from the 400-Hz group in that all of the trends seen in Figs. 1 and 2 are found in the data for the 200- and 800-Hz CF groups.

Figure 1 displays the data for the L-R task as mean percent “Right” judgments as a function of ITD/IPD *re: CF* at π (IPD is the interaural phase difference computed based on CF). The bold vertical black line in the approximate middle of the figure indicates an ITD/IPD approximately equivalent to π . The results in Fig. 1 and for each CF group are similar in that for $\text{ITD} < \pi$, percent “Right” judgments for all bandwidths and CFs increased toward 100% as ITD increased. For broader bandwidths the proportion of “Right” judgments tended to be near 100% for large ITDs, whereas for narrow bandwidths large ITDs $> \pi$ produce more “Left” judgments ($\% \text{ “Right”} < 0.5$), with the proportion of “Left” judgments increasing ($\% \text{ “Right”}$ decreasing) as ITD/IPD approached 2π .

Figure 2 shows the MOVE data plotted as the percent “Right Move” judgments versus the two ITD/IPD values presented in the two successive observation intervals of each trial. For all CF groups when the ITDs for both stimuli were less than π , the listeners indicated that the image moved right (high proportion of “Right Move” judgments) as the ITD increased from the

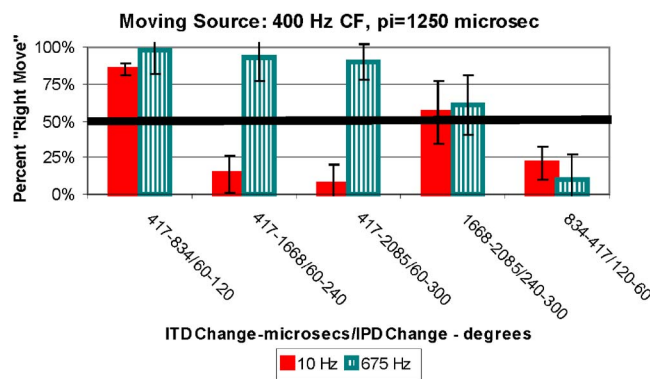


Fig. 2. (Color online) Average data (with standard error of the mean) for the 400-Hz CF group for the MOVE task indicating percent “Right Move” judgments as a function of the ITD (μs)/IPD (relative to π at CF). The first ITD/IPD is that for the first stimulus presented and the second ITD/IPD is for the second stimulus. Different bars represent different bandwidth conditions indicated in the legend.

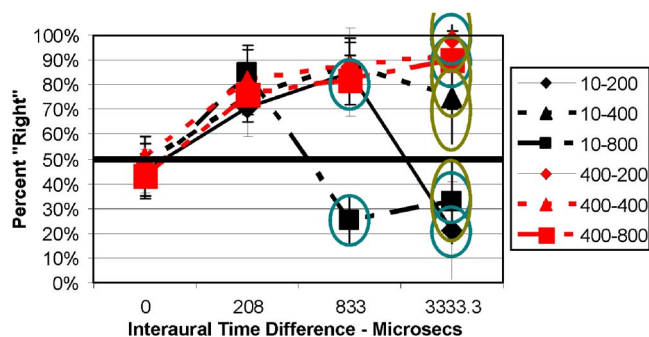


Fig. 3. (Color online) Average data (with standard error of the mean) for the MIXED group for the L-R task indicating percent “Right” judgments as a function of ITD (μs). Red curves for the 400-Hz-wide noise; black curves for 10-Hz-wide noise. Different curve and symbol types for different CF conditions as indicated in the legend. Data within blue circles are for $\text{ITD} > \pi$ at CF. Data within green ellipses are those for $\text{ITD} > 2\pi$. See Table 2.

first to the second observation interval. For the broadband noise the percent “Move Right” judgments were greater than 50% for all conditions in which the second stimulus contained the larger ITD and less than 50% when the second stimulus contained the smaller ITD (far right data in Fig. 2). This is consistent with the data in Fig. 1 (and those for the other two CF groups) in that for the broadband stimuli the images were always on the side of the head to which ITD led (right side) and larger ITDs produce images more lateralized toward the leading ear than smaller ITDs.

A different pattern of results exists for the narrow-band noise for all CF groups. Whether the listener reported the image as moving right or left depended on the relationship of the ITD to the value obtained for π at CF. When the smaller $\text{ITD} < \pi$ and the larger $\text{ITD} > \pi$, listeners indicated that the image moved left (% “Right Move” $< 50\%$). This is also consistent with the data from Fig. 1 in that images for narrow-band stimuli for $\text{ITD} < \pi$ were on the right and those for $\text{ITD} > \pi$ were on the left. Thus, when a stimulus with an $\text{ITD} < \pi$ was followed by one with $\text{ITD} > \pi$, the image would be predicted to move left (not right) as the data in Fig. 2 suggest. The trends for both narrow-band and broadband noises for all CF groups are far less evident for the conditions in which the ITD is very large (the next to the far right data in Fig. 2). In these cases for both the narrow-band and broadband cases for all CF groups the percent “Move Right” judgments are near 50%.

In the MIXED condition, Fig. 3, the left-right judgments follow the same trends as shown in Fig. 1 (and the other CF groups). When the ITD was greater than π (the blue circled data) the images were always judged to be right when the bandwidth was broad, but left when it was narrow, even when the ITD was greater than 2π (data within green ellipse). Thus, randomizing CF within a block of trials (MIXED group) did not have a major effect on left-right judgments relative to keeping CF fixed within a block of trials (CF groups).

Two audio examples are provided for listening over stereo headphones:

Mm. 1. Audio 1 file (180 kB). This is a file of type “wav”.

Mm. 2. Audio 2 file (180 kB). This is a file of type “wav”.

4. Discussion

All of the main effects of the data from Figs. 1–3 and the other CF groups are consistent with past psychophysical data. That is, use of untrained listeners who were not provided feedback and were presented high uncertainty stimulus conditions produces lateralization results for processing narrow-band and broadband stimuli that are similar to the data in the literature obtained with trained listeners in minimal uncertainty tasks with feedback. As stated in the Introduction, if there is a switch in the activation of one IC to the other as ITD is increased beyond π at CF for broadband stimuli, then one might assume broadband stimuli would be perceived on different sides of the head, depending on the relative magnitude of the ITD. The current data along with

those from other studies are not entirely consistent with such an interpretation of these IC results. Thus, listening experience, feedback, and condition complexity are probably not relevant to teasing out these relationships between physiology and perception.

There are, however, indications in the data suggesting that laterality is different when ITDs greatly exceed the biologically relevant ITD. The diffuseness of an image may be broader for large ITDs as compared to small ITDs. If the image is diffuse, then even though it might be perceived as being on one side of the head rather than the other, it might be more difficult to determine if such a diffuse image moved left or right. Thus, difficulty in making a left-right movement judgment might be an indirect measure of diffuseness. The MOVE data (Fig. 2 and for the other CF groups) for conditions in which both stimuli had ITDs much larger than the biologically possible ITDs indicate more difficulty in making motion judgments than when the ITDs were smaller. This is consistent with the possibility that such conditions produce diffuse images (although the attempt to measure diffuseness in the P-D task did not produce such results). It is also possible that these large ITDs are already lateralized as far right as possible, although why this would be frequency dependent and apply to both narrow-band and wideband stimuli is not clear. Thus, it is possible that the arguments (based on IC data) that large ITDs may be processed differently than small ITDs are partially reflected in lateral judgments of inexperienced human listeners.

The conditions used in this set of experiments and in a great deal of the psychophysical and physiological literature are unnatural in several ways. First, the ITDs are often much larger than occur in nature. Second, the interaural level difference (ILD) is usually 0 dB while ITD is varied. This does not occur in nature as ITD and ILD covary. Third, the ITDs are either only ongoing ITDs or the rise times are long compared to the ITD, making the use of ITD-onset cues highly problematic (as was the case in the present experiments). For most real world sounds there are interaural onset cues for sound sources at different azimuths and these are known to affect sound localization judgments (see Buell *et al.*, 1991). Thus, attempts to model (psychophysically or physiologically) data based on such artificial conditions may not be appropriate for describing ITD processing in the real world. To the extent that existing models do reveal relevant processing schemes for ITD, the current possible disagreement in the literature between human psychophysical data and recordings in the IC suggest that either the IC is not *the* neural circuit that forms *the* basis for psychophysical judgments based on ITD *and/or* that current psychophysical models or new models need to consider the processing that takes place within IC.

Acknowledgments

This research was supported by NIDCD grants awarded to William A. Yost and to Stanley Sheft. We thank Joe Boomer and Jim Collier, the Parmlly Hearing Institute lab technicians, and Elizabeth Chrobak, an undergraduate student, for their assistance.

References and links

- Blauert, J. (1997). *Spatial Hearing* (MIT Press, Cambridge).
- Buell, T. N., Trahiotis, C., and Bernstein, L. R. (1991). "Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays," *J. Acoust. Soc. Am.* **90**, 3077–3084.
- Harper, N. S., and McAlpine, D. (2004). "Optimal neural population coding of an auditory spatial cue," *Nature (London)* **430**, 682–686.
- Jeffress, L. A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.* **41**, 35–39.
- Kuhn, G. (1987). "Physical acoustics and measurements pertaining to directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–26.
- McAlpine, D., Jiang, D., and Palmer, A. R. (2001). "A neural code for low-frequency sound localization in mammals," *Nat. Neurosci.* **4**, 396–401.
- McFadden, D., Jeffress, L. A., and Russell, W. E. (1973). "Individual differences in sensitivity to interaural differences in time and level," *Percept. Mot. Skills* **37**, 755–761.
- Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichman, R., Griffiths, T. D., and McAlpine, D. (2006). "Representation of interaural time delay in human auditory midbrain," *Nat. Neurosci.* **9**, 1096–1098.
- Trahiotis, C., and Stern, R. M. (1989). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," *J. Acoust. Soc. Am.* **86**, 1285–1293.

- Trahiotis, C., Bernstein, L. R., and Akeroyd, M. A. (2001). "Manipulating the 'straightness' and 'curvature' of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay," *J. Acoust. Soc. Am.* **109**, 320–330.
- Wright, B. A., and Fitzgerald, M. B. (2001). "Different patterns of human discrimination learning for two interaural cues to sound-source location," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12307–12312.
- Yost, W. A. (1981). "Lateral position of sinusoids presented with interaural intensive and temporal differences," *J. Acoust. Soc. Am.* **70**, 397–409.

Acoustic micro-Doppler radar for human gait imaging

Zhaonian Zhang and Philippe O. Pouliquen

*Department of Electrical and Computer Engineering, The Johns Hopkins University, 3400 North Charles Street,
105 Barton Hall, Baltimore, Maryland 21218
zz@jhu.edu, philippe@alpha.ece.jhu.edu*

Allen Waxman

*BAE Systems, Advanced Information Technologies, 6 New England Executive Park, Burlington, Massachusetts 01803
allen.waxman@baesystems.com*

Andreas G. Andreou

*Department of Electrical and Computer Engineering, The Johns Hopkins University, 3400 North Charles Street,
105 Barton Hall, Baltimore, Maryland 21218
andreou@jhu.edu*

Abstract: A portable acoustic micro-Doppler radar system for the acquisition of human gait signatures in indoor and outdoor environments is reported. Signals from an accelerometer attached to the leg support the identification of the components in the measured micro-Doppler signature. The acoustic micro-Doppler system described in this paper is simpler and offers advantages over the widely used electromagnetic wave micro-Doppler radars.

© 2007 Acoustical Society of America

PACS numbers: 43.60.Lq, 43.35.Yb, 43.28.We [JC]

Date Received: November 15, 2006 **Date Accepted:** December 19, 2006

1. Introduction

The velocity of a moving object relative to an observer can be estimated by measuring the frequency shift of a wave radiated or scattered by the object, known as the Doppler effect. If the object itself contains moving parts, each moving part will result in a modulation of the base Doppler frequency shift, known as the micro-Doppler effect (Chen and Ling, 2002).

For example, the frequency spectrum of acoustic or electromagnetic waves scattered from a walking person is a complex time-frequency representation of human gait. It includes not only the Doppler shifted components from the velocity of the entire body but also the micro-Doppler components from the motion of the arms and legs. The acquisition of human gait signatures is important in diverse applications, ranging from rehabilitation engineering to human biometrics and surveillance. Previous studies of gait acquisition have employed electromagnetic micro-Doppler systems, such as continuous-wave (CW) X-band radars operating at 10.5 GHz (Geisheimer *et al.*, 2001; Otero, 2005). In other studies, ultrasound was used to study human gait (Sabatini and Cholla, 1998) in conjunction with infrared trigger mechanisms for pulsed ultrasound (Weir and Childress, 1997), but they have not employed micro-Doppler signature acquisition.

In this paper, we report on a portable acoustic micro-Doppler system operating in the 40 kHz acoustic frequency range and present experimental results from its application in human gait imaging.

2. Principle of operation

Given an acoustic wave transmitted by an observer, the frequency of the received wave due to a simple single-point scatterer is $f=f_0(1+2v/c)$ (Tipler, 1991), where f_0 is the frequency of the transmitted acoustic wave, v is the velocity of the scatterer relative to the observer, and c is the speed of sound. The Doppler frequency shift due to the scatterer is $f_{\text{Doppler}}=f_0 2v/c$, which is proportional to the velocity of the scatterer relative to the observer.

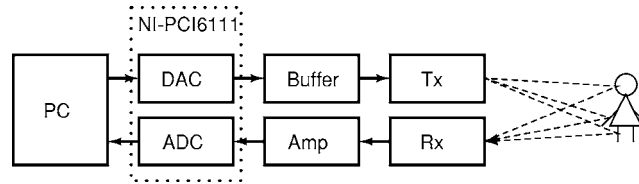


Fig. 1. Block diagram of the ultrasound micro-Doppler system. A personal computer (PC) is used to control a National Instruments (NI) PCI6111 data acquisition (DAQ) card. A continuous 40 kHz sine wave is digitally synthesized by the DAQ card and sent to a buffer to drive an ultrasonic transducer (Tx). The received signal is digitized at 1 Msamples/s and stored on the PC.

In the case of an articulated body such as a walking person, the torso, each arm, and each leg has its own velocity, and even when the torso's velocity is constant, the velocity of the limbs changes over time. The Doppler signature f_{Dsig} for such a complex object has multiple time-dependent frequency shifted components and is defined as

$$f_{\text{Dsig}}(t) = f_0 \sum_i \frac{2v_i(t)}{c}, \quad (1)$$

where $v_i(t)$ is the velocity of the torso or an individual limb as a function of time.

A two-dimensional representation of human gait can be obtained from the returned Doppler signal by applying the short-time Fourier transform (STFT) to the received signal as follows:

$$\text{STFT}(t, f) = \int x(t + \tau) g(\tau) \exp(-j2\pi f\tau) d\tau, \quad (2)$$

where $x(t)$ is the received signal, $g(t)$ is a sliding window function (e.g., a Hamming window), t is time, and f is frequency. In this time-frequency plot, the horizontal axis is time, the vertical axis is frequency, and the magnitude of the short time Fourier transform output at each point is represented by the hue of the point's color (or the intensity in the case of a gray-scale representation).

3. Experimental setup and results

A block diagram of our experimental setup is illustrated in Fig. 1 with a picture of the system in Fig. 2. The acoustic wave produced by the transmitter (Tx) is directed at a walking human and is reflected by the head, torso, and limbs. The reflected signal is received by another ultrasonic transducer (Rx), amplified through a variable gain chain, and then digitized by the data acquisition card at 1 Msamples/s. The digitized data are stored on the PC for subsequent analysis in MATLAB.

We tested the operation of the acoustic micro-Doppler radar system both indoors in a 30-foot-long corridor and outdoors in a parking lot. The ultrasonic transducers were placed at knee height, and a volunteer was instructed to walk toward or away from the transducers. We collected 10 s of data at a time and performed the short-time Fourier transform in MATLAB using a 50 ms Hamming window and 1/2 overlap between adjacent transform windows.

Figure 3 shows the spectrogram of a person walking towards the radar. An accelerometer was attached to the person's ankle and the output was acquired simultaneously with the acoustic signal from the radar receiver. Acquired accelerometer data were processed in MATLAB to calculate the Doppler frequency shift.

4. Discussions

An acoustic radar is capable of resolving motions of objects whose dimensions are equal to or larger than the wavelength of the acoustic waves. In air, the resolution of the system will be approximately 9 mm and 1 mm at 40 kHz and 340 kHz respectively, for which commercial

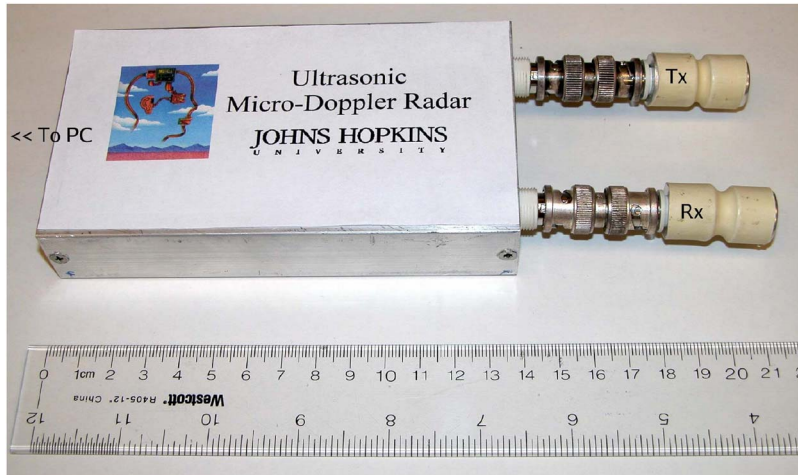


Fig. 2. (Color online) Photograph of the acoustic micro-Doppler radar system.

off-the-shelf transducers are available. Analytical expressions are available to calculate the absorption of sound in still air, given the acoustic frequency, atmospheric pressure, temperature, and relative humidity. (Evans *et al.*, 1972; Bass *et al.*, 1990) For frequencies above 10 kHz, the absorption of sound in dry air can be approximated as a linear function of the square of the frequency. (Crocker, 1998) For the frequencies of interest here, the attenuation is about 1.3 dB/m at 40 kHz and 20 dB/m at 340 kHz, therefore the range of a 340 kHz system is inherently determined to be 15 times smaller than that of a 40 kHz system.

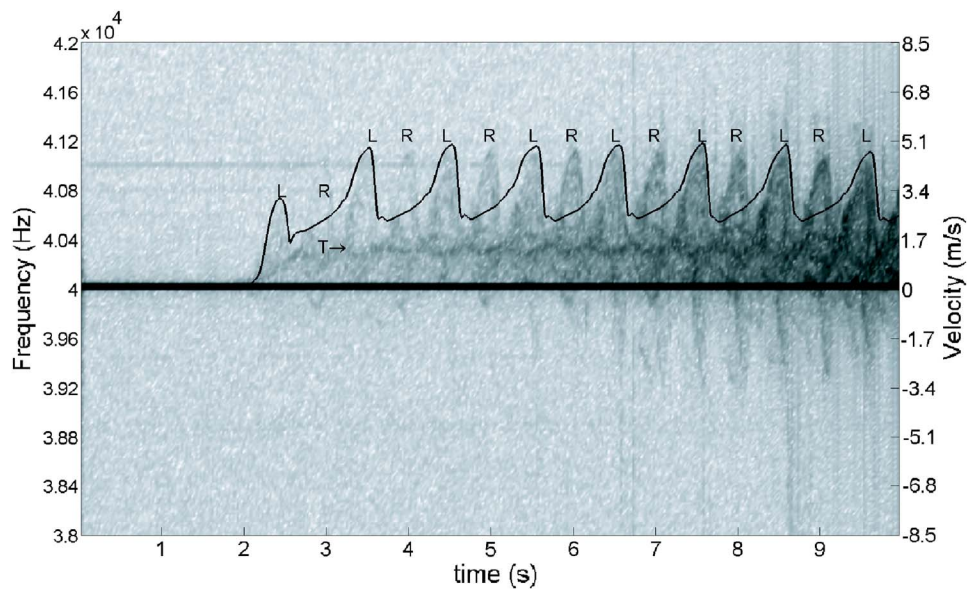


Fig. 3. Micro-Doppler spectrogram of a person walking towards the radar. An accelerometer is attached to the left leg and the velocity derived from the accelerometer signal is superimposed on the spectrogram. The letters 'L' and 'R' mark regions of the spectrogram representing the motion of the left and right legs, respectively. The right arrow 'T' points to a region of the spectrogram around 40.35 kHz, corresponding to a strong Doppler return due to the motion of the torso (walking speed approximately 1.5 m/s). The features below 40 kHz are Doppler shifts caused by arms swinging away from the transducers.

The velocity resolution is determined by the sampling rate of the ADC and the window size of the STFT. Given a sampling rate of 1 Msamples/s and a window size of 50 000 samples in our system, the frequency resolution is 20 Hz, corresponding to a velocity resolution of 0.085 m/s when 40 kHz transducers are used.

5. Conclusion

A continuous-wave ultrasound micro-Doppler radar system for human gait analysis is reported in this paper. The use of ultrasound rather than microwaves makes the data acquisition and signal processing tasks easier to perform with standard audio frequency digital signal processing hardware and provides immunity from electromagnetic interference sources. The system is simple and can operate in natural environments that are complementary to microwave micro-Doppler radars, such as underwater. The micro-Doppler system can also be used in robotics for intelligent mapping of dynamic environments.

Preliminary experimental data (not shown in this paper) suggest that we are also able to detect multiple humans walking in the observation field. It is also worthwhile noting that each volunteer in our study had a somewhat unique gait signature, suggesting that human gait signatures may provide useful information in biometric identification systems.

Acknowledgments

This work was supported by MASINT project “Decentralized-Fusion, On-Demand Activation, Awareness Sensor Network” NMA401-02-9-2002 under a subcontract by Honeywell and by NSF Grant No. IIS-0434161.

References and Links

- Bass, H. E., Sutherland, L. C., and Zuckerwar, A. J. (1990). “Atmospheric absorption of sound: Update,” *J. Acoust. Soc. Am.* **88**(4), 2019–2021.
- Chen, V. C., and Ling, H. (2002). *Time-Frequency Transforms for Radar Imaging and Signal Analysis* (Artech House, Boston).
- Crocker, M. J. (ed.) (1998). *Handbook of Acoustics* (Wiley, New York).
- Evans, L. B., Bass, H. E., and Sutherland, L. C. (1972). “Atmospheric absorption of sound: Theoretical predictions,” *J. Acoust. Soc. Am.* **51**(5B), 1565–1575.
- Geisheimer, J., Marshall, W., and Greneker, E. (2001). “A continuous-wave (CW) radar for gait analysis,” in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computer* (IEEE, Pacific Grove, CA), Vol. 1, pp. 834–838.
- Otero, M. (2005). “Application of a continuous wave radar for human gait recognition,” in *Proc. SPIE: Signal Processing, Sensor Fusion, and Target Recognition XIV*, Vol. 5809, Orlando, FL, pp. 538–548.
- Sabatini, A., and Colla, V. (1998). “A method for sonar based recognition of walking people,” *Rob. Auton. Syst.* **25**, 117–126.
- Tipler, P. A. (1991). *Physics for Scientists and Engineers*, 3rd ed. (Worth, New York).
- Weir, R. F., and Childress, D. S. (1997). “A new method of characterizing gait using a portable, real-time, ultrasound ranging device, in *Proceedings of 19th International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, pp. 1810–1812.

Ultrasonic wave generation due to human footsteps on the ground

Alexander Ekimov and James M. Sabatier

*National Center for Physical Acoustics, The University of Mississippi, 1 Coliseum Drive,
University, Mississippi 38677
aekimov@olemiss.edu, sabatier@olemiss.edu*

Abstract: Human footsteps generate broadband frequency vibrations in the ground/floor and sound in the air from a few Hertz up to ultrasonic frequencies due to striking and sliding contacts between a foot and the ground/floor. The high-frequency (above 1 kHz) vibrations from footsteps were detected on a building floor, but were not detected on the outdoor ground, even at 1 m from a walker. This paper presents results of ultrasound registration from footsteps on the ground at greater distances. Results are based on sound measurements in air, since the sound absorption in air is less than vibration absorption in the ground.

© 2007 Acoustical Society of America

PACS numbers: 43.35.Yb, 43.20.Tb, 43.20.Ye [NX]

Date Received: December 4, 2006

1. Introduction

Walking people generate vibrations and sound in a broadband frequency range from a few Hertz up to ultrasonic frequencies due to interactions between the foot and the supporting surface.^{1,2}

There are two characteristic frequency bands in the vibration and sound responses of footstep signatures.^{1,2} The first frequency band is generated by a force normal to the supporting surface and is concentrated in a low-frequency range below 500 Hz. The second frequency band is generated by the tangential (friction) force and is located in a high-frequency range, above 1 kHz up to ultrasonic frequencies. Indoor tests have demonstrated such high-frequency bands in the vibration and sound responses from human footsteps,^{1,2} while outdoor tests have shown only the low-frequency vibration responses (below 1 kHz).² In buildings, footstep vibration magnitudes in the high-frequency range are comparable and independent of walking styles (regular, soft, and stealthy)¹ at distances close to the detector (1 m). However, on the ground/floor surfaces, different walking styles result in different vibration signatures in the low-frequency range that limit the maximum ranges for the seismic method of footstep detection. For example, the stealthy walking style is undetectable on the ground even a few meters from a seismic detector.^{2,3} Sound measurements can provide the high-frequency response of footsteps on the ground, but previous experiments have not measured sound responses from human footsteps in outdoor environments.

This article presents test results of measuring the high-frequency response in footstep sounds. Earlier work was based on footstep-induced vibrations in the ground;² here tests are based on sound measurements in air, since the sound absorption in air is significantly less than vibration absorption in the ground.⁴ Low-frequency, ambient acoustic noise can significantly decrease the dynamic range (30 dB or more⁵) of the footstep signal in ultrasonic frequencies when measured by broadband microphone;^{1,5} therefore, a narrow-band (resonant) ultrasonic microphone was used in these tests. The ultrasonic microphone had maximum sensitivity in the ultrasonic frequency range and rolled-off in sensitivity in the low-frequency range.

This article presents test results, using both microphones and accelerometers, from measurements of human footsteps in a broadband frequency range (vibrations) and in a narrow-band ultrasonic frequency range (sound pressure) outdoors.

2. Comparison of attenuations of sound in air and vibrations in the ground at ultrasonic frequencies

Vibrations in the ground and sound in air, generated by human footsteps, attenuate over distance by geometric spreading loss due to propagation and energy loss due to absorption.^{6,7} The resulting sound pressure $P(\omega, R, \alpha_s, t)$ and vibration $V(\omega, R, \alpha_v, t)$ magnitudes at a distance R from a walker may be presented by

$$P(\omega, R, \alpha_s, t) = P_0(\omega, t) \times A_{\alpha_s}(R, \alpha_s) \times A_s(R), \quad (1)$$

$$V(\omega, R, \alpha_v, t) = V_0(\omega, t) \times B_{\alpha_v}(R, \alpha_v) \times B_v(R), \quad (2)$$

where $P_0(\omega, t)$ and $V_0(\omega, t)$ are the sound and the vibration footprint signatures at the walker's location, $\omega = 2\pi f$, f is a frequency, t is time, $A_{\alpha_s}(R, \alpha_s)$ and $B_{\alpha_v}(R, \alpha_v)$ are the sound and the vibration absorption functions, respectively, α_s and α_v are the air and the ground absorption coefficients, respectively, and $A_s(R)$ and $B_v(R)$ are the attenuation functions due to geometric spreading for the sound and the vibration, respectively.

Atmospheric absorption of sound is caused by friction losses in the transmission medium.⁶ The result of sound attenuation $A_{\alpha_s}(R, \alpha_s)$ by atmospheric absorption is presented as a function of a distance R and the air absorption coefficient α_s in dB/m:

$$A_{\alpha_s}(R, \alpha_s) = 10^{-(1/20) \times \alpha_s R}. \quad (3)$$

Methods and equations for the calculation of α_s as a function of frequencies, humidity, temperature, and air pressure are given in Refs. 8 and 9. The maximum value of absorption α_s (in dB/m) in air at room temperature over all humidities for frequencies up to 50 kHz is given by¹⁰

$$\alpha_s(f) = 0.000\,033 \times f, \quad (4)$$

where f is the frequency of sound in Hz. $f = c_s / \lambda_s$, where c_s is the speed of sound in air ($c_s = 343$ m/s at 20 °C), and λ_s is the sound wavelength. Equation (4) may be rewritten as

$$\alpha_s(f) = 0.0114 \times \lambda_s^{-1}. \quad (5)$$

Vibration attenuation $B_{\alpha_v}(R, \alpha_v)$ due to absorption in the ground is described by an equation similar to Eq. (3), where α_s is changed to α_v . The absorption coefficient α_v is defined as the energy lost due to seismic wave propagation through a distance equal to a complete wavelength λ_v . Values of α_v for common Earth materials⁷ range from 0.25 to 0.75 dB per λ_v :

$$\alpha_v(f) = (0.25 - 0.75) \times \lambda_v^{-1}. \quad (6)$$

From Eqs. (3), (5), and (6), it follows that the largest values of absorption $A_{\alpha_s}(R, \alpha_s)$ and $B_{\alpha_v}(R, \alpha_v)$ are independent of frequency at the distance scale of $R = \lambda$, and as a result the higher frequency waves attenuate faster than the lower frequency waves versus distance R . Comparison of absorption coefficients α_s [Eq. (5)] and α_v [Eq. (6)] for the same wavelengths shows that the seismic absorption coefficient α_v is greater than the sound absorption coefficient α_s . For example, using Eqs. (5) and (6), a sound wave in air with $\lambda = 0.01$ m will be attenuated due to absorption by 1.14 dB per 1 m, and the vibration wave in the ground with the same wavelength will be attenuated by 25–75 dB per 1 m.

In the case of the same geometric spreading loss due to propagation $A_s(R) = B_v(R)$ (which may be valid for short distances between a walker and a detector location on the ground), the resulting attenuation is much less for sound waves than for seismic vibrations due to the difference in absorption losses.

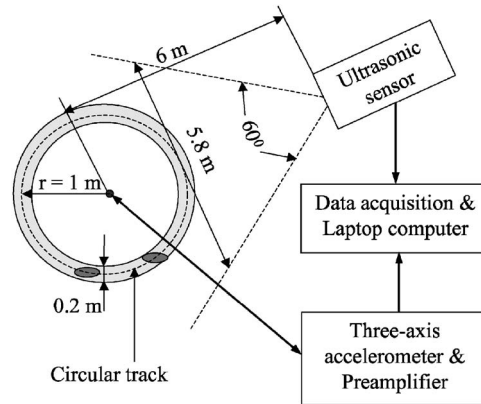


Fig. 1. Setup for the measurement of human footstep vibration and sound signatures on the ground.

3. Setup for measurements of human footstep vibration and sound signatures on the ground

Since sound in air has less absorption than vibrations in the ground, a new concept was pursued for high-frequency measurement of footstep signatures in outdoor tests. A resonant ultrasonic ceramic transducer (microphone) was used for the high-frequency sound measurements. The resonant receiver was chosen because of the high level of the ambient noise floor in the low-frequency range when compared with ultrasonic frequencies.⁵ However, the resonant receiver naturally filtered low frequencies and, as a result, increased the dynamic range of measurements in the high frequencies.

Outdoor tests were conducted in a grassy area on the University of Mississippi campus. The test setup (which was a modification of the test setup described in Refs. 1 and 2) consisted of a circular track marked on the ground at a constant distance from an accelerometer to the subject during the measurements (Fig. 1). In these experiments a person walked on the circular track, which was 0.2 m wide. The median radius (r) of the track was 1 m and, depending on the length of the person's stride, 9 to 11 steps were required to complete a full circle.

A three-axis accelerometer (PCB 356B18) was threaded into the top of a 10-cm spike hammered into the ground at the center of the circle (Fig. 1). The accelerometer was calibrated over the frequency range of 0.2 Hz to 16.5 kHz and had a sensitivity of 1 V/g . A battery-powered signal conditioner (PCB model 480B21) amplified signals from the accelerometer. Only the component of acceleration normal to the ground surface is presented and discussed in this article. An ultrasonic ceramic sensor (25OSR) was attached to a tripod of 1.2-m height and placed 6 m away from the center of the test track (Fig. 1). The ultrasonic ceramic sensor (UCS) had a resonance frequency of 25.5 kHz and a typical bandwidth (at -6 dB) of 1 kHz . The directivity (at -6 dB) was 60° . The UCS was calibrated at its resonance frequency of 25.5 kHz and had a sensitivity of $-18\text{ dB re: } 1\text{ V/Pa}$. In the configuration presented in Fig. 1, the beam pattern of the ultrasonic sensor covered all of the test area. A receiving preamplifier (not shown in Fig. 1) amplified signals from the UCS. Data recording and processing were conducted using a two-channel, 16-bit data acquisition board (DAQ) (Echo Indigo IO) and a laptop computer with Sound Technology software (LAB432).

4. Test results: Footstep vibration and sound signatures on the ground

Footstep acoustic signatures for regular, soft, and stealthy walking styles (styles were described in Ref. 1) were measured using the setup shown in Fig. 1. A man walked at a constant speed for all test styles. A DAQ with a sampling rate of 64 kHz and a 32-kHz antialiasing filter acquired signals from the accelerometer and the UCS. The data (sound pressure and acceleration) for a full circle of ten footsteps for each walking style were taken and merged together in one data

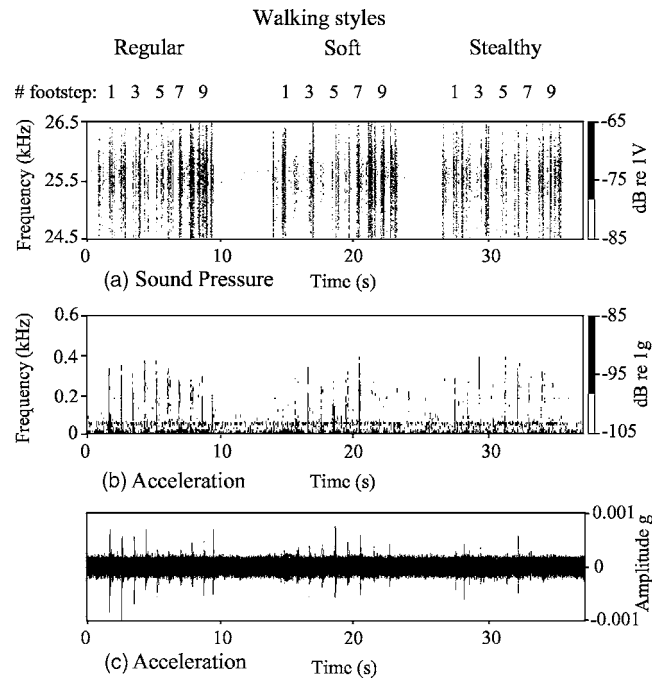


Fig. 2. Footstep signatures of ten regular, soft, and stealthy footsteps on the ground: (a) and (b) are the spectrograms of sound pressure at a distance of 5–7 m from a walker and acceleration at a distance of 1 m from a walker, respectively, and (c) is the time domain signal of acceleration. The sampling rate was 64 kHz and the FFT size was 8192.

file. The results of processing are presented in Figs. 2(a)–2(c). The walking styles and odd footsteps on the test track are marked in these figures. Spectrograms of the sound and acceleration responses of the ground in the frequency band 24.5–26.5 kHz for the sound and 20–600 Hz for the acceleration are presented in Figs. 2(a) and 2(b). The time domain signal of the acceleration is presented in Fig. 2(c).

The high-frequency vibrations above 600 Hz were not detectable; all data are presented in the frequency range below 600 Hz. The fast Fourier transform (FFT) size was 8192, which corresponded to 7.81 Hz in the spectral line resolution and 128 ms in the time resolution. Each Fourier spectral line resolution is equal to the sampling rate divided by the FFT size. The qualitative comparison of the sound pressure and vibration responses of the walking styles shows a stable level of sound pressure signals for all walking styles in the distance range of 5–7 m. The vibration response was less detectable for soft and stealthy styles even at 1 m from the detector.

In Figs. 3(a) and 3(b), the average spectra of ten footsteps for each walking style are presented for sound pressure and vibration signals, respectively. The sampling rate was 64 kHz and the FFT size was 32 768, which corresponded to 1.95 Hz in the spectral line resolution and 512 ms in the time resolution. The maxima of sound pressure responses have comparable magnitudes for all walking styles. The signal-to-noise ratio is close to 15 dB. This effect supports the conclusion (in Ref. 1) that the reason for comparable sound magnitudes for different styles of walking in the high-frequency range is the friction between a foot and the ground. This conclusion was presented in Ref. 1 for footstep vibration signatures in the high-frequency range for different styles of walking in buildings.

The maxima of the ground vibration response to footsteps are in the frequency band near 19 Hz for all walking styles. The magnitudes of the sound pressure at 25.5 kHz and the acceleration at 19 Hz, taken from Figs. 3(a) and 3(b), are presented in Table 1. A comparison of

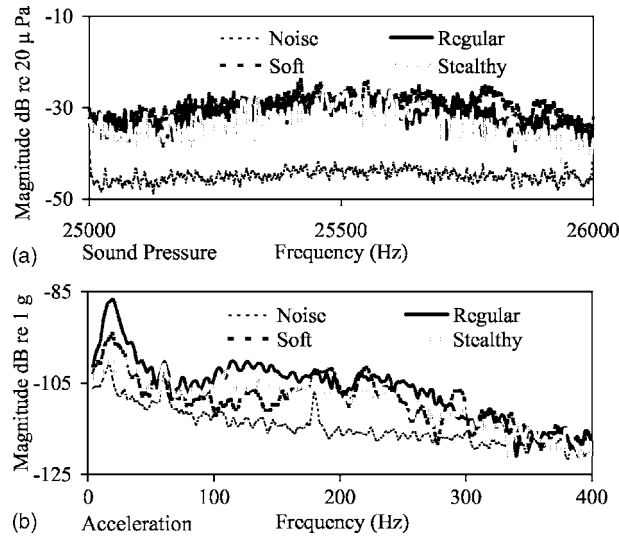


Fig. 3. Average Fourier spectra of ten regular, soft, and stealthy footsteps on the ground, and the background noise floor: (a) the sound pressure at 5–7 m from a walker and (b) the acceleration at 1 m from a walker. The sampling rate was 64 kHz and the FFT size was 32 768.

the walking styles and maxima of the vibration responses shows that the vibration maxima are related to the walking styles [see Table 1 and Fig. 3(b)]. Soft and stealthy walking reduced the vibration response of the ground by 7.5 and 12.5 dB, respectively, relative to regular walking. As a result, the distance range for footstep detection was reduced. The sound pressure response in the high-frequency range had no similarly strong dependence on walking style; the dispersion of magnitudes is within 1 dB.

5. Conclusions

Human footstep vibration and sound signatures of three distinct styles of walking (regular, soft, and stealthy) with the same speed of motion were measured on an outdoor ground surface. Results indicated a strong dependence between walking styles and vibration magnitudes in the low-frequency range (below 400 Hz). The seismic response from footsteps in the frequency range above 600 Hz was not detected even at 1 m from the detector.

The measurement of human footstep sound signatures in narrow-band ultrasonic frequencies was conducted at expanded ranges of up to 7 m from the detector. Tests are based on sound measurements, since the sound absorption in air is significantly less than vibration absorption in the ground. This allows human footstep detection at increased ranges. These tests demonstrate evidence of ultrasonic sound wave generation due to friction of a footstep on the ground. The magnitude of ultrasonic signals in air had no strong dependence on the human

Table 1. Sound pressure (at 5–7 m) and vibration (at 1 m) levels of walking styles.

Walking style	Maximum sound pressure at 25.5 kHz		Maximum vibration at 19 Hz	
	dB re: 20 μ Pa	dB re Regular	dB re: 1 g	dB re Regular
Regular	-28	0	-86.5	0
Soft	-27	+1	-94	-7.5
Stealthy	-28	0	-99	-12.5

footstep style (regular, soft, and stealthy), unlike the seismic vibration signals. Outdoor tests showed the possibility of human footstep signature measurements on a grassy ground at 7 m from the ultrasonic detector with a signal-to-noise ratio near 15 dB.

Acknowledgments

This work was supported by the Department of the Army, Army Research Office, under Contract No. W911NF-04-1-0190. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the sponsor.

References and links

- ¹A. Ekimov and J. M. Sabatier, "Vibration and sound signatures of human footsteps in buildings," *J. Acoust. Soc. Am.* **120**, 762–768 (2006).
- ²A. Ekimov and J. M. Sabatier, "Broad frequency acoustic response of ground/floor to human footsteps," *Proc. SPIE* **6241**, OL1–OL8 (2006).
- ³K. M. Houston and D. P. McGaffigan, "Spectrum analysis techniques for personnel detection using seismic sensors," *Proc. SPIE* **5090**, 162–173 (2003).
- ⁴L. L. Beranek, *Acoustical Measurements* (AIP, Woodbury, NY, 1988).
- ⁵H. E. Bass and L. N. Bolen, "Ultrasonic background noise in industrial environments," *J. Acoust. Soc. Am.* **78**, 2013–2016 (1985).
- ⁶L. B. Evans, H. E. Bass, and L. C. Sutherland, "Atmospheric absorption of sound: theoretical predictions," *J. Acoust. Soc. Am.* **72**, 1565–1575 (1972).
- ⁷P. Kearey, M. Brooks, and I. Hill, *An Introduction to Geophysical Exploration*, 3rd ed. (Blackwell Science, Padstow, 2002).
- ⁸American National Standard, "Method for the calculation of the absorption of sound by the atmosphere," ANSI S1.26-1995 (ASA-23-1995).
- ⁹International Standard, "Acoustics-Attenuation of sound during propagation outdoors—Part 1: Calculation of the absorption of sound by the atmosphere," ISO 9613-1:1993(E).
- ¹⁰L. B. Evans and H. E. Bass, "Tables of absorption and velocity of sound in still air at 68°F," Report WR72-2, Wyle Laboratories, Huntsville, AL (1972).

Acoustic measures of low-frequency noise in extended high-frequency audiometry

Nicolas Schmuziger

*Department of Otorhinolaryngology, University Hospital, CH-4031 Basel, Switzerland and
Department of Otorhinolaryngology, Cantonal Hospital, CH-5001 Aarau, Switzerland
nicolas.schmuziger@ksa.ch*

Matthias Brechbuehl

*Norsonic Brechbuehl AG, CH-3452 Grünenmatt, Switzerland
matthias.brechbuehl@norsonic.ch*

Rudolf Probst

*Department of Otorhinolaryngology, University Hospital, CH-4031 Basel, Switzerland and
Department of Otorhinolaryngology, University Hospital, CH-8091 Zürich, Switzerland
rudolf.probst@usz.ch*

Abstract: A very high signal-to-noise ratio is required for equipment designed for extended high-frequency audiometry because listeners with almost no hearing ability in the extended high-frequency range may have normal hearing sensitivity in the lower frequencies. Two commercially available systems designed for pure-tone audiometry were evaluated both in the conventional and extended high-frequency range. Unwanted lower frequency signals greater than the noise floor occurred predominantly at presentation levels of approximately 110 dB SPL or higher. Test tones in the extended high-frequency range should be restricted to levels that are not associated with lower frequency noise.

© 2007 Acoustical Society of America

PACS numbers: 43.80.Vj, 43.58.-e [BLM]

Date Received: December 1, 2006 **Date Accepted:** January 8, 2007

1. Introduction

Extended high-frequency audiometry is a suitable method for the early detection of ototoxicity (e.g., Fausti *et al.*, 1999). Several audiometers are available that include the extended high-frequency range. A very high signal-to-noise ratio is required for such equipment because listeners with almost no hearing ability in the extended high-frequency range, particularly for the highest frequencies, may have normal hearing sensitivity for the lower frequencies. According to the specifications of the International Electrotechnical Commission (IEC), “no test subject shall detect any unwanted sound from the transducer coinciding with the presentation of the test tone, even at the maximum setting of the hearing level control (IEC, 1994).”

Reference equivalent threshold sound pressure levels (RETSPLs) correspond to the mean/modal hearing thresholds of a “sufficiently large number of ears of otologically normal persons of both sexes aged between 18 and 30 years,” according to IEC 60645-1 (2001). They are lower than 10 dB SPL in the frequency range from 0.5 to 6 kHz and increase up to 56 dB SPL at 16 kHz using the circumaural earphone HDA 200 from Sennheiser (ISO 389-5, 1998). Hearing thresholds of patients receiving aminoglycosides are often poorer than these RETSPLs, particularly in the extended high-frequency range, considering that presbycusis affects the highest frequencies first. Therefore, test tones in the extended high-frequency range have to be presented often near to the maximum setting of the hearing level control, particularly at 14 and 16 kHz. Such sound pressure levels are typically around 100 to 120 dB SPL, making unwanted low frequency noise or sound more likely.

In the literature, only a few incidental remarks may be found concerning the issue of the effects of low-frequency noise on thresholds in the extended high-frequency range. Frank

(1990) reported that unwanted sounds below the test frequencies were not detected within the 80-dB dynamic range of the signal analyzer testing tone levels of 120 dB SPL at frequencies of 10, 12, 14, 16, 18, and 20 kHz using a commercially available Beltone 2000 audiometer (Beltone Electronics). Unwanted sound or signals above the test frequencies were present at harmonics of these frequencies and were at least 35 dB below the presentation level of the test tones.

In evaluating equipment for an experiment on the effects of noise exposure (Schmuziger *et al.*, 2004), the first author noted that there was noise present in the signals generated by some equipment. Based upon this observation, a survey of commercially available equipment used for testing in the conventional and extended high-frequency range was undertaken. The low-frequency noise present during the output of signals in the extended high-frequency range was measured in two audiometers that were selected because they were available commercially on the local market at the time of the study. This article discusses the results of the survey and their clinical implications.

2. Material and methods

2.1 Audiologic equipment

Two commercially available systems designed for pure-tone audiometry both in the conventional and the extended high-frequency range were evaluated: Madsen Itera II (GN Otometrics, Denmark) and GSI 61 (Grason-Stadler, United States of America). They are diagnostic two-channel audiometers that, according to the manufacturers, meet or exceed the international standard for equipment for extended high-frequency audiometry according to IEC 60645-4 (1994). Frequencies in the extended high-frequency range included 8, 9, 10, 11.2, 12.5, 14, and 16 kHz according to ISO 389-5 (1998). According to IEC 60645-4 (1994), the frequency of 8 kHz is considered to be both the highest frequency in the conventional range and the lowest frequency in the extended high-frequency range. Additionally, the GSI 61 has an option for 18 and 20 kHz.

2.2 Calibration

Audiometers, equipped with Sennheiser HDA 200 earphones, were calibrated according to the regulations of the Swiss Federal Office for Metrology and according to ANSI S3.6-1996 (1996), ISO 389-1 and 389-5. Reference equivalent threshold sound pressure levels for 18 and 20 kHz on the GSI 61 audiometer were specified by the manufacturer.

2.3 Sound level measurements

Sound level measurements were performed by Norsonic Brechbuehl AG, Switzerland, in a double-walled sound treated booth. Only one channel/headphone combination was evaluated for each audiometer. The selection of this combination was counterbalanced across equipment. Measurements were made using an IEC 318 artificial ear, which includes a Brüel and Kjaer (B&K) 4153 artificial ear with a B&K type 1 flat plate adaptor (DB0843), and a B&K 4134 $\frac{1}{2}$ -in. microphone. The artificial ear was connected to a 01 dB-Metravib Orchestra signal analyzer with a dynamic range of more than 100 dB with the option for signal analysis by real-time narrow band fast Fourier transform (FFT, 4096 points, Hanning Window).

Tones were presented initially at the maximum output of each audiometer, which was predetermined by the manufacturer and reduced in 5-dB steps. Median maximum output levels for all test frequencies in the extended high-frequency range were 109 dB SPL (range: 106–115) for the GSI 61 and 116 dB SPL (range 101.6–126) for the Itera II. The results were stored on the hard disc of the analyzer for later analysis on the PC using the dBFA software suite from 01 dB-Metravib.

3. Results

Unwanted lower frequency signals at audible levels were more prominent for the Madsen Itera II than for the GSI 61. All of the lower frequency signals were audible to the first author, whose

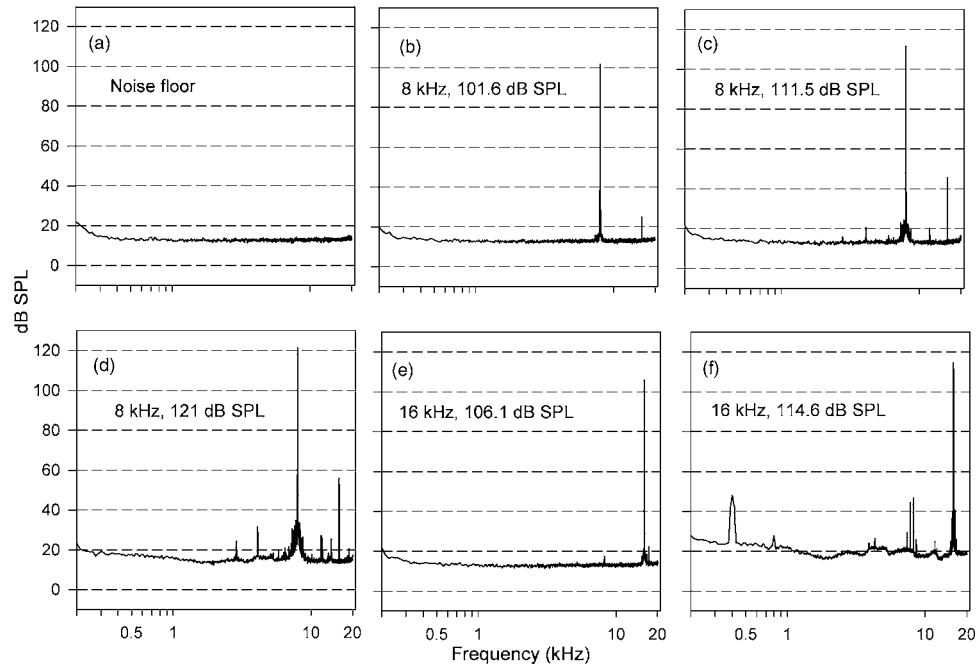


Fig. 1. Spectral analysis of the acoustic output of the Madsen Itera II generating 8- and 16-kHz tones at different levels specified in the title of each spectrum. The noise floor of the audiometer (signal analysis without test tone) is depicted in panel (a).

audiometric thresholds are plotted in Fig. 2. The noise floor from the audiometers was approximately 13 dB SPL for frequencies above 0.3 kHz and is shown in Fig. 1(a) for the Madsen and in Fig. 2(a) for the GSI.

3.1 Madsen Itera II

Unwanted lower frequency sounds occurred particularly at 8 and 16 kHz. Unwanted signals below a test frequency of 8 kHz presented at approximately 121 dB SPL were present at harmonics of 4.1 kHz with a level of 31.6 dB SPL. Moreover, peaks occurred at 2.9 and 0.3 kHz

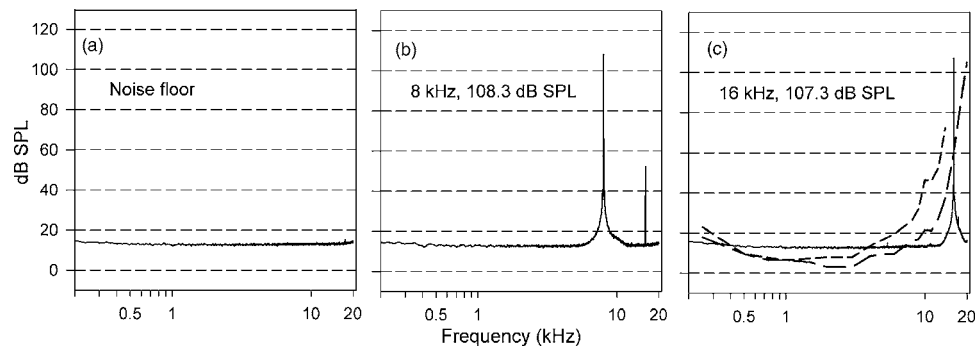


Fig. 2. Spectral analysis of the acoustic output of the GSI 61 generating 8- and 16-kHz tones at different levels specified in the title of each spectrum. Gray medium-dashed line: Hearing threshold levels of the first author (dB SPL). Dark gray long-dashed line: RETSPLs for 0.25–20 kHz. The noise floor of the audiometer is depicted in panel (a).

with levels of 24.3 and 19.3 dB SPL, respectively [Fig. 1(d)]. Reducing the level of the signal to 111.5 dB SPL or lower reduced the occurrence of these unwanted signals to a few dB over the noise floor, as depicted in Figs. 1(b) and 1(c).

A presentation level of 114.6 dB SPL at 16 kHz was associated with unwanted signals of close to 50 dB SPL at the harmonic frequency of 8 kHz, but such signals were also present near 0.4 kHz, as shown in Fig. 1(f). These unwanted signals could not be detected above the noise floor when the level was reduced to 106.1 dB, as demonstrated in Fig. 1(e).

3.2 GSI 61

No unwanted signals could be measured for any frequencies with our equipment, even at the maximum output of the audiometer. However, the listening check by the first author revealed weakly audible low-level noise at lower frequencies when a 16-kHz signal was presented at 107.3 dB SPL, 18 kHz at 109.1 and 113 dB SPL, and 20 kHz at 115 dB SPL. The audiometric thresholds of the first author and RETSPLs are plotted in Fig. 2(c), along with the output of the signal analyzer for a 16-kHz tone at 107.3 dB SPL.

4. Discussion and conclusions

The International standard IEC 60645-4 (1994) for equipment for extended high-frequency audiometry states that “No test subject shall detect any unwanted sound from the transducer..., even at maximum setting of the hearing level control.” Because many listeners with almost no hearing ability at the highest frequencies, particularly at 14 kHz and above, have normal hearing sensitivity for the lower frequencies, these requirements are very difficult to achieve for commercially available clinical audiometers. This is a serious limitation for the clinical use of extended high-frequency audiometry because false-positive responses may occur. The most common clinical use of extended high-frequency audiometry is monitoring of potential aminoglycoside-induced hearing loss, which typically affects hearing first in the highest frequency range with later progression to the lower frequency regions (Fausti *et al.*, 1999). Thresholds in the extended high-frequency range potentially could be underestimated because a patient might be responding to lower frequency noise produced by the hardware. Therefore, perception of lower frequency noise or off-frequency signals could result in missing early signs of ototoxicity.

Spectral analysis of the output of the audiometer does not guarantee that the requirements according to IEC 60645-4 (1994) are met in all cases, because the dynamic range of most analyzers is not wide enough to cover the full range of differences between extended high- and low-frequency hearing. The unwanted signals may not be detectable above the noise floor of the audiometer. The signal analyzer used in this study had an excellent dynamic range of more than 100 dB, which is greater than analyzers available for clinical use (Frank, 1990). Nevertheless, our results with the GSI 61 audiometer demonstrated that even when no unwanted signals were measured above the noise floor of the audiometer, some low-level noise at lower frequencies was audible by the first author performing listening checks when test tones at 16 kHz and above were presented at the maximum output of the audiometer.

Unwanted lower frequency signals or off-frequency signals occurred more often for the Madsen Itera II than for the GSI 61, probably because the maximal output levels of the test tones were generally higher and less restricted for the Madsen. The output levels of test tones in the extended high-frequency range should be restricted to meet the requirements of IEC 60645-4 (1994). According to IEC 60645-1 (2001), “objective acoustical measurements may be impracticable for testing for the presence of unwanted sound from the audiometer. Therefore, subjective tests shall be performed using at least two otologically normal test subjects whose hearing threshold levels shall not exceed 10 dB for the test frequencies 250 Hz to 8 kHz.” Similar subjective tests could be advantageous for extended high-frequency audiometry, using otologically normal test subjects with normal hearing in the conventional frequency range and presumably impaired hearing in the extended high-frequency range. Stelmachowicz *et al.* (1989) obtained auditory thresholds in the 8- to 20-kHz range from 240 subjects ranging in age from 10 to 60 years and found that the largest changes in sensitivity with age occurred between

40 and 59 years. Considering their results, it seems reasonable to assume that anyone over the age of 40 years is likely to have hearing loss in the extended high frequencies. Providing that such subjects have normal hearing in the conventional frequency range, they would probably be suitable to participate in such subjective tests. Subjects with hearing loss in the extended high-frequency range are more suitable than subjects with normal hearing in this frequency range because low-frequency sound or noise will usually only occur at test tone levels near the maximum output of the audiometer and one would not want to expose subjects with normal hearing to such high levels. Even more importantly, such test tone levels well above the individual's threshold could mask unwanted lower-frequency sound.

Manufacturers of equipment for extended high-frequency audiometry should be responsible for performing the subjective tests needed to check the accuracy of their equipment and potentially to restrict the maximum output of their audiometers to levels that are unlikely to produce unwanted lower-frequency sound. Moreover, such subjective tests should be regularly performed during the clinical use of such equipment, at least at the time of regular calibrations.

Further efforts are needed by the manufacturers to design commercially available audiometers for extended-high-frequency audiometry with improved signal-to-noise ratio. The introduction of a steady background of low-level noise restricted to the conventional frequency range to the test earphone could be helpful for masking the off-frequency energy or unwanted sound. It could prevent listeners from responding to these low-frequency sounds when high-frequency test signals are presented. Further studies could be advantageous to determine the effective dynamic range of the measurement system combining acoustical measurements of the test tones and masking studies with noise restricted to the conventional frequency range.

Acknowledgments

This study was supported by grants from the Society of Hearing-Impaired People of Basel, Switzerland. Dr. Frances P. Harris provided very helpful comments on an earlier version of this manuscript. The authors gratefully acknowledge Albert Stücheli from Diatec, Switzerland, and Dr. Martin Kompis, Head of the Audiology from the Department of Otorhinolaryngology, University Hospital Berne, Switzerland, for their collaboration.

References and links

- American National Standards Institute (1996). "American National Standard Specification for Audiometers," ANSI S3.6 (ANSI, New York).
- Fausti, S. A., Henry, J. A., Helt, W. J., Phillips, D. S., Frey, R. H., Noffsinger, D., Larson, V. D., Fowler, C. G. (1999). "An individualized, sensitive frequency range for early detection of ototoxicity," *Ear Hear.*, **20**, 497–505.
- Frank, T. (1990). "High-frequency hearing thresholds in young adults using a commercially available audiometer," *Ear Hear.*, **11**, 450–454.
- International Electrotechnical Commission (1994). "Audiometers-Part 4: Equipment for extended high-frequency audiometry," IEC 60645-4 (IEC, Geneva).
- International Electrotechnical Commission (2001). "Electroacoustics—Audiological equipment - Part 1: Pure-tone audiometers," IEC 60645-1 (IEC, Geneva).
- International Organization for Standardization (1998). "Acoustics-Reference zero for the calibration of audiometric equipment-Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," ISO 389-1 (ISO, Geneva).
- International Organization for Standardization (1998). "Acoustics-Reference zero for the calibration of audiometric equipment-Part 5: Reference equivalent threshold sound pressure levels for pure tones in the frequency range 8 kHz to 16 kHz," ISO/TR 389-5 (ISO, Geneva).
- Schmuziger, N., Probst, R., and Smurzynski, J. (2004). "Test-retest reliability of pure-tone thresholds from 0.5 to 16 kHz using Sennheiser HDA 200 and Etymotic Research ER-2 earphones," *Ear Hear.*, **25**, 127–132.
- Stelmachowicz, P. G., Beauchaine, K. A., Kalberer, A., and Jesteadt, W. (1989). "Normative thresholds in the 8- to 20-kHz range as a function of age," *J. Acoust. Soc. Am.* **86**, 1384–1391.

E. Moran

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

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

Dissertation Abstract

Distribution, patchiness, and behavior of Antarctic zooplankton, assessed using multi-frequency acoustic techniques [43.30.Sf, 43.20.Fn, 43.30.Ft]—Gareth L. Lawson, *Joint Program in Oceanography/Applied Ocean Science and Engineering, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Cambridge, MA, June 2006 (Ph.D.)*. The physical and biological forces that drive zooplankton distribution and patchiness on the Antarctic continental shelf were examined, with particular emphasis on the Antarctic krill, *Euphausia superba*. Acoustic, video, and environmental data were collected during fall and winter surveys of the Marguerite Bay region in 2001 and 2002. Improved parametrization of a theoretical model of krill target strength was achieved through direct measurement of all model parameters. Methods were developed and verified for

acoustically distinguishing krill aggregations from other zooplankton, and estimating krill length, abundance, and biomass. Application of these methods to multifrequency survey data demonstrated strong seasonal, interannual, and spatial variability in the distribution of both krill and overall zooplankton biomass. Highest krill biomass was consistently associated with regions close to land where temperatures at depth were cool. The vertical position and density of individual krill aggregations varied with time of day, food availability, and the occurrence of predators, suggesting that aggregation and diel vertical migration represent a balance between avoiding visual predators and accessing shallowly distributed food resources. These findings have important implications to the fields of zooplankton acoustics and Antarctic krill ecology, especially in relation to the interactions of the krill with its predators.

Advisors: Peter H. Wiebe and Timothy K. Stanton

USA Meetings Calendar

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

2007

- 4–8 June 153rd Meeting of the Acoustical Society of America, Salt Lake City, UT [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>].
- 22–24 Oct. NoiseCon 2007, Reno, NV [Website is: www.inceusa.org/nc07]
- 27 Nov.–2 Dec. 154th Meeting of the Acoustical Society of America, New Orleans, LA (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: <http://asa.aip.org>].

2008

- 29 June–4 July Acoustics 08 Joint Meeting of the Acoustical Society of America, European Acoustics Association and the Acoustical Society of France, Paris, France [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>].
- 28 July–1 Aug. 9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBen, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBen 9, P. O. Box 1609, Groton, CT 06340-1609; Tel.: 860-572-0680; Email: icben2008@att.net; WWW: www.icben.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).

REVISION LIST

New Associates

Alexander, Julia B., Flat 6, 23 Down St., London W1J 7AR, UK.
Barboza, Michael, Paulus, Sokolowski and Sartor, LLC, 67A Mountain Blvd. Ext., Warren, NJ 07928.

- Beard, Paul C., Medical Physics, University College London, Gower St., London WC1E 6 BT, UK.
- Berkenpas, Eric J., Hydro Technologies, 700 Automation Dr., Unit C, Windsor, CO 80550.
- Boek, Wilco, Underwater Technology, TNO Defence, Security and Safety, Oude Waalsdorperweg 63, The Hague 2597AK, The Netherlands.
- Buaka Muanke, Paulin, Mechanical Engineering, Univ. of Massachusetts, Lowell, 1 University Ave., Lowell, MA 01854.
- Buret, Marc M., Supreme Environmental Research Ltd., Rm. 2394, Universal Trade Ctr., 3-5A Arbuthnot Rd., Hong Kong, China.
- Cook, Justin W., Wyle Laboratories, Inc., 128 Maryland St., El Segundo, CA 90245.
- Crothers, Sarah, 332 Walden Creek Way, Greenville, SC 29615.
- Curra, Francesco P., Ctr. for Industrial and Medical Ultrasound, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105.
- Denes, Samuel L., Bioacoustics, Hubbs-SeaWorld Research Inst., 2595 Inghram St., San Diego, CA 92109.
- Ferreira, Tiago M. D., Rue Dr. Sa Carneiro, No. 41A, 3 Esq. Frente, Sao Joao da Madeira, 3700-255, Portugal.
- Fujimori, Hidetoshi, JAMSTEC/MEDID, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan.
- Gedamke, Jason, Australian Government Antarctic Div., Australian Ctr. for Appl. Marine Mammal Sci., Chennal Highway, Kingston TAS 7050, Australia.
- Gedemer, Linda A., 5930 Penfield Ave., Woodland Hills, CA 91367.
- Grassi, Elena, Univ. of Maryland, 3355 AV Williams Bldg., College Park, MD 20742.
- Grimes, Douglas K., 3550 Shirley St., Walled Lake, MI 48390.
- Hagler, Gina, 9710 Traville Gateway Dr., #255, Rockville, MD 20850.
- Hanjun, Liu, Communication Sciences and Disorders, Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208.
- Javitch, Ronald A., P.O. Box 67, Station H, Montreal QC H3G 2K5, Canada.
- Jenkins, Adam C., The Greenbusch Group, 1900 West Nickerson St., Ste. 201, Seattle, WA 98119.
- Jonsson, Joshua L., Maxxon Corp., P.O. Box 253, Hamel, MN 55340.
- Keane, Michael F., Keane Acoustics, 144 Annwood Rd., Palm Harbor, FL 34685.
- Ko, Wing P., Transportation Business Group, CH2M Hill, 555 South Flower St., Ste. 3550, Los Angeles, CA 90071.
- Lang, William H., NSF, Ocean Sciences, 4201 Wilson Blvd., Arlington, VA 22230.
- Latranyi, Erik M., Hendrick Mfg., Perforating Div., 1 Seventh Ave., Carbondale, PA 18407.
- Laugier, Pascal P., Lab. Imagerie Parametrique, Univ. Pierre et Marie Curie Paris 6, Paris 75006, France.
- Leff, Brian M., TaylorMade-adidas Golf Co., 5545 Fermi Ct., Carlsbad, CA 92008.
- Leighty, Katherine A., Disney's Animal Kingdom, Animal Programs Administrations, 1200 N. Savannah Cir., East, Bay Lake, FL 32830.
- Levins, Jennifer, 35 South White Horse Pike, Apt. 112, Audubon, NJ 08106.
- Li, Zhaung, 900 Pump Rd., Apt. 76, Richmond, VA 23238.
- Lim, Hock Siang, Bk. 750 Woodlands Ave. 4, #12-327, Singapore 730750, Singapore.
- Makin, Simon J., Psychology, Univ. of Reading, Harry Pitt Bldg., Earley Gate, Reading, Berkshire, RG6 6AL, UK.
- Manouchehri, Daneh, PSOMAS, 3187 Red Hill Ave., Ste. 250, Costa Mesa, CA 92626.
- McDonnell, Deborah, 16707 Governor Bridge Rd., Bowie, MD 20716.
- Metrani, Prabhakara N., Monarch Self Adhesive Tapes & Foames I. Pvt. Ltd. 302 V. P. Rd., Jalhalli, Bangalore 560013, India.
- Miura, Masayoshi, Nijusseikigaoka Kakinoki-chou 48, Matsudo-shi, Chiba-ken 271-0088, Japan.
- Nakajima, Hirofumi, Honda Research Inst. Japan Co., Ltd., Honcho 8-1 Wako-shi, Saitama 351-0188, Japan.
- Nathevon, Nicholas, Biology Dept., Univ. Jean Monnet, 23 rue Michelin, Saint-Etienne, Cedex 02, 42100, France.
- Neale, William T. C., Kineticotp LLC, 44 Cook St., Ste. 510, Denver, CO 80206.
- Nickell, Joe R., AV Plus Design, Inc., 10500 Mohawk Ln., Leawood, KS 66206.
- Park, Gyu-Tae, Next Future Co., Ltd., 148-1 Mabuk-dong, Giheung-gu, Yongin City, Gyeonggi-do, 446-798, Korea.
- Parker Stetter, Sandra, School of Aquatic and Fishery Sci., Univ. of Washington, Box 355020, Seattle, WA 98195-5020.
- Paul, John D., School of Electrical and Electronic Engineering, Univ. of Nottingham, Tower Bldg., University Park, Nottingham NG7 2RD, UK.
- Peterson, David R., 56 Ridge Dr., Greenbrier, AR 72058.
- Philip, Trudy L., Johns Hopkins Univ., Applied Physics Lab., Natl. Security Tech. Dept., 11100 Johns Hopkins Rd., MS 8-220, Laurel, MD 20723-6099.
- Raimond, Andrew P., Psychology, Univ. of Reading, Harry Pitt Bldg., Earley Gate, Reading, Berkshire, RG6 6AL, UK.
- Sagren, Anders K., Bank of Brains, Leo Fender Rd. 1, Uppsala SE-75237, Sweden.
- Scardigli, Giovanni, TECNAV Systems s.r.l., Piazza Dante 19/20, Livorno, Toscana, 57121, Italy.
- Schimmelpenningh, Julia C., 12 Park St., North Brookfield, MA 01535.
- Seland, Bernt S., Alte Aue 24, Hamburg 21129, Germany.
- Sexton, Kenneth D., The Sexton Company LLC, 860 E St., NE, Salem, OR 97301.
- Steiner, Gerald, Inst. of Electrical Measurement and Measurement Signal Processing, Kopernikusgasse 24/4, Graz, Styria 8010, Austria.
- Su, Zuhre, Russ Berger Design Group, 4006 Beltline Rd., Ste. 160, Addison, TX 75001.
- Sussman, Elyse, Albert Einstein College of Medicine, Neuroscience, 1410 Pelham Pkwy, S., Bronx, NY 10461.
- Swist, Lawrence P., P.O. Box 263, Greenwood Lake, NY 10925.
- Turgeon, Martine, Psychology Dept., Lancaster Univ., Fylde College, Lancaster LA1 4YF, UK.
- Unger, Jeffrey S., Unger Technologies, Inc., 297 North 9th St., Noblesville, IN 46060.
- Vaccaro, Michael G., 932 Winterhaven Dr., Gambrills, MD 21054.
- Voskuil, Duane M., Voskuil Violins, 1002 North 8th St., Bismarck, ND 58501.
- Wang, Chung-Wu, 4F, No. 19, Ln. 18, Sec. 1, Hang-Chou S. Rd., Taipei, Taiwan 100, ROC.
- Wang, Yi (Kevin), Sound and Vibration Lab., Emerson Climate Technologies, 1765 Campbell Rd., Sidney, OH 45365.
- Wearing, Joseph B., 430 Navaro Place, Unit 220, San Jose, CA 95134.
- Wenceslas, Benson, Universal Specialties LLC, P.O. Box 807, Wadikabir, Muscat, PC 117, Oman.
- Woolworth, David S., Oxford Acoustics, Inc., 356 CR 102, Oxford, MS 38655.
- Yuan, Jiahong, Linguistics, Univ. of Pennsylvania, 255 S. 36th St., Williams Hall 609, Philadelphia, PA 19104.
- Zhang, Yang, Speech-Language-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr., SE, Minneapolis, MN 55455.
- Zinin, Pavel, HIGP SOEST, 2525 Correa Rd., Honolulu, HI 96822.

New Students

- Akhoun, Idrick, 4 place Jean Jaures, Lyon F-69007, France.
- Alghassi, Hedayat, Box 40 - 6335 Thunderbird Crescent, Vancouver, BC V6T 2G9, Canada.
- Allen, George P., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117.
- Anderson, Eileen B., 7817 Delray Dr., Little Rock, AR 72227.
- Anick, Jason P., Acoustics, Univ. of Hartford, 200 Bloomfield Ave., Hartford, CT 06117.
- Austin, Melanie, 128 Montreal St., Victoria BC V8V 1Y8, Canada.
- Bach, Roger A., 504 West 37th St., Hays, KS 67601.
- Barroso, Celia, 3556 Maplewood Ave., Los Angeles, CA 90066.
- Beaudry, Alan R., Acoustics and Music, Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117.
- Beyersdorr, Aaron L., 218 Porter Hall, 60 South College Dr., Platteville, WI 53818.
- Bodson, Anais, Universit Bochum, Universitätsstrasse ND 6/33, Bochum D-44780, Germany.
- Bonner, Derek M., 3744 Bianco Dr., Fort Collins, CO 80521.
- Bostron, Jason Bostron, 1164 Day Rd., Sykesville, MD 21784.

- Boswell, Kevin M., Oceanography and Coastal Sciences, Louisiana State Univ., LSU School of the Coast and Environment, Baton Rouge, LA 70803.
- Boyle, Kelly S., Dept. of Zoology, Univ. of Hawaii, 2538 McMarthy Mall, Edmondson 152, Honolulu, HI 96822.
- Boyle, Mark V., 65 Bridge St., Kilrea, Coleraine, Belfast, Derry, BT5 15RR, Ireland.
- Brekke, Edmund, Yishun St. 612, Blk. 605, #04-329, 760604, Singapore.
- Brunett, Acacia, Juniata College, Box 279, Huntingdon, PA 16652.
- Buckley, Kristi A., 8201 Edgewater Dr., Frisco, TX 75034.
- Burgos, Julian M., Univ. of Washington, 1122 NE Boat St., Seattle, WA 98105.
- Cao, Xiang, Hearing Lab., Univ. of Pennsylvania, 3401 Walnut St., Ste. 328C, Philadelphia, PA 18104.
- Carlin, Michael A., 72 White St., Apt. 5, Clinton, NY 13323.
- Chang, Chiung-Yun, Speech and Hearing Science, Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210.
- Charles, Daniel L., Graduate Program in Acoustics, Pennsylvania State Univ., 217 Applied Science Bldg., State College, PA 16802.
- Chen, Nancy F., Research Lab. of Electronics, MIT, 77 Massachusetts Ave., Rm. 36-545, Cambridge, MA 02139.
- Chester, Ryan T., Physics Dept., Brigham Young Univ., N283 ESC, Provo, UT 84602.
- Clair, Shaun E., Penn State Univ., 217 Applied Science Bldg., State College, PA 16802.
- Clark, Meghan, 3-212 Lamoine Village, Macomb, IL 61455.
- Crossley, Jennifer F., 114 Rhyddings Terrace, Swansea, W. Glamorgan, SA2 0DS, UK.
- Defoe, Jeff J., 133 Alexander Ave., RR 3, Belle River ON N0R 1A0, Canada.
- DeFreitas, Celeste, 11531 Janette Ln., Garden Grove, CA 92840.
- Del Giudice, Alex A., UCSD, Linguistics, 9500 Gilman Dr., #108, La Jolla, CA 92093-0108.
- DeMeo, William J., 535 Lipoa Parkway, Ste. 149, Kihei, HI 96753.
- Dick, David A., Univ. of Hartford, 200 Bloomfield Ave., Box 3768, West Hartford, CT 06117.
- Dobie, Stephen B., 2854 Owl Ave., Palm Harbor, FL 34683.
- Doyle, Jeffrey E., 28 Chestnut St., Belmont, MA 02478.
- Driscoll, Nancy A., 105 Chestnut Circle, Northport, NY 11768.
- Duke, Cole V., 160 S. Viewcrest Dr., Bountiful, UT 84010.
- Duke, Mikila, 225 South 900 West, Provo, UT 84601.
- Dunn, Alexandra L., 231 Blackburn St., #7, Santa Cruz, CA 95060.
- Edwards, Scott S., CETA, Univ. of Hartford, Box 1902, 200 Bloomfield Ave., West Hartford, CT 06117.
- Enyo, Yumiko, Linguistics Dept., Univ. of Hawaii, Manoa, 569 Moore Hall, 1890 East-West Rd., Honolulu, HI 96822.
- Epp, Bastian, Univ. of Oldenburg, Carl-von-Ossietzky Str., Oldenburg 26129, Germany.
- Epps, Sarah A., Marine Science, Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Ctr., MS 39529.
- Eshelman, Evan J., 6793 Mayflower Hill, Waterville, ME 04901.
- Ettlinger, Marc, UC Berkeley, Linguistics, 1203 Dwinelle Hall, Berkeley, CA 94720.
- Fisher, Ethan Z., Electrical and Computer Engineering, Ben Gurion Univ. of the Negev, Ben Gurion Rd., Beer-Sheva, Negev 84001, Israel.
- Friez, Nathan M., 1020 East 64th St., Sioux Falls, SD 57108.
- Fujisawa, Kazuhiro, Earthquake Research Inst., 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.
- Garnder, Joshua T., 13820 South 44th St., Apt. 1030, Phoenix, AZ 85044.
- Gaudreau, Marc-Andre, Ecole de Technologie Superieure, Genie mecanique, 1100 rue Notre Dame, Montreal QC H3C 1K3, Canada.
- Gautham, Mysore J., Music, Stanford Univ., The Knoll, 660 Lomita Dr., Stanford, CA 94305.
- Giacomoni, Clothilde, Box 3500, 200 Bloomfield Ave., West Hartford, CT 06117.
- Goetze, Stefan, Dept. of Communications Engineering, Univ. of Bremen, Otto-Hahn-Allee 1, Bremen 28359, Germany.
- Graetzer, Simone, 10/196 The Ave., Parkville, Melbourne VIC 3052, Australia.
- Hamilton, Christopher B., 72 West Chapel Rd., Asheville, NC 28803.
- Hamilton, Jamie L., 6596 S. Bellaire Circle, Centennial, CO 80121.
- Hay, Todd A., 1213 Kinney Ave., Austin, TX 78704.
- Heigl, Werner M., Apache Corporation, E&P Technology, 2000 Post Oak Blvd., Ste. 100, Houston, TX 77056.
- Helgadottir, Sigrun R., 2920 Colgate Rd., Madison, WI 53705.
- Hilton, Marshall S., 1001 Connecticut St., Lawrence, KS 66044.
- Hornbecker, Alex M., 200 Bloomfield Ave., Box 3931, West Hartford, CT 06117.
- Howell, Jonathan, Linguistics Dept., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853-4701.
- Hung, Renee, 15F-1 No. 15, Ln. 56, Xing-zhong Rd., Nangan, Taipei 115, Taiwan.
- Husztly, Csaba, Gyongyvira ut 15D, Budapest H1125, Hungary.
- Hwang, Jiwon, 83 Emily Dr., South Setauket, NY 11720.
- Ignatiuk, Daniel A., Box 1512, 200 Bloomfield Ave., West Hartford, CT 06117.
- Iloreta, Jonathan I., 965 The Alameda, Berkeley, CA 94707.
- Inverso, Danielle, ASLP, Gallaudet Univ., 700 Florida Ave., NE, Washington, DC 20002.
- Iyer, Ananth N., Temple Univ., Speech Lab., 1947 N. 12th St., Philadelphia, PA.
- Jeong, Jaehoon, 737 Olokele Ave., #1203, Honolulu, HI 96816.
- Jewett, Cayla M., 24915 SE 192nd St., Maple Valley, WA 98038.
- Jiang, Haisheng, Simon Fraser Univ., Linguistics, 8888 Univ. Dr., Burnaby BC V5A 1S6, Canada.
- Jin, Cheng, Zhejiang Univ., Room 209, New Bldg. of State Key Lab. of Fluid Power Transmission & Control, 38 Zheda Rd., Hangzhou, Zhejiang, 310027, P.R. China.
- Jing, Yun, 2425 21st St., Apt. 6, Troy, NY 12180.
- Kappus, Brian A., UCLA, Physics, 425 Portola Plaza, Los Angeles, CA 90095.
- Katseff, Shira, 1623 Josephine St., Berkeley, CA 94703.
- Khan, Sameer, UC Los Angeles, Linguistics, 3125 Campbell Hall, Los Angeles, CA 90095.
- Kilpatrick, Cynthia D., Linguistics, UCSD, 9500 Gilman Dr., #0108, La Jolla, CA 92093-0108.
- Kim, Miran, Linguistics, SUNY at Stonybrook, S201 SBS Bldg., Stonybrook, NY 11794-4376.
- Knoll, Monja A., Psychology Dept., Univ. of Portsmouth, King Henry Bldg., King Henry I St., Portsmouth PO1 2DY, UK.
- Krysl, Robb, 156 Porter St., Loft #155, East Boston, MA 02128.
- Kuntzman, Michael L., 230 Clover Ln., Monroe, IN 46772.
- Lacsamana, Cheryl M., 2341 Commonwealth Dr., Charlottesville, VA 22901.
- Le, Duy Nam, Inst. of Numerical Simulation, Univ. of Technology Hamburg, Schwarzenbergstrabe 95, Hamburg 21073, Germany.
- Lee, Hye-Sook, 306 East State St., Apt. 305, Ithaca, NY 14850.
- Lee, Nelson, 12210 Candy Ln., Saratoga, CA 95070.
- LePage, Tom, Rensselaer Polytechnic Inst., Architectural Acoustics, 110 8th St., Troy, NY 12180.
- Lindsay, Ken, 180 Ohio St., Ashland, OR 97520.
- Magliula, Elizabeth A., 969 West Main Rd., Apt. 3203, Middletown, RI 02842.
- Maki, David N., 610 Crestland Ave., Murfreesboro, TN 37130.
- Martei, Margaret, 7452 Matflower Hill Dr., Waterville, ME 04901.
- Matsugu, Yuka, East Asian Studies, Univ. of Arizona, Learning Services Bldg. 102, Tucson, AZ 85721-0105.
- Mayr, Andreas R., Hans Glueck Strasse 21, Peissenberg, Bavaria, 82380, Germany.
- Mazevski, Annette G., Dept. of Audiology & Speech-Lang. Path., Gallaudet Univ., 800 Florida Ave., NE, MTB 205, Washington, DC 20005.
- McGee, Kathryn F., 9228D Regents Rd., La Jolla, CA 92037.
- Miller, Nicola A., 11 St. Devenick's Place, Aberdeenshire AB15 9LN, Scotland.
- Mueller-Blenkle, Christina M., Okerstr. 36, Berlin 12049, Germany.
- Munson, Cheyenne, Dept. of Psychology, Univ. of Iowa, E11 SSH, Iowa City, IA 52242.
- Neitzel, Richard L., Environmental & Occupational Health Sci., Univ. of Washington, 4225 Roosevelt Way, NE, Ste. 100, Seattle, WA 98105.
- Novak, Cathleen C., 11706 Elm St., Omaha, NE 68144.
- Oberreuter, Jorge, Univ. Austral de Chile, Escuela de Ingenieria Acustica, General Lagos 2086 - Campus Miraflores, Valdivia, Chile.

- Ocuppaugh, Jaclyn L., Dept. of Linguistics and Languages, Michigan State Univ., East Lansing, MI 48823.
- Okobi, Anthony O., Div. of Biology and Medicine, Brown Univ., Box 8461, Providence, RI 02912.
- Pacini, Aude F., Hawaii Inst. of Marine Biology, Marine Mammal Research Program, P.O. Box 1106, Kailua, HI 96734.
- Paiva, Adam V., Metropolitan Acoustics, 40 West Evergreen Ave., Philadelphia, PA 19118.
- Pappafotis, Jason M., 1110 Odom Ct., Smyrna, TN 37167.
- Patton, Jake S., 4305 NE 59th Terrace, Kansas City, MO 64119.
- Peters, Mils, CIRMMT, McGill Univ., Schulich School of Music, 555 Sherbrooke St., West, Montreal QC H3A 1E3, Canada.
- Pietruch, Rafal W., ul. Mila 12, Fabianki 87-811, Poland.
- Porter, Scott P., Graduate Program in Acoustics, Pennsylvania State Univ., 217 Applied Science Bldg., State College, PA 16802.
- Poulsen, Andrew J., 426 Trapelo Rd., #3, Belmont, MA 02478.
- Puikkonen, Panu, Dept. of Physics and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602.
- Rackers, Luke, 1074 S. Dahlia St., Apt. G-535, Glendale, CO 80246.
- Reeves, Olivia J., 4645 Stanton Rd., Colorado Springs, CO 80918.
- Rodriguez, Erik A., 2709 Redwood Ave., McAllen, TX 78501.
- Roe, Maren Helene, Broeset 157-12, Trondheim 2-7050, Norway.
- Rohrkasse, Sarah M., 2670 Darke Ct., Cincinnati, OH 45233.
- Roon, Kevin, New York Univ., Linguistics, 719 Broadway, 4th Floor, New York, NY 10013.
- Rypina, Irina I., 285 Sunrise Dr., Apt. 7, Key Biscayne, FL 33149.
- Schenck, William E., 5007 Congressional Way, Lawrence, KS 66049.
- Scherma, Amy C., ME and Acoustics, Univ. of Hartford, Box 2136, 200 Bloomfield Ave., West Hartford, CT 06117.
- Sell, Alexander W., One Univ. Heights, Asheville, NC 28804.
- Sellon, Jonathan Blake, 97 Easton Rd., Westport, CT 06880.
- Shen, Yi, 1610 S. Dorchester Dr., Apt. 3, Bloomington, IN 47401.
- Shilman, Molly S., UCLA, Linguistics, 3125 Campbell Hall, Los Angeles, CA 90095-1543.
- Shport, Irina A., Linguistics Dept., Univ. of Oregon, 1290 Univ. of Oregon, Eugene, OR 97403-1290.
- Shunsuke, Sakaguchi, Graduate School of Life Sci. and Sys. Eng., Kyusyu Inst. of Technology, 2-4 Hibikino, Wakamatsuku, Kitakyusyu, Fukuoka, 808-0196, Japan.
- Silver, Jonathan, 24 Cedar Hill Rd., Newtown, CT 06470.
- Slotman, Michael J., Physics Dept., Bethel Univ., 3900 Bethel Dr., #2293, St. Paul, MN 55112.
- Smith, Charles J., Univ. of Hartford, Box 3302, 200 Bloomfield Ave., West Hartford, CT 06117.
- Smith, Joel E., 3520 State Route 681 North, Albany, OH 45710.
- Smith, Stephanie M., Honors Tutorial College, 35 Park Place, Athens, OH 45701.
- Steele, Daniel L., 350 Memorial Dr., Cambridge, MA 02139.
- Stoakes, Hywel M., Univ. of Melbourne, School of Lang. and Ling., Cnr Gratten and Swanston St., Melbourne VIC 3010, Australia.
- Sugiyama, Yukiko, Linguistics Dept., Univ. at Buffalo, 609 Baldy Hall (North Campus), Buffalo, NY 14260.
- Symmonds, Alexandra J., Acoustical Engineering, Univ. of Hartford, 200 Bloomfield Ave., Box 1248, West Hartford, CT 06117.
- Torres, Carlos E., Malaya St. U2 Santa Clara, Guaynabo, PR 00969.
- Toscano, Joseph C., Psychology Dept., Univ. of Iowa, E11 SSH, Iowa City, IA 52242.
- Tu, Hsiao-Wei, Psychology Dept., Univ. of Maryland, College Park, Biology/Psychology Bldg., College Park, MD 20770.
- Tu, Juan, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105.
- Udovychchenkov, Ilya A., 285 Sunrise Dr., Apt. 7, Key Biscayne, FL 33149.
- Urban, Radoslaw G., Dept. of Vertebrate Zoology, Univ. of Wroclaw, Inst. of Zoology, Sienkiewicz St. 21, Wroclaw 50-335, Poland.
- Vogel, Carl K., Acoustics, Univ. of Hartford, Box 802, 200 Bloomfield Ave., West Hartford, CT 06117.
- Wagner, Alicia J., Architectural Engineering, Univ. of Nebraska, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182.
- Watanabe, Yuko, Linguistics Dept., Univ. of Arizona, Douglass 200E, Tucson, AZ 85721.
- Wilkie, Sonia L., 1501 SanMateo Ave., Mildura VIC 3500, Australia.
- Worf, Alan, Lewis & Clark College, 0615 SW Palatine Hill Rd., MSC2253, Portland, OR 97219.
- Wrobel, Pat L., 8582 Tullamore, Tinley Park, IL 60477.
- Xavier, John V., Luther College, Physics, 700 College Dr., Decorah, IA 52101.
- Xie, Zhengchao, 1120 8th St., Apt. C19, Tuscaloosa, AL 35401.
- Zendel, Benjamin R., Rotman Research Inst., Univ. of Toronto, Psychology, 3560 Bathurst St., Toronto ON M6A 2E1, Canada.
- Zybura, Jack A., Box 2726, 200 Bloomfield Ave., West Hartford, CT 06117.

New Electronic Associates

- Bader, Christopher, Kratyls Technologies, 298 High St., Medford, MA 02155.
- Batke, Jan-Mark, Digital Audio Laboratory, Karl Wiechert Allee 24, Hannover 30625, Germany.
- Bentall, Nathan, 32 Deanfield Rd., Oxford OX2 9DW, UK.
- Berghuis, Andrew P., Dept. Primary Industries & Fisheries, 16-32 Enterprise St., Bundaberg QLD 4670, Australia.
- Blomme, Erik, KATHO, VHTI, Doorniksesteenweg 145, Kortrijk WVI 8500, Belgium.
- Braun, Christopher B., Hunter College, Psychology, 695 Park Ave., New York, NY 10021.
- Cho, Tongjun, SK E&C, 192-18 Kwanhun-dong, Jongno-gu, Seoul 110300, Korea.
- Davis, Sara, NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677.
- Dorge, Filip, Achelmsstraat 37, Peer, Limburg 3990, Belgium.
- Doyle, Kenneth J., 40 NewCt. Rd., Topsham, Exeter, Devon, EX3 0BT, UK.
- Dunn, Charlotte A., Bahamas Marine Mammal Research Org., P.O. Box AB-20714, Marsh Harbour, AB Bahamas.
- Foster, Alvin, AMF Mail, 995 Blue Hill Ave., Boston, MA 02124.
- Gaines, Gerald C., 207 1/2 Wilson Ave., Swannanoa, NC 28778.
- Graebner, John E., 2 Woodland Rd., Short Hills, NJ 07078.
- Heintze, Olaf, Inst. of Composite Structures and Adaptive Systems, German Aerospace Ctr., Lilienthalplatz 7, Braunschweig 38108, Germany.
- Hunt, Randy E., Indiana Univ., Southeast, Biology, 4201 Grant Line Rd., New Albany, IN 47150.
- Johnson, Steven A., 1916 East Browning Ave., Salt Lake City, UT 84108.
- Krakauer, Alan H., Evolution and Ecology, Univ. of California, Davis, One Shields Ave., 2320 Storer Hall, Davis, CA 95616.
- Lafont, Jean-Pierre, Sente du bas du dos d'ane, Evrecquemont, Yvelines 78740, France.
- Luis Miguel, Ortega, Sedaam, Cerrada de Hidalgo 102, Pachuca, Hidalgo, 42000, Mexico.
- Mahajan, Sanjay K., 28834 W. King William Dr., Farmington Hills, MI 48331.
- McEowen, Edwin, 5986 E. Chapine Rd., Columbia City, IN 46725.
- McKnight, Ian J., Powerix Technologies LLC, 2000 Hogback Rd., Ste. 13, Ann Arbor, MI 48105.
- Menzel, Christoph, P.O. Box 1787, 53 Northwood Ln., New London, NH 03257.
- Metz, Kenneth H., 715 N. White Tail Dr., P.O. Box 949, Franktown, CO 80116.
- Nikolic, Igor, Norsonic AS, Postboks 24, Lierskogen, Buskerud, 3421, Norway.
- Rines, Steven H., Grabower Allee 2, Ludwigslust MV 19288, Germany.
- Romero, Iaia, Sun Power Energy, P.O. Box 123, Molins de Rei, Barcelona 08750, Spain.
- Sanchez-Garcia, Antonio, SAES, Advanced Studies & DSP, Carretera de la Algameca s/n, Cartagena, Murcia, 30205, Spain.
- Schuhmann, Jeremy S., 5206 Saratoga Ave., Chevy Chase, MD 20815.
- Slaughter, Julie C., Etrema Products, Inc., 2500 North Loop Dr., Ames, IA 50010.
- Smuelders, David M., Delft Univ. of Technology, Geotech., P.O. Box 5028, Delft 2600GA, The Netherlands.
- Stein, Ben P., American Inst. of Physics, Media & Gov't Relations, One Physics Ellipse, College Park, MD 21044.
- Stratman, James A., Laird Technologies, RF Engineering, 3425 North 44th St., Lincoln, NE 68504.
- Thomson, Philip, 64 Krantzcke Ct., Nicholls ACT 2913, Australia.

Woods, Terry D., FDA/Ctr. for Devices & Radiological Health, DSFM, 9200 Corporate Blvd., HFZ-170, Rockville, MD 20850.

Yoshikazu, Koike, Faculty of Engineering, Shibaura Inst. of Technology, 3-7-5 Toyosu, Kotoh-ku, Tokyo 135-8548, Japan.

Ziolkowski, Marcin, NDT Consultants Limited, Middlemarch House, Siskin Dri., Coventry, Warwickshire, CV3 4FJ, UK.

New Corresponding Electronic Associates

Uguru, Joy O., Univ. of Nigeria, Humanities Unit, Nsukka, Enugu 4100001, Nigeria.

Uskokovic, Milan, Horvacanska 17, Zagreb 10000, Croatia.

Reinstated

R. F. DeLaCroix, W. A. Friedl, M. D. Sneedon—*Members*

C. W. Beamer, IV, J. P. Boyczuk, B. D. Dushaw, R. H. Gifford, K. M. Kurowski—*Associates*

M. S. Kushwaha—*Electronic Associate*

Associates Elected Members

G. A. Clark, S. Cochran, M. A. Bahtiarian, C. Giguere, G. Haiat, S. R. Hertz, J. Hiatt, I. Hoffman, K. A. Johnson, W. Li, Y. Liu, X. Luo, D. J. Moretti, G. J. Orris, V. G. Petnikov, S. D. Pfeiffer, J. C. Preisig, K. D. Rolt, S. G. Saucier, R. Shrivastav, C. A. Storch, D. T. Walter, M. J. Wilkinson, D. G. Zeddies

Member to Electronic Associate

S. A. Rybak, K. A. Woo, S-C. Wooh

Student to Associates

B. E. Anderson, D. T. Bradley, E. Cosharek, T. Hunter, G. Kawai, S. M. Lulich, J. L. Miksis-Olds, A. J. Miller, K. Nagao, Y. Naka, J. N. Oswald, K. W. Schwartz, B. M. Shafer, J. Voix, I. L. Wilson, P. D. Henderson, T. Kozasa, E. T. Kusel, S. Rollins, R. D. White

Student to Electronic Associate

L-W. Hsieh, C. H. Wiese

Associate to Electronic Associate

E. Aden, S. B. Anderson, X. Chen, D. K. Dacol, K. Hata, J. A. Meyers, P. H. F. Nicholson, G. Petculescu, C-M. Wu

Electronic Associate to Associate

L. Bian, M. H. Burk, A. G. Petculescu

Associate to Student

A. M. Sinclair

Resigned

J. Binek, R. E. Douglass, H. Endo, A. Hasegawa, J. L. Jones, Jr., J. Rye, W. M. Sprinkle—*Members*

N. E. Amos, C. H. Chew, N. Chubachi, A. Fujikawa, G. P. Hentschel, D. Hofer, E. A. Kamp, J. Martini, E. McKeown, C. O. M. C. Piehler, P. T. Saussus, B. C. Schwartz, M. Schwartz, P. L. Schrock, S. Shull—*Associates*

C. Champagne, T. Jibiki, G. D. Lynde—*Electronic Associates*

D. Dorado, S. Loe, B. P. Ortega—*Students*

Deceased

R. L. Grason, J. Igarashi, N. Rott, F. N. Spiess, H. F. Tiersten—*Fellows*

D. B. Callaway, S. Feldman, W. L. Gessert—*Members*

P. R. Market, J. R. Ramsey, W. B. Schmidt—*Associates*

M. P. Beddoes—*Electronic Associate*

Fellows	911
Member	2450
Associates	2588
Students	1110
Electronic Associates	445
	7504

ACOUSTICAL STANDARDS NEWS

Susan B. Blaeser, Standards Manager

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

George S. K. Wong

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

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please, contact Susan B. Blaeser.

Comments are welcomed on all material in Acoustical Standards News.

This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, flyers on the draft international documents available, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: <http://asa.aip.org>.

Standards Meetings Calendar—National

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

Monday, 4 June 2007

- 7:00 p.m.—ASACOS Steering Committee

Tuesday, 5 June 2007

- 8:30 a.m.—ASA Committee on Standards (ASACOS)
- 10:30 a.m.—Standards Plenary Group (includes U.S. TAG to ISO/TC 43, TC 43/SC 1, and TC 29)
- 2:00 p.m.—Accredited Standards Committee S12, Noise

Wednesday, 6 June 2007

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

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

Tuesday, 19 June 2007

- 1:00 p.m.—Accredited Standards Committee S2, Mechanical Vibration, Shock and Condition Monitoring (includes U.S. TAG to ISO/TC 108 and its subcommittees)

New ASA Standards Website

The new ASA Standards website was launched this past fall. Here, you will find useful information related to the four national standards committees administered by ASA, as well as information on ASA-administered U.S. TAGs to various international standards committees. Current standards activities, meeting details, forms, and information on participation can also be found here. Visit us at <http://asa.aip.org/> and click the “Standards Info” button for more information. Please let us know what you think and what would make this site useful to you.

Hawaii Working Group Meetings

Several Working Groups met during the Honolulu meeting in December.

These are photos from a few of them with apologies to those we did not get to photograph. All working groups are reminded that ASA makes meeting rooms available at each ASA meeting. Please be sure to let the Secretariat know when you schedule a working group meeting.



S12/WG 41 Model Community Noise Ordinances.



S1/WG 5 Band Filter Sets.



Measurement subgroup of S12/WG 46 Acoustical Performance Criteria for Relocatable Classrooms.



S12/WG 11 Hearing Protector Attenuation and Performance.



S12/WG 15 Measurement and Evaluation of Outdoor Community Noise.

Call for Participation

Participation in each of the four Accredited Standards Committees and nine U.S. Technical Advisory Groups administered by ASA is open to all directly and materially affected parties. The members of the Committees and U.S. TAGs are companies, organizations, and government agencies. New members, particularly those representing the user interest category, are always welcome. Contact the Secretariat for information or visit our website.

Standards News from the United States

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

American National Standards Call for Comment on Proposals Listed

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

ARI (Air-Conditioning and Refrigeration Institute)

New Standards

BSR/ARI 370-200x, Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment (new standard)

Applies to the outdoor portions of factory-made commercial and industrial Large Outdoor Refrigerating and Air-Conditioning Equipment, including heat pumps, used for refrigerating or air-conditioning of spaces.

ASSE (ASC A10) (American Society of Safety Engineers)

New Standards

BSR A10.46-200x, Hearing Loss Prevention in Construction and Demolition Workers (new standard)

This standard applies to all construction and demolition workers with potential noise exposures (continuous, intermittent, and impulse) of 85 dBA and above.

TIA (Telecommunications Industry Association)

New Standards

BSR/TIA 1083-200x, Telecommunications—Telephone Terminal Equipment—Handset—Magnetic Measurement Procedures and Performance Requirements (new standard)

This standard defines measurement procedures and performance requirements for the handset generated 106 audio band magnetic noise of wireline telephones. A telephone complies with this standard if it meets the 107 requirements in this standard when manufactured and can be expected to continue to meet these 108 requirements when properly used and maintained.

Final Actions on American National Standards

ASA (ASC S12) (Acoustical Society of America)

New National Adoptions

ANSI S12.5-2006/ISO 6926:1999, Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels [Identical national adoption and revision of ANSI S12.5-1990 (R1997)]: 18 December 2006

ASTM (ASTM International)

Reaffirmations

ANSI/ASTM F956-1991 (R2006), Specification for Bell, Cast, Sound Signalling [Reaffirmation of ANSI/ASTM F956-1991 (R2001)]: 26 December 2006

ANSI/ASTM F957-1991 (R2006), Specification for Gong, Sound Signaling [Reaffirmation of ANSI/ASTM F957-1991 (R2001)]: 26 December 2006

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards.

The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

AGMA (American Gear Manufacturers Association)

BSR/AGMA 6032-200x, Standard for Marine Gear Units: Rating [Revision of ANSI/AGMA 6032-A94 (R2006)]

This standard provides rating practices for marine main propulsion, pump, and ship generator set service. The formulas evaluate gear tooth capacity as influenced by the major factors which affect gear tooth pitting and gear tooth fracture. It also addresses bearings, clutches, lubricating oil systems, shafts, and certain aspects of vibration. Project Need: To update the existing standard to reflect current design and gear rating practices. Stakeholders: Designers, manufacturers, and users of gear drives for marine service.

ISA (ISA)

BSR/ISA 60534-8-1 (75.07.03)-200x, Control Valve—Aerodynamic Noise Prediction (New standard)

Defines equipment, methods, and procedures for obtaining laboratory measurements of sound pressure levels radiated by control valves and/or associated piping configurations, including fixed restrictions, through which compressible fluids are passing. Project Need: Safety requirements, such as occupational health standards, require that human exposure to noise be limited. There are also data indicating that noise levels above certain levels could lead to pipe failure or affect associated equipment. Stakeholders: Consumers, manufacturers, regulatory bodies.

BSR/ISA 60534-8-4 (75.07.02)-200x, Control Valve—Hydrodynamic Noise Prediction (New standard)

Establishes a method to predict the noise generated in a control valve by liquid flow and the resulting noise level measured downstream of the valve and outside of the pipe. Project Need: Safety requirements, such as occupational health standards, require that human exposure to noise be limited. There are also data indicating that noise levels above certain levels could lead to pipe failure or affect associated equipment. Stakeholders: Consumers, manufacturers, regulatory bodies.

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

BSR C63.2-200x, Electromagnetic Noise and Field Strength Instrumentation, 10 kHz to 40 GHz—Specifications (New standard)

Incorporates CISPR 16, Part 1-1 as the US National Specification for field-strength instrumentation while keeping the unique requirements that are not in CISPR 16-1-1 and are still used by certain USA organizations. The USA requirements that are not in CISPR 16-1-1 are summarized and discussed in this Standard. Project Need: To incorporate CISPR 16-1-1 as the US national specification for field-strength instrumentation, with the addition of unique requirements that are not in CISPR 16-1-1 and are still

used by certain USA organizations. Stakeholders: EMC test laboratories, EMC test equipment manufacturers, product certifiers, product manufacturers.

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

This standard applies to both wireless communications devices (WDs) and hearing aids. It sets forth uniform methods of measurement and parametric requirements for the electromagnetic and operational compatibility and accessibility of hearing aids used with wireless communications devices, including cordless, cellular, Personal Communications Service (PCS) phones, and Voice over Internet Protocol (VoIP) devices. This version is focused on existing services, which are in common use. Project Need: To amend the currently published document, ANSI C63.19, with changes to selected paragraphs in the standard. Stakeholders: EMC test laboratories, wireless communications device manufacturers, hearing aid manufacturers.

BSR/IEEE 1478-200x, Standard for Environmental Conditions for Transit Rail Car Electronic Equipment (Revision of ANSI/IEEE 1478-2001)

This standard specifies baseline environmental conditions under which transit rail car electronic equipment shall both operate and/or survive. Special requirements anticipated for applications that have environmental conditions outside these baseline conditions are neither covered nor intended herein. Environmental conditions include temperature, humidity, atmospheric pressure, water, corrosive elements, and vibration/shock. These conditions depend upon equipment location such as under car, interior, truck mounted, and carbody mounted. Project Need: To provide a technical review and revision to several comments from one of the entities using the standard. Stakeholders: Rail transit industry throughout the U.S.

Standards News from Abroad

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

Newly Published ISO and IEC Standards

Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization.

ISO Standards

ACOUSTICS (TC 43)

ISO 140-18:2006, Acoustics—Measurement of sound insulation in buildings and of building elements—Part 18: Laboratory measurement of sound generated by rainfall on building elements

ISO 389-5:2006, Acoustics—Reference zero for the calibration of audiometric equipment—Part 5: Reference equivalent threshold sound pressure levels for pure tones in the frequency range 8 kHz to 16 kHz

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 10813-1/Cor1:2006, Vibration generating machines—Guidance for selection—Part 1: Equipment for environmental testing—Corrigendum

INDUSTRIAL FANS (TC 117)

ISO 13347-4/Cor1:2006, Industrial fans—Determination of fan sound power levels under standardized laboratory conditions—Part 4: Sound intensity method—Corrigendum, FREE

SMALL CRAFT (TC 188)

ISO 14509-2:2006, Small craft—Airborne sound emitted by powered recreational craft—Part 2: Sound assessment using reference craft, \$61.00

ISO Draft Standards

ACOUSTICS (TC 43)

ISO/DIS 3382-1, Acoustics—Measurement of room acoustic parameters—Part 1: Performance rooms (23 March 2007)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 10816-7, Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts—Part 7: Rotodynamic pumps for industrial applications, including measurements on rotating shafts (19 March 2007)

FLUID POWER SYSTEMS (TC 131)

ISO/DIS 15086-3, Hydraulic fluid power—Determination of the fluid-borne noise characteristics of components and systems—Part 3: Measurement of hydraulic impedance (8 March 2007)

IEC Standards

ELECTROACOUSTICS (TC 29)

IEC 60118-4 Ed. 2.0 b:2006, Electroacoustics—Hearing aids—Part 4: Induction loop systems for hearing aid purposes—Magnetic field strength

IEC 61672-3 Ed. 1.0 b:2006, Electroacoustics—Sound level meters—Part 3: Periodic tests

WIND TURBINE GENERATOR SYSTEMS (TC 88)

IEC 61400-11 Ed. 2.1 en:2006, Wind turbine generator systems—Part 11: Acoustic noise measurement techniques

ULTRASONICS (TC 87)

IEC 60565 Ed. 2.0 b:2006, Underwater acoustics—Hydrophones—Calibration in the frequency range 0.01 Hz to 1 MHz

IEC Draft Standards

44/540/FDIS, IEC 61310-1, Safety of machinery—Indication, marking and actuation—Part 1: Requirements for visual, acoustic and tactile signals (12 January 2007)

2/1416/FDIS, Amendment 1 to IEC 60034-14 Ed.3, Mechanical vibration of certain machines with shaft heights 56 mm and higher—Measurement, evaluation and limits of vibration severity (02 February 2007)

U.S. National Committee of the IEC

U.S. Proposals for Initiation of International Standards

TC 59—Performance of Household Electrical Appliances

The following proposal for the initiation of an International Standard has been submitted to the International Electrotechnical Commission:

TC 59: Performance of Household Electrical Appliances.

Title: Household and similar electrical appliances—Test Code determination of airborne acoustical noise for food waste disposers.

Scope: This technical standards proposal covers the requirements for the outline drawings and dimension of the above 10 Pin Micro Size Multimedia Card.

BOOK REVIEWS

P. L. Marston

Physics Department, Washington State University, Pullman, Washington 99164

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

Editorial Policy: *If there is a negative review, the author of the book will be given a chance to respond to the review in this section of the Journal and the reviewer will be allowed to respond to the author's comments. [See "Book Reviews Editor's note," J. Acoust. Soc. Am. 81, 1651 (May 1987).]*

Sounds in the Sea: From Ocean Acoustics to Acoustical Oceanography

Herman Medwin and colleagues

Cambridge University Press, Cambridge, 2005. 643 pp. Price: \$100 ISBN: 978-0-521-82950-2

I entered "Medwin" into the *Journal of the Acoustical Society of America's* search engine as I was preparing to write this review and was not surprised to get 269 citations. When sorted by date, the first is a 1953 paper by Hank Medwin and Isadore Rudnick on vorticity in acoustic fields. The last listed and most recent is a 2007 paper on the acoustic signature of bubbles from marine seeps by Ira Leifer and Dajun Tang in which they cited a paper by Hank. This list will continue to grow over the years as young investigators pour over his works and find that Hank laid the foundation for many of the subareas of acoustics including physical acoustics, ocean acoustics, and acoustical oceanography. One of his last contributions to the field before his passing is this book.

The book consists of four parts with 24 chapters. The first part is entitled "Fundamentals," written by Hank and having nine chapters in which he develops the fundamentals of ocean acoustics with emphasis on the physics. Part II deals with studies near the sea surface and consists of Chaps. 10 and 11. Part III covers bioacoustical studies and Chaps. 12–21 are included. Part IV on studies of ocean dynamics presents Chaps. 18–21 while Part V includes Chaps. 22–24 on acoustical studies of the ocean bottom.

Chapter 1 deals with sound propagation in a simplified sea. But not until 25 pages into the book is the wave equation developed. Instead, he begins with a description of Colladon and Sturm's experiment in 1827 in which they successfully measured the speed of sound in Lake Geneva. That priority given to experimental results is a hallmark of his texts. Again before the wave equation is derived, the reader is provided the World War II data collected by Eckart on Lloyd's mirror effect at sea and told to derive the interference pattern from the direct and scattered fields. Like all good teachers, he is inviting his students to consider how seemingly complicated effects can be explained by the application of first principles. Another example of Hank's teaching style is the first problem at the end of Chap. 1 asking the student to use a sketch and Huygens wavelets to derive the law of reflection for plane waves incident at a plane interface.

Chapter 2 on transmission and attenuation along ray paths is a basic description of rays and the effect of temperature gradients in ocean. Without depending on the eikonal and transport equations, he develops ray paths and tubes with emphasis on conservation of energy and continues with a description of sound channels in the low-latitudes and in the Arctic. His efficient and effective development is very reminiscent of his other texts, *Acoustical Oceanography: Principles and Applications* by C. S. Clay and H. Medwin and *Fundamentals of Acoustical Oceanography* by H. Medwin and C. S. Clay. Chapter 3 on sound sources and receivers again presents Hank's bias toward experimental results with the simplest theory needed to explain the data. Chapter 4 on nonlinear acoustics provides the reader with the simplest equations needed to explain cavitation, parametric sources, and explosives. Chapter 5 might be entitled ambient noise but instead is "Interpreting ocean sounds," reflecting Hank's belief that "Signals (or noise) are in the ear of the listener." He develops the concepts of sampling theory, frequency filtering, power spectra, and Wenz curves for ocean "noise."

Chapters 6–9 provide the "basic tools and analytical concepts" needed for sound scattering from marine bodies, bubbles, fish, the rough sea surface, and the ocean floor. In Chap. 8, normal modes in ocean waveguides are efficiently developed with a number of experimental descriptions to support the theory. Backscatter and forward scatter from the ocean surface is de-

scribed with an emphasis on the inverse problem. These chapters along with the earlier ones provide a sound foundation (sorry) for the remainder of the book where his colleagues span the field of acoustical oceanography.

David Farmer provides a survey of acoustical studies of the upper ocean boundary layer in Chap. 10. Acoustic sensing of wave directional spectra and bubble clouds are described. This mathematical level of the survey is consistent with the first nine chapters and accessible to the student wanting an introduction to the field. Again, experiments are presented with easily understood equations. Jeff Nystuen writes Chap. 11 on using underwater sound to measure raindrop size distribution. Measuring rainfall at sea is important for climatological studies but difficult for traditional sensors. Different raindrop sizes produce different sounds and acoustic rain gauges for ocean moorings are described.

Van Holliday and Tim Stanton provide Chap. 12 on active acoustical assessment of plankton and micronekton while John Horne and Michael Jech wrote Chap. 12 on fish backscatter. These chapters are very valuable to biological oceanography practitioners and students by giving them a fundamental description of the acoustic physics involved with scattering from these organisms. Orest Diachok in Chap. 14 on absorption spectroscopy provides a relatively new technique for remotely sensing the location and number of fish in the ocean.

Doug Cato, Michael Noad, and Robert McCauley provide Chap. 15 on passive acoustic monitoring of marine animals. By measuring vocalizations, one can study behavior, distributions, and movement. In addition, acoustical methods for estimating population sizes and rates of increase in numbers are presented in this enlightening chapter. Joe Blue and Ed Gerstein have Chap. 16 on the acoustical causes of collisions between marine animals and vessels. They postulate that the acoustic conditions near the sea surface are not allowing marine mammals to detect the approach of a vessel. The largest man-made cause of death for whales is shipstrike and this chapter provides the readers with the fundamentals of the problem and recommendation for mitigation. Ching-Sang Chiu and Chris Miller provide Chap. 17 on whale monitoring using acoustics. The technique described in this chapter detects, classifies, and counts the vocalizations of whales in an automatic manner. Blue whales are used as an example for a dereverberation technique for reconstruction and characterization of the marine mammal calls.

Bob Spindel wrote Chap. 18 on ocean acoustic tomography. After developing the radon transform and Fourier-slice theorem, Bob describes tomography for sensing the ocean mesoscale and other features. Chapter 19 by David Dowling and Heechun Song describes time reversal acoustics in the ocean. The technique exploits the multipath/multimode nature of ocean propagation. Daniela Di Iorio and Ann Gargett contribute Chap. 20 on studies of turbulent processes using Doppler and scintillation techniques. Part IV concludes with Chap. 21 on very high frequency coastal acoustics.

Part V on the ocean bottom has three chapters. Chapter 22 by David Palmer and Peter Rona presents acoustical imaging of deep hydrothermal vents. Nick Makris wrote Chap. 23 on using acoustics to image underwater mountain ranges. Mike Buckingham wrote Chap. 24 on acoustical remote sensing of the seabed using propeller noise from a light aircraft.

This is a text that every acoustical oceanographer should have on his/her shelf. The book, besides serving as a very good reference, is appropriate as a textbook for an upper level undergraduate or first year graduate course in ocean acoustics. Hank and his colleagues have worked hard to present this important material at a student's level.

JAMES H. MILLER

*Professor of Ocean Engineering and Oceanography,
University of Rhode Island, Narragansett, RI 02882
miller@uri.edu*

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice

11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

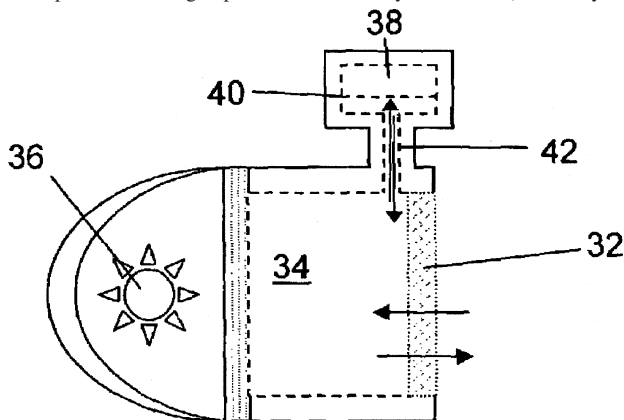
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
DIMITRI DONSKOY, *Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey 07030*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

7,106,445

43.20.Ye PHOTOACOUSTIC GAS SENSOR UTILIZING DIFFUSION

Robert E. Uber, assignor to Mine Safety Appliances Company
12 September 2006 (Class 356/432); filed 6 August 2003

A photoacoustic gas sensor based on cyclical heating and cooling of a gas sample 34 by a flashing light 36 detects a specific gas in environmental air. Air enters 34 by diffusion through porous plug 32 into illuminated volume 34. Pulsed heating occurs by selective photon absorption by the subject gas, if present. The light pulse rate is usually 8–20 Hz (the Golay cell



30

principle). The disadvantages are often that light impinging on a pressure sensing microphone within volume 34 will cause a false signal, and that the volume large enough to include the microphone will require greater amounts of air to diffuse into 34, requiring a long time to detect the subject gas. The author argues that a smaller sample chamber 34 is acceptable, and the microphone can be shielded from lamp light, since the pressure changes need

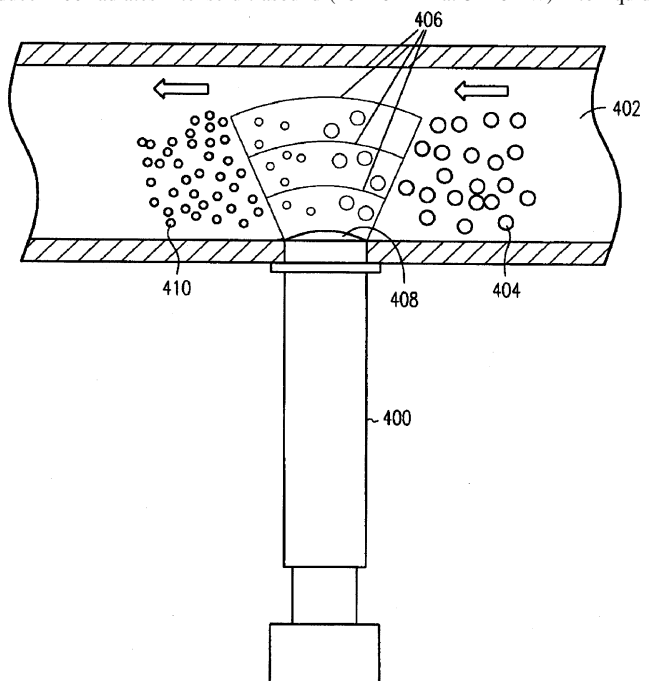
only pass through capillary 42 to diaphragm 40 of condenser microphone 38. A claimed advantage is that capillary 42 results in a faster response time because the gas sample need not also diffuse into the microphone volume 38–40.—AJC

7,101,691

43.25.Yw ALCOHOL PRODUCTION USING SONICATION

Michael T. Kinley *et al.*, assignors to UltraForce Technology LLC
5 September 2006 (Class 435/161); filed 26 August 2004

In preparation for fermentation of ground corn and other grains, transducer 400 radiates intense ultrasound (10–20 kHz at 3–10 kW) into liquid



402 containing ground grain particles 404. Intense cavitation in region 406 shakes off or strips away starch granules from protein bodies and fibers 410 to better expose them to yeast for digestion. The ultrasound level is limited so as to not degrade desirable proteins or damage yeast. Stages of such "sonication" are applied after grinding, steeping, and cooking cycles of the 402–410 liquid. Benefits include faster fermentation, increased ethanol yield (1%–10% increase), decreased additives and heating energy needed, and better stillage and wet cake by-products. (A bushel of #2 Yellow Dent Corn yields about 2.7 gallons of ethanol.)—AJC

7,109,637

43.35.Ns THIN-FILM BULK ACOUSTIC OSCILLATOR AND METHOD OF MANUFACTURING SAME

Eiju Komuro *et al.*, assignors to TDK Corporation
19 September 2006 (Class 310/320); filed in Japan 9 July 2003

This patent discloses refinements of the basic technique for manufacturing a film bulk acoustic resonator, or FBAR. The primary claim seems to be that polishing the faces of the resonator improves the Q.—JAH

7,091,869

43.35.Sx OPTOACOUSTIC MEASURING ARRANGEMENT AND USE THEREOF

Martin Forster and Peter Nebiker, assignors to Siemens Building Technologies AG
15 August 2006 (Class 340/628); filed in 9 October 2000

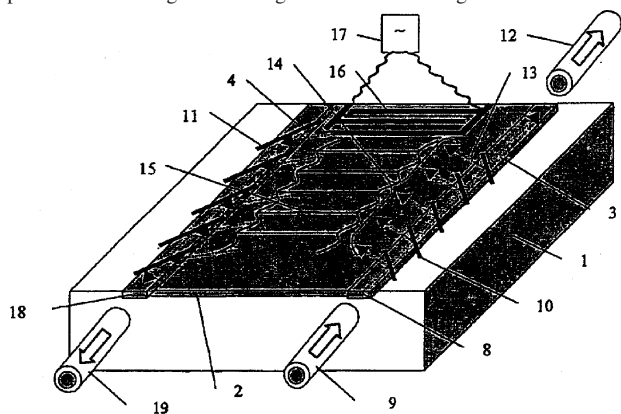
The device is an optoacoustic measuring arrangement for detection of gases and/or aerosols. The arrangement includes a measuring cell, a reference cell, microphones assigned to each cell, an electronic evaluation circuit connected to the microphones in which a subtraction of the signals of the microphones occurs, and a radiation source for applying a modulated signal to the measuring cell. The modulation frequency of the radiation source coincides with the resonant frequency of the measuring cell. This system can be used as a smoke or gas alarm or a fire hazard alarm, wherein each of the measuring cells is exposed to radiation of a wavelength which is absorbed by the substance to be detected, producing an optoacoustic effect as a result.—DRR

7,092,139

43.35.Sx ACOUSTO-OPTICAL VARIABLE FILTER

Andrei Vladimirovich Tsarev, Novosibirsk, Russian Federation
15 August 2006 (Class 359/285); filed in Russian Federation 4 August 2000

We have here an acousto-optical tunable filter consisting of planar optical waveguide 2 for propagation of light beam 14, arranged sequentially along the path of the light radiation, a means for forming a light beam, transducer 16 of acoustic waves suited to generate acoustic wave 15 that is capable of interacting with the light beam and causing diffraction within a



predetermined wavelength range, and a means for selecting the diffracted portion of the light beam. These light-beam-forming means and selecting means are implemented in the form of stripe waveguides 3 and 4, respectively, deployed in the plane of the planar waveguide, each provided with a set of unit reflectors 10 and 11 that overlap an aperture of corresponding stripe waveguide 3. The relative position and inclination angle of the reflectors are selected to maintain phase difference at the operating radiation wavelength for any pair of beams 13 reflected from different unit reflectors 10, which value is, essentially, a multiple of 2π .—DRR

7,090,333

43.35.Ty FOCUSED ACOUSTIC ENERGY IN THE PREPARATION OF PEPTIDE ARRAYS

Mitchell W. Mutz and Richard N. Ellson, assignors to Picoliter Incorporated
15 August 2006 (Class 347/46); filed 15 October 2002

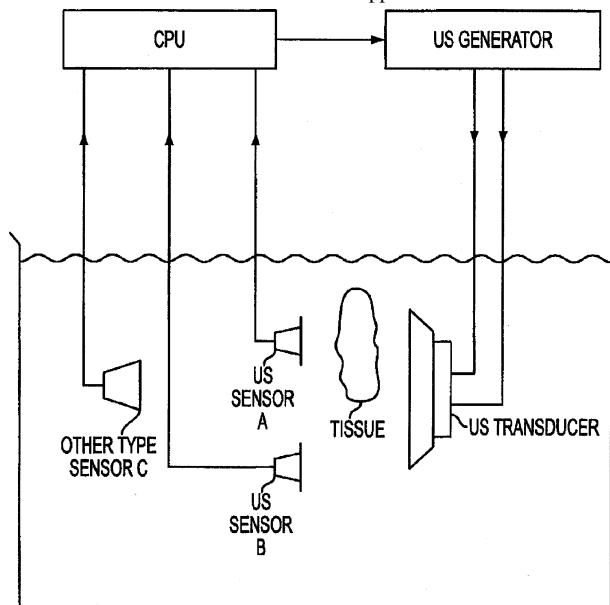
The patent document describes the use of focused acoustic ejection (FAE) technology in the preparation of peptide arrays. The arrays are prepared by acoustically ejecting peptide-containing fluid droplets from individual reservoirs toward designated sites on a substrate for attachment thereto.—DRR

7,090,974

43.35.Wa ULTRASOUND-MEDIATED HIGH-SPEED BIOLOGICAL REACTION AND TISSUE PROCESSING

Wei-Sing Chu, assignor to American Registry of Pathology
15 August 2006 (Class 435/6); filed 10 July 2001

We are presented with methods of fixing and processing tissue and samples on a membrane through the use of ultrasound radiation. Ultrasound in the frequency range of 0.1–50 MHz is used and the sample of the tissue receives 0.1–200 W/cm² of ultrasound intensity. The use of ultrasound requires much shorter times. The associated apparatus consists of transducers



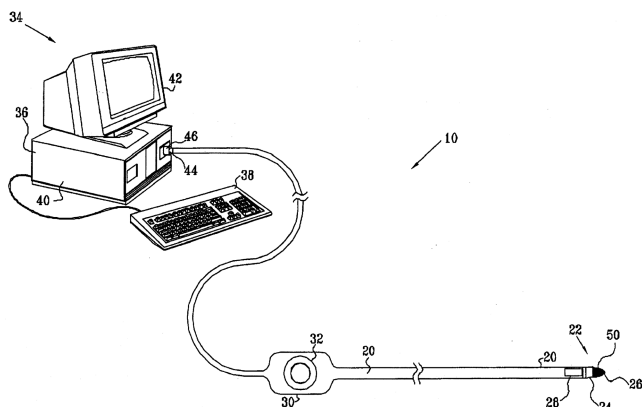
of one or more heads for producing the ultrasound radiation, a central processing unit, and optionally includes one or more sensors. Sensors can include those to measure and monitor ultrasound or temperature. This monitoring system is said to allow one to achieve accurate and optimum tissue fixation and processing without overfixation and tissue damage.—DRR

7,090,639

43.35.Yb ULTRASOUND CATHETER CALIBRATION SYSTEM

Assaf Govari, assignor to Biosense, Incorporated
15 August 2006 (Class 600/437); filed 29 May 2003

A means is provided for calibrating a probe having a position sensor and an ultrasonic transducer. A test fixture includes an ultrasonic target disposed, therein, at a known position. A computer receives a position signal generated by the position sensor while the transducer is in alignment with



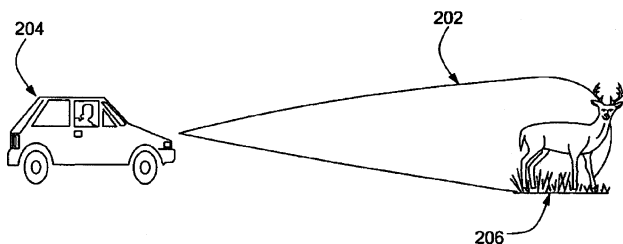
the ultrasonic target. It determines the orientation of the probe in the frame of reference of the test fixture and determines calibration data for the probe in response to the probe's orientation.—DRR

7,106,180

43.35.Yb DIRECTIONAL ACOUSTIC ALERTING SYSTEM

Frank Joseph Pompei, Wayland, Massachusetts
12 September 2006 (Class 340/384.7); filed 30 August 2002

The patent suggests the use of an audio modulated high-level ultrasonic carrier to produce a narrow-beam warning or alerting signal in a specific direction. "The parametric array of audio transducers operates by employing the nonlinear interaction between high-frequency sound components and the propagation medium to generate at least one highly directional



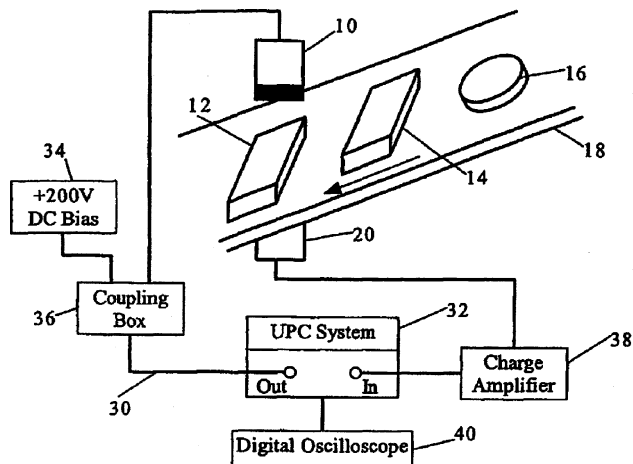
beam of lower frequency sound within the propagation medium." This basic "parametric" technique has received some recent television coverage, particularly in the area of possible riot control and a variety of military applications. The figure shows the technique in a more benign application.—JME

7,107,852

43.35.Yb METHOD OF INSPECTING FOOD STUFFS AND/OR ASSOCIATED PACKAGING

David Arthur Hutchins *et al.*, assignors to University of Warwick
19 September 2006 (Class 73/598); filed in United Kingdom 27 February 2002

This is a system to perform measurements on liquids, meat, viscous sugar, starch-based materials, or other food stuffs through the use of air-



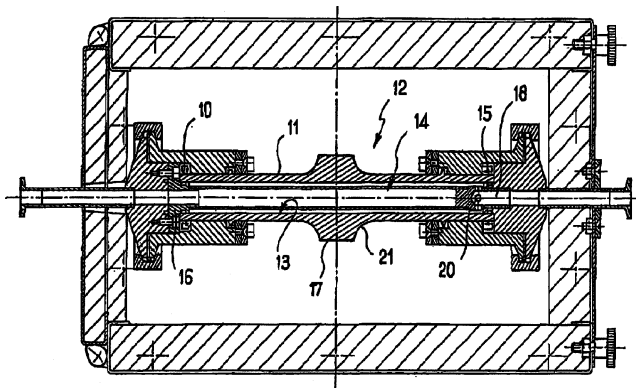
coupled ultrasound. The technique uses ultrasonic transducers (essentially capacitive transducers with polymer membranes) to generate ultrasonic signals in air and to receive these signals after they have passed through the material under test. Ultrasonic pulse compression is then applied to increase the sensitivity of signals transmitted through the materials. Such a method can be programmed to detect variations in the consistency of starch or sugar-based (viscous liquid) food stuffs, to detect liquid levels in polymer-based soft-drink bottles, or to detect foreign objects within food stuffs (such as bones within fish or meat).—DRR

7,101,485

43.35.Zc LIQUID PROCESS AND CONTINUOUS FILTERING DEVICE USING HIGH POWER DENSITY ULTRASOUNDS

Dominique Dubruque *et al.*, assignors to Sodeva
5 September 2006 (Class 210/748); filed in France 13 March 2001

This is a continuous filtration device in a liquid path. A modular ultrasonic treatment unit 12, which is in the form of a tubular metal body 11 with a cylindrical internal surface 13 and of circular cross section, is open at both its feed end 10 and its discharge end 15. The external surface of the said tubular metal body has, in the vicinity of the nodal zone, a collar 17 which is coaxial with the said tube and which projects radially; this collar being equipped at its periphery with an ultrasonic converter which projects radially. The frequency of the ultrasonic signal is equal to the vibration frequency of the said collar 17 and to the longitudinal vibration frequency of the said tubular metal body 11. A filtering cylindrical element 14, with a



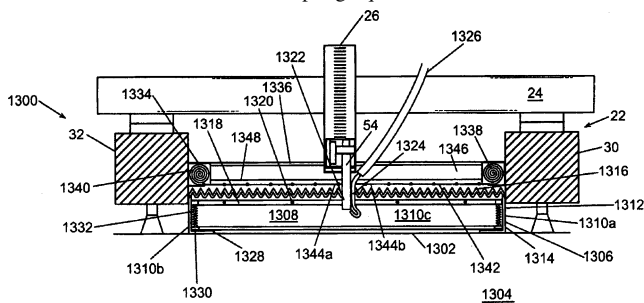
minimum filtration area of about 50 cm² and preferably about 80 cm² with a mesh opening less than about 20 μm, is arranged inside the tubular metal body 11. The nominal ultrasonic power density dissipated inside the tubular metal body is greater than about 2 W/cm² of filtering area.—DRR

7,104,132

43.35.Zc ACOUSTIC COUPLING WITH A FLUID BATH

Dennis William Mueller, assignor to MetScan Technologies, LLC
12 September 2006 (Class 73/620); filed 6 December 2004

Nondestructive inspection of the laminated skins of high performance aircraft requires a portable and flexible coupler for ultrasound transducers operating at 1–200 MHz. A liquid coupler at transducer 54 is claimed that transfers ultrasound by rod 1322 into liquid 1308 and onto the surface 1302 of skin 1304. Fresh bubble-free coupling liquid 1308 is introduced under rod



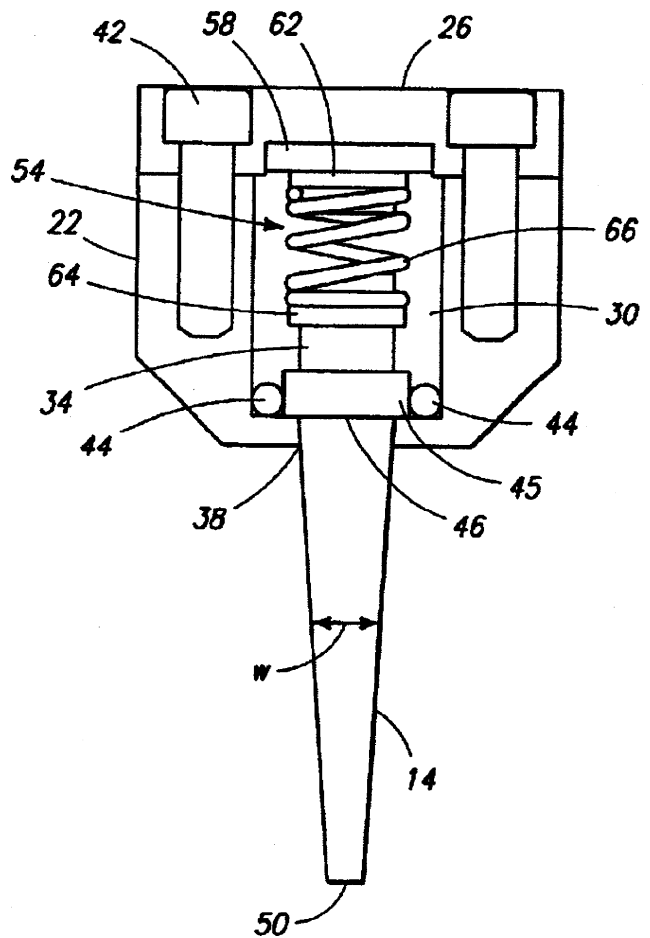
54 via pipe 1326–1324. A closed container 1310a–1310b retains a wide fluid pool 1308. Pliable materials 1328, 1330 maintain a liquid seal. A pleated cover 1316–1338 allows lateral movements of ultrasound emitting rod 54 for inspection of laminated skin 1302 over a wide skin surface area 1304 for a single instrument position 1300.—AJC

7,105,985

43.35.Zc MEGASONIC TRANSDUCER WITH FOCUSED ENERGY RESONATOR

Mark J. Beck and Richard B. Vennerbeck, assignors to Product Systems Incorporated
12 September 2006 (Class 310/323.19); filed 26 September 2001

This patent is concerned with the design of an ultrasonic cleaning probe for semiconductor cleaning. The authors maintain that the use of hard,



crystalline materials and a rectangular cross-section for the horn are novel, which is very doubtful. The design studies for this were done in the 1950s.—JAH

7,108,003

43.35.Zc ULTRASONIC CLEANING APPARATUS

Shigenori Kitahara, assignor to Tokyo Electron Limited
19 September 2006 (Class 134/184); filed in Japan 16 January 2002

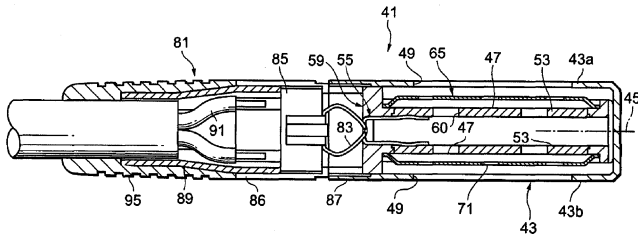
A number of ultrasonic transducers are arranged in the bottom wall of a cleaning tank. An ultrasonic oscillator powers the transducers through a switching unit, which switches the drive between a mode in which all the ultrasonic transducers are powered, and thus are excited, and a second mode in which only a part of the ultrasonic transducers are powered. The first mode is used when cleaning substrates not resistant to vibration and the second mode is used to dislodge substrates resistant to vibration.—DRR

7,106,869

43.38.Ar CONDENSER MICROPHONE

Hiroshi Kanda *et al.*, assignors to Sony Corporation
12 September 2006 (Class 381/174); filed in Japan 4 July 2002

By rearranging the form factor of a capacitor microphone diaphragm as a rectangle rather than the more conventional circular form, a fairly large overall diaphragm area can be accommodated into a narrow stem. The narrower body size may be an advantage in some applications, and of course,



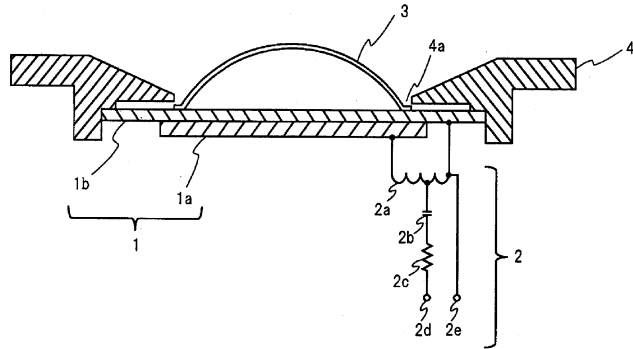
the larger total diaphragm area will improve the noise floor. The user, however, must be aware that the polar response of the microphone will be different in each orthogonal pickup plane.—JME

7,079,661

43.38.Fx SPEAKER FOR SUPER-HIGH FREQUENCY RANGE REPRODUCTION

Shoji Tanaka, assignor to Matsushita Electric Industrial Company, Limited
18 July 2006 (Class 381/190); filed in Japan 30 January 2002

Monomorph piezoelectric ceramic circular element 1a is bonded to thin (0.15 mm) circular brass element 1b to which a polyethylene terephthalate dome 3 is also attached as shown. Frame 4 is made from a rigid polystyrol



resin. Voltage boosting circuit 2 is connected between the free side of the piezoelectric element and the brass element. The best performance, from about 20 kHz to more than 100 kHz, appears to be achieved when the dome diameter is about 0.6–0.7 of the element diameter.—NAS

7,109,642

43.38.Fx COMPOSITE PIEZOELECTRIC APPARATUS AND METHOD

Walter Guy Scott, North Palm Beach, Florida
19 September 2006 (Class 310/334); filed 29 November 2004

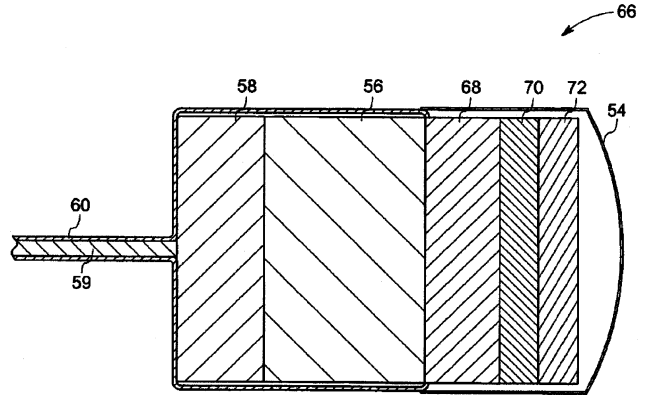
This patent teaches a variety of techniques for fabricating 1–3 piezoelectric composites using many tricks borrowed from the MEMS community. The text is densely packed with ideas and variations on those ideas, to the point where it is difficult to follow what is actually being accomplished in the fabrication process described. Nevertheless, it is quite informative to read, and mercifully brief.—JAH

7,105,986

43.38.Hz ULTRASOUND TRANSDUCER WITH ENHANCED THERMAL CONDUCTIVITY

Douglas Glenn Wildes et al., assignors to General Electric Company
12 September 2006 (Class 310/327); filed 27 August 2004

This ultrasonic transducer consists of a composite structure of backing material with enhanced thermal conductivity for use in the transducer. The



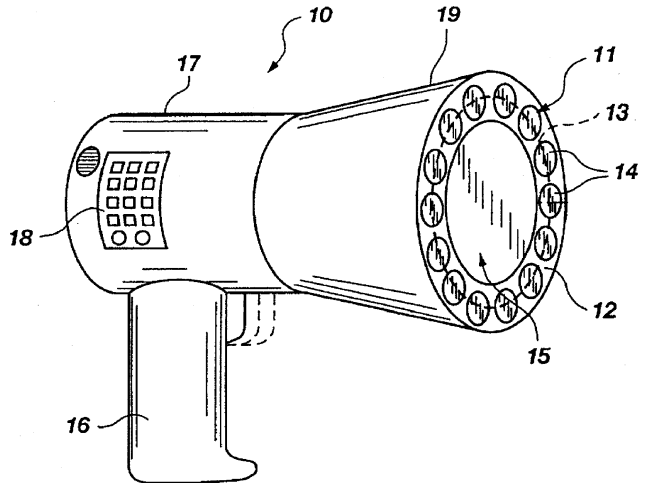
composite structure includes multiple layers of backing material alternatively arranged with multiple thermal conductive elements, in which the thermally conductive elements are configured to transfer heat from the center of the transducer to multiple points on the composite structure of the backing material.—DRR

7,088,830

43.38.Ja PARAMETRIC RING EMITTER

Elwood G. Norris et al., assignors to American Technology Corporation
8 August 2006 (Class 381/77); filed 18 March 2002

This sound emitting device generates one or more new sonic or subsonic signals by beating together two or more ultrasonic signals whose difference in frequency corresponds to the desired new sonic or subsonic frequency. The device includes one or more emitter segments positioned around a central open section structured with a diagonal width greater than



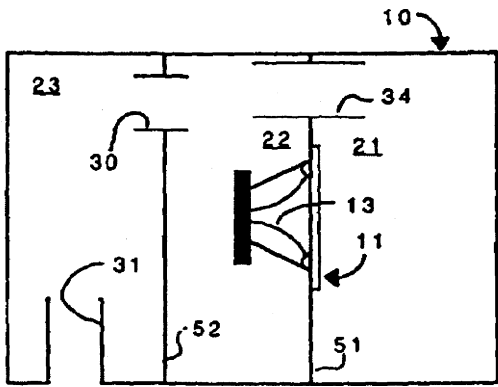
the cross-sectional diagonal of the parametric emitting perimeter. The ultrasonic frequency signal source and sonic/subsonic frequency generator are coupled together to a modulating circuit for mixing an ultrasonic signal with an electrical signal corresponding to the new sonic or subsonic frequency.—DRR

7,103,193

43.38.Ja BANDPASS WOOFER ENCLOSURE WITH MULTIPLE ACOUSTIC FIBERS

James J. Croft III, assignor to American Technology Corporation
5 September 2006 (Class 381/345); filed 14 September 2001

The intriguing title of this patent turns out to be a typographical error; it should read, "...multiple acoustic filters." What we have is another bandpass enclosure. Bandpass designs enjoyed great popularity about twenty



years ago and are still featured in a number of surround sound systems. Over the years, however, it has become apparent that their theoretical advantages are often overshadowed by practical shortcomings. The patent describes a triple-tuned arrangement (shown) which can be expanded to a quadruple-tuned system by adding another tuned chamber at the output of vent 31. Whether or not this geometry really represents an improvement over prior art, it certainly can be made to yield bandpass response.—GLA

7,106,865

43.38.Ja SPEAKER DIAGNOSTICS BASED UPON DRIVING-POINT IMPEDANCE

Peter M. Pavlov *et al.*, assignors to Motorola, Incorporated
12 September 2006 (Class 381/59); filed 15 December 2004

In loudspeaker manufacturing, quality control tests usually include impedance vs. frequency. The curve is easy to run, does not require a special acoustic environment, and can disclose a variety of defects. Impedance testing is also commonly used to verify that an installed loudspeaker has not

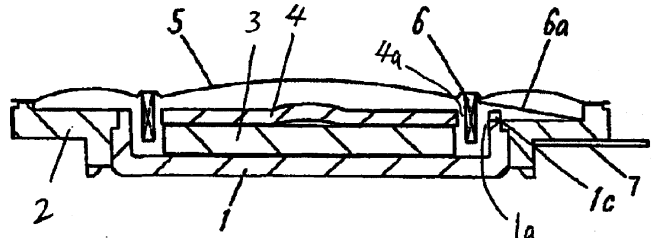
degraded over time. Since both of these applications are known prior art, trying to patent the process is a bit slippery. The 20 claims of this patent describe two features that might be considered nonobvious. The first is a method for performing a series of diagnostic functions. The second is the embodiment shown in which a second pickup coil is used to derive “driving point impedance.”—GLA

7,106,879

43.38.Ja LOUDSPEAKER

Hiroshi Yano, assignor to Matsushita Electric Industrial Company, Limited
12 September 2006 (Class 381/396); filed in Japan 7 February 2002

This shallow loudspeaker design includes notched portions in outer frame 1 to provide extra vertical clearance for voice coil leads 6a. If the



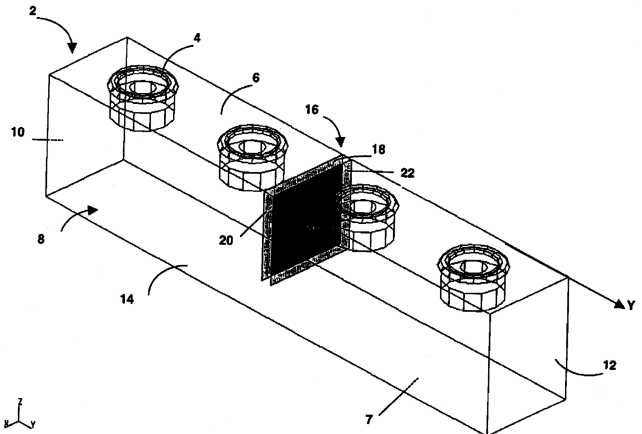
notched portions must be located asymmetrically, then a third notch can be added to balance the driving force on the voice coil.—GLA

7,111,706

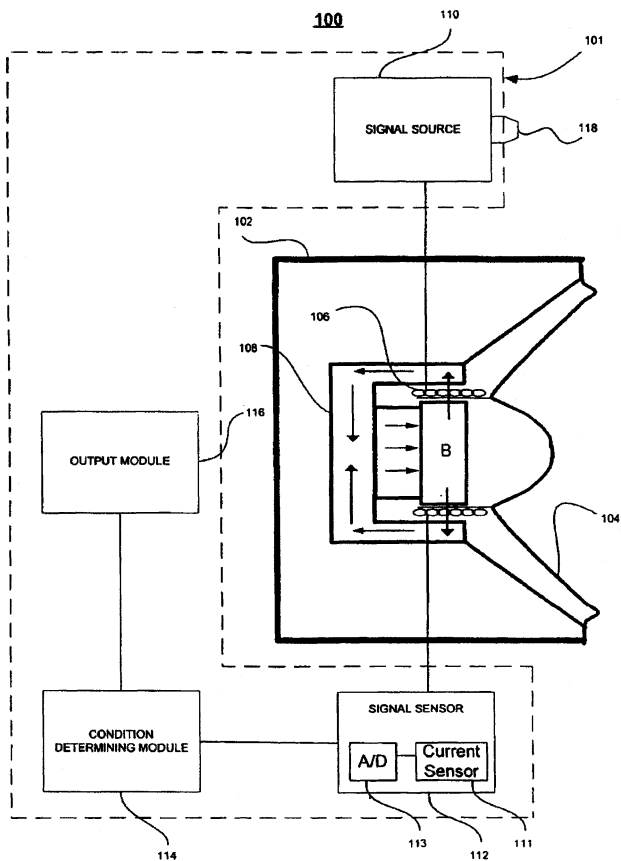
43.38.Ja ENCLOSURE AND AUDIO-VISUAL APPARATUS COMPRISING SAME

Gilles Bourgoïn, assignor to Thomson Licensing
26 September 2006 (Class 181/156); filed in France 24 January 2001

The inventor must be the only amateur loudspeaker builder in the world who is not aware of Thiele-Small alignments for vented boxes. The illustration shows four loudspeakers 4 in a simple box 2. A rectangular vent 18 is tuned to about 155 Hz. If low-Q loudspeakers are used in this oversized box, the result is an underdamped vent resonance producing a peak near 150 Hz. This situation has been analyzed at length, and popular “how to” books explain how vent damping can be added to smooth out response.



In this case, however, flexible plates are inserted to form dividers 20, 22. The patent asserts that “the plates 20, 22 produce an attenuation and limit the influence of the vent around the bass-reflex resonant frequency $f(0)$ without changing the value thereof, by partly absorbing the vibrations whose



frequency is close to the mechanical resonant frequency $f(M)$." Wrong. It is apparent now that the box houses a pair of lossy, highly nonlinear, passive radiators, with additional air-mass loading that substantially lowers the system resonance.—GLA

52

7,106,876

43.38.Kb MICROPHONE FOR SIMULTANEOUS NOISE SENSING AND SPEECH PICKUP

Richard J. Santiago, assignor to Shure Incorporated
12 September 2006 (Class 381/369); filed 14 October 2003

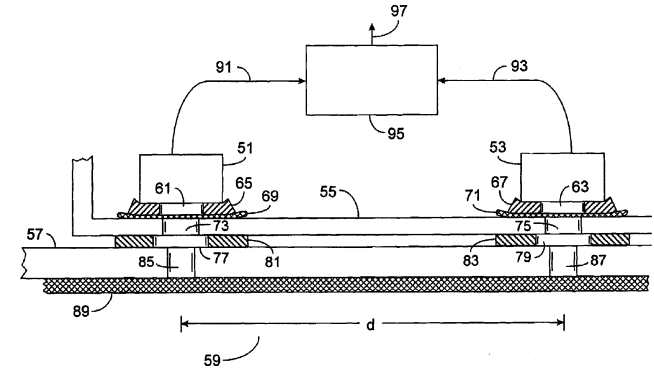
A dual-cartridge microphone simultaneously produces an omnidirectional pattern and a cardioid directional pattern from one microphone housing. The directional pattern is said to be suitable for picking up speech in hands-free cellular phone use and the omnidirectional pattern for ambient noise sensing—both in a vehicle.—DAP

7,110,553

43.38.Kb DIRECTIONAL MICROPHONE ASSEMBLY FOR MOUNTING BEHIND A SURFACE

Stephen D. Julstrom and Robert B. Schulein, assignors to Etymotic Research, Incorporated
19 September 2006 (Class 381/86); filed 3 February 1998

This well-written patent describes the mounting of dual omnidirectional elements laterally behind an acoustically transparent surface. The use of a pair of omnielements offers a number of advantages over a single element in producing a first-order directional pattern. Among these are



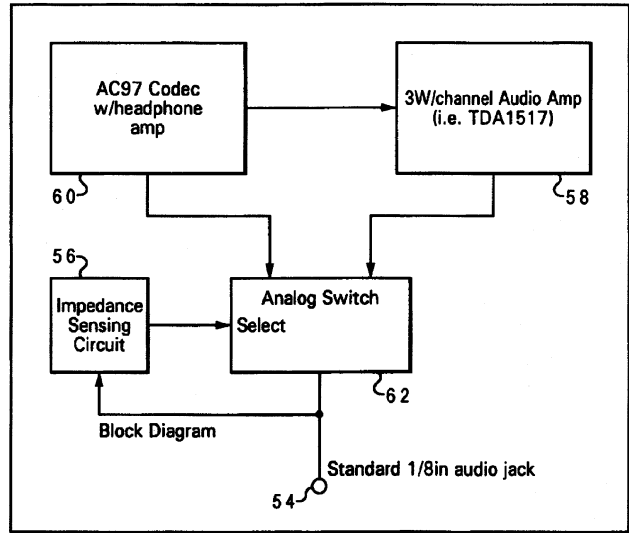
precise designing of the gradient distance, steering of the primary axis, and matching of low-frequency response. In the typical automobile cab, for example, there are ample opportunities for installing multiple assemblies as needed. The patent further presents relevant response curves for a number of applications.—JME

7,099,481

43.38.Lc METHOD AND SYSTEM FOR AUTOMATICALLY DETECTING AND POWERING PC SPEAKERS

Brian A. Baker et al., assignors to Lenovo (Singapore) Pte. Limited
29 August 2006 (Class 381/58); filed 15 May 2001

Integrated circuits are cheaper than switches or connectors. It follows that the speaker and line output jacks on a typical computer could be replaced by a universal output jack that senses what has been plugged in



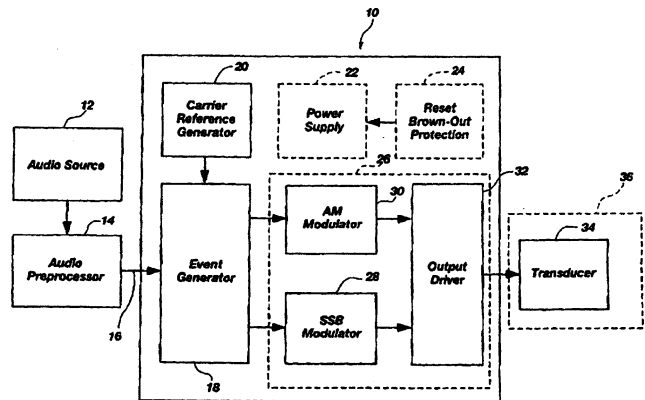
(headphone, powered speaker, passive speaker) and makes appropriate internal connections. Although the patent suggests a somewhat iffy practical implementation, the basic idea itself is worth considering.—GLA

7,109,789

43.38.Lc MODULATOR-AMPLIFIER

Michael E. Spencer, assignor to American Technology Corporation
19 September 2006 (Class 330/10); filed 21 January 2003

The invention described here is intended mainly for use with parametric loudspeakers—those that modulate an ultrasonic beam to produce sound from empty air. This is a long patent containing 38 illustrations, and it is somewhat difficult to follow. Nevertheless, those interested in this field are



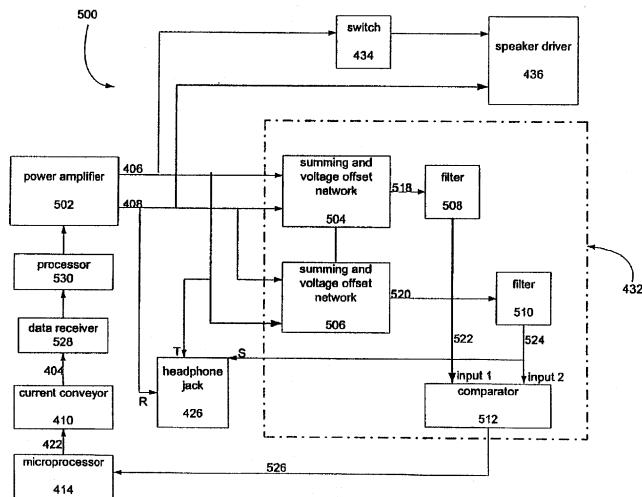
advised to obtain a copy. A switching-type combination modulator and amplifier is disclosed that operates in binary or trinary fashion. The main performance goals are improvements in power efficiency and audible signal quality.—GLA

7,110,839

43.38.Lc AUDIO SYSTEM FOR MINIMIZING THE CHANCE THAT HIGH POWER AUDIO SIGNALS MAY BE DIRECTED TO A HEADPHONE JACK

Bradley C. Wood, assignor to Harman International Industries, Incorporated
19 September 2006 (Class 700/94); filed 1 October 2001

This patent is a good example of how microchips have replaced knobs and wires in consumer electronics. Consider a typical computer sound package consisting of left and right loudspeakers plus a subwoofer that also houses electronic components. Signal lights, controls, and a headphone jack are normally located in the right loudspeaker. According to the patent, there are two things wrong with this arrangement. The first is the need for a bulky,



multiple-conductor cable between the subwoofer and the right loudspeaker. The second is that during the insertion of a headphone plug, speaker-level audio is momentarily connected to the plug tip before the mechanical changeover switch is activated. The patent discloses how a clever assortment of current sources, comparators, filters, and summing networks can reduce the number of wires to four and automatically switch to headphone operation almost instantaneously.—GLA

7,107,074

43.38.Md COMPUTER SYSTEM WITH WIRELESS AUDIO SIGNAL TRANSMITTER MODULE

Rong-Jung Lee, assignor to Mitac Technology Corporation
12 September 2006 (Class 455/557); filed in People's Republic of China 29 January 2003

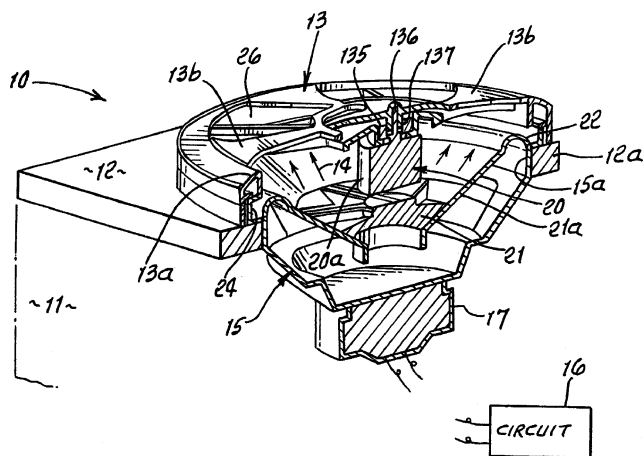
This patent describes the use of local two-way rf transmission between a personal computer and multiple users/listeners. In addition to dispensing with the usual collection of interconnecting cables, multiple channel capability increases the scope of "user multiplexing."—JME

7,085,392

43.38.Ne LOUDSPEAKER SOUND MODULATION APPLIANCE

Joseph K. Garretson, assignor to Scosche Industries, Incorporated
1 August 2006 (Class 381/162); filed 23 March 2004

Further development of the invention disclosed in United States Patent 6,978,031, reviewed in J. Acoust. Soc. Am. 120(2), 571 (2006), but not cited in this patent, provides a variant of an automobile aftermarket product that



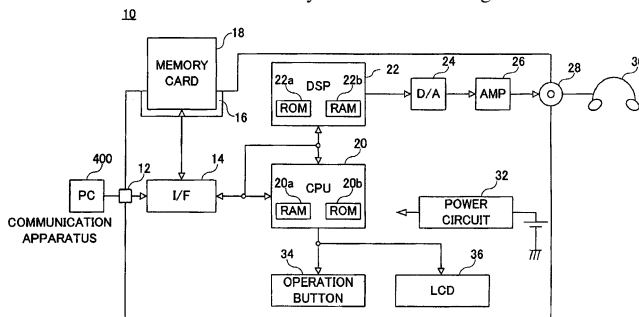
features a spinner element 13 mounted over loudspeaker diaphragm 15. The spinner is spun by motor 20. Item 17 is described as an actuator mechanism, but what it actuates is not clear to this reviewer. The idea is to modulate the sound emanating from the device.—NAS

7,103,431

43.38.Ne AUDIO PLAYER

Toru Kamimura et al., assignors to Sanyo Electric Company, Limited
5 September 2006 (Class 700/94); filed in Japan 25 August 1999

In this device, a sound reproducing apparatus includes a slot 16 into which a memory card 18 is inserted. Decoding software is incorporated in an audio file recorded in the memory card. The recording software is loaded



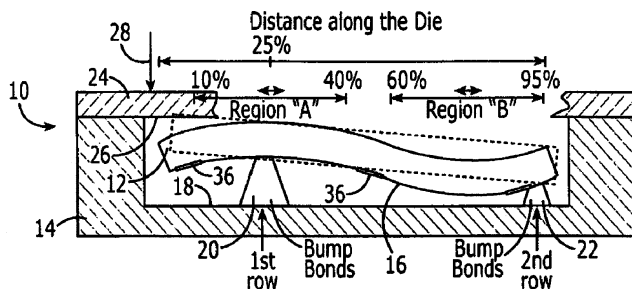
into a RAM 22b by a CPU 20 when it is not possible to decode the sound data by the decoding software stored on a ROM 22a. Therefore, a DSP 22 decodes sound data read from the same audio file in accordance with the decoding codeware in the audio file.—DRR

7,100,451

43.38.Rh SURFACE ACOUSTIC WAVE SENSING SYSTEM AND METHOD FOR MEASURING PRESSURE AND TEMPERATURE

Leland P. Solie, assignor to Sawtek, Incorporated
5 September 2006 (Class 73/703); filed 26 August 2004

SAW device 16 on substrate 12 is mounted on bumps 20, 22 located at 25% and 75% along evacuated package 14. Flexible membrane 24 closes the package volume and allows external pressure 28 to impinge on the back side of SAW substrate 12, deforming 12 into shape 16. Compression occurs about 20, increasing the SAW propagation speed in region A, while tension



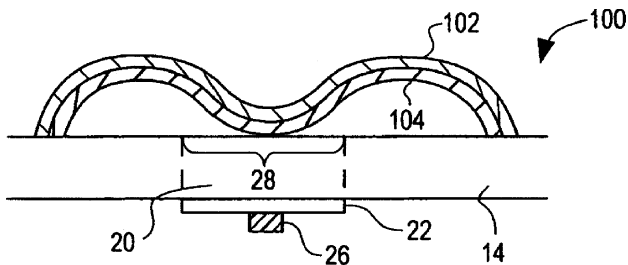
from position 16 to bump 22 decreases the propagation speed in region B. A SAW pulse is introduced in center electrode 36 and received at the electrodes at either end of the substrate. The difference in pulse arrival times at the two outer electrodes 36-36 is a measure of pressure 28, while the average propagation time is a measure of the temperature at substrate 12.—AJC

7,106,310

43.38.Rh ACOUSTIC WAVE TOUCH ACTUATED SWITCH

Terence J. Knowles *et al.*, assignors to TexZec, Incorporated
12 September 2006 (Class 345/177); filed 20 November 2001

A low cost, rugged, explosion-proof and touch-sensitive switch is claimed where a high-Q surface acoustic wave resonator cavity 20 is



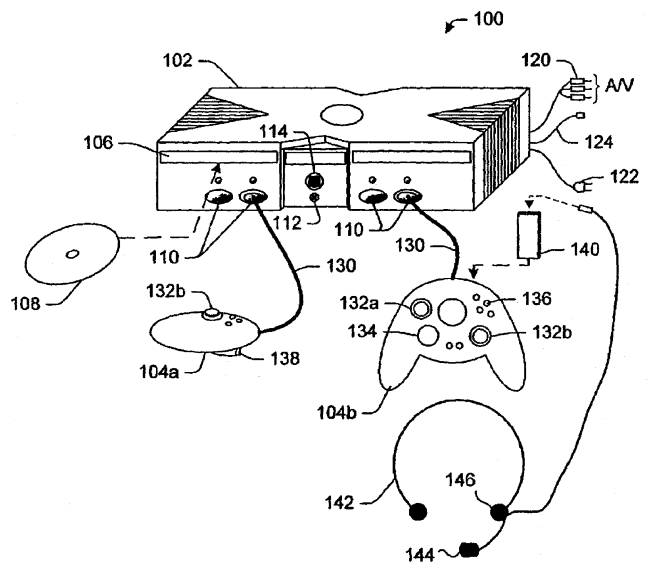
covered with a touch-deformable bubble seal 102 carrying acoustical absorber 104 on its interior surface. Pressed (on) condition is shown. Associated circuitry detects the change from resonant (off) to damped (on) states.—AJC

7,090,582

43.38.Si USE OF MULTIPLE PLAYER REAL-TIME VOICE COMMUNICATIONS ON A GAMING DEVICE

Damon V. Danieli and Roxana G. Arama, assignors to Microsoft Corporation
15 August 2006 (Class 463/35); filed 3 April 2003

The patent pertains to a multiplayer electronic game system that facilitates voice communication between players using one or more multiplayer electronic gaming devices, including voice communication over a network that conveys data between the coupled gaming devices. Verbal communications directed to one or more players are sampled and compressed in real-time, producing data packets that are transmitted to another game console. The data packets are decompressed and converted into an analog signal that drives the headphone of the intended recipient. Players can selectively mute voice communications to or from a specific player.—DRR

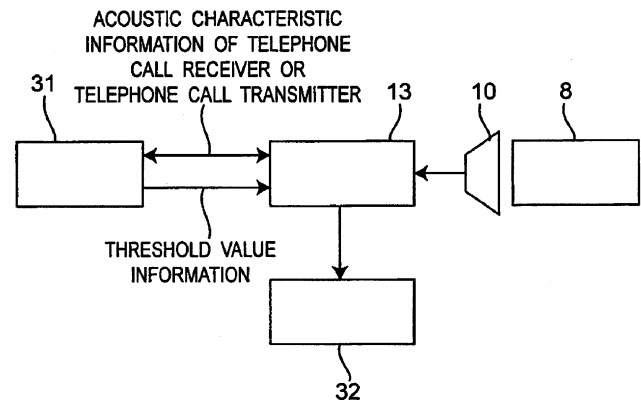


7,092,512

43.38.Si CALL CONTROL DEVICE AND CALL CONTROL METHOD

Makoto Kato and Masahiko Hashimoto, assignors to Matsushita Electric Industrial Company, Limited
15 August 2006 (Class 379/387.01); filed in Japan 10 April 2002

The object of this device is to provide easy connection and disconnection to a telephone line by detecting the presence of a phone user's ear or mouth instead of using a complex mechanism. The distance between a



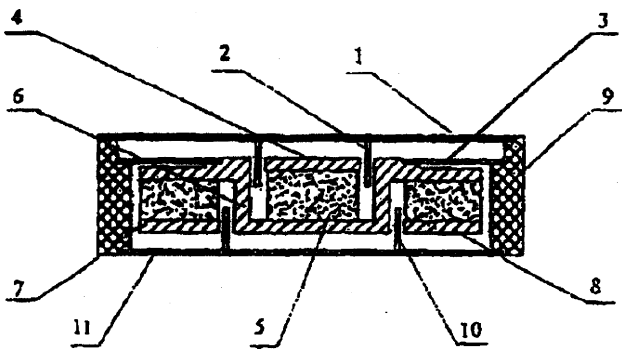
receiver and the call recipient's ear or the call sender's mouth is detected using acoustic impedance and reflective wave properties of the ear or mouth to control call operations, namely the connection and/or disconnection of the communication line.—DRR

7,099,489

43.38.Si TWIN MAGNETIC LOOP MULTIFUNCTIONAL VIBRATOR-SPEAKER TRANSDUCER

Zheng Min Pan *et al.*, assignors to American Audio Components Incorporated
29 August 2006 (Class 381/423); filed in People's Republic of China 15 April 2003

This patent describes a dual-function transducer for use in cellular telephones. Two nested magnets 5, 7 and bowl-shaped "magnetic transfer" 6 provide two concentric magnetic gaps in a single thin package. Voice coil



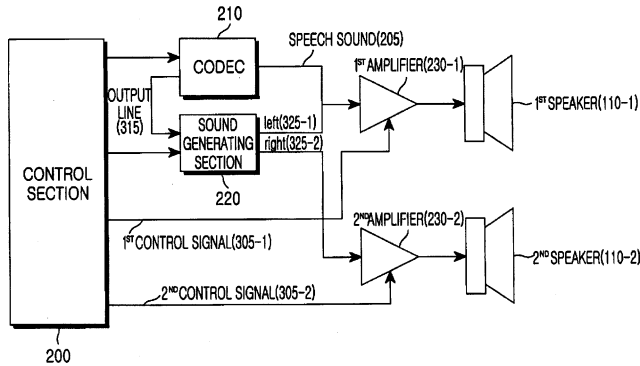
2 drives diaphragm 1 in conventional fashion to reproduce audio signals. Vibration coil 10 is fastened to rigid plate 11 and, thus, drives the entire magnetic assembly, which is supported by flexible spider 3. In an alternative embodiment, magnets, gaps, and coils are elliptical rather than circular.—GLA

7,103,393

43.38.Si SOUND OUTPUT SYSTEM AND METHOD OF A MOBILE COMMUNICATION TERMINAL

Seon-Woong Hwang, assignor to Samsung Electronics Company, Limited
 5 September 2006 (Class 455/569.1); filed in South Korea 29 December 2001

In this cellular telephone, normal speech communication is routed through amplifier 230-1 to miniature loudspeaker 110-1. When the unit is used to reproduce stereo music, a second channel consisting of amplifier



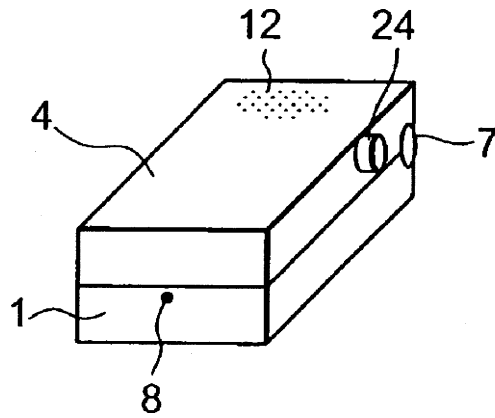
230-2 and loudspeaker 110-2 is brought into play. Thus, stereo playback capability does not compromise telephone usage either functionally or electronically.—GLA

7,103,395

43.38.Si FOLDING MOBILE TERMINAL DEVICE CAPABLE OF RECEIVING AND TRANSMITTING SOUND WHILE FOLDED

Ukyo Mori, assignor to NEC Corporation
 5 September 2006 (Class 455/575.1); filed in Japan 11 October 2001

A major concern in contemporary cellular telephone design is how to carry on a conversation when the case is closed. This NEC patent describes a no-nonsense approach. When the unit is closed, the user's voice is picked up at microphone opening 8 and incoming signals are reproduced by loud-



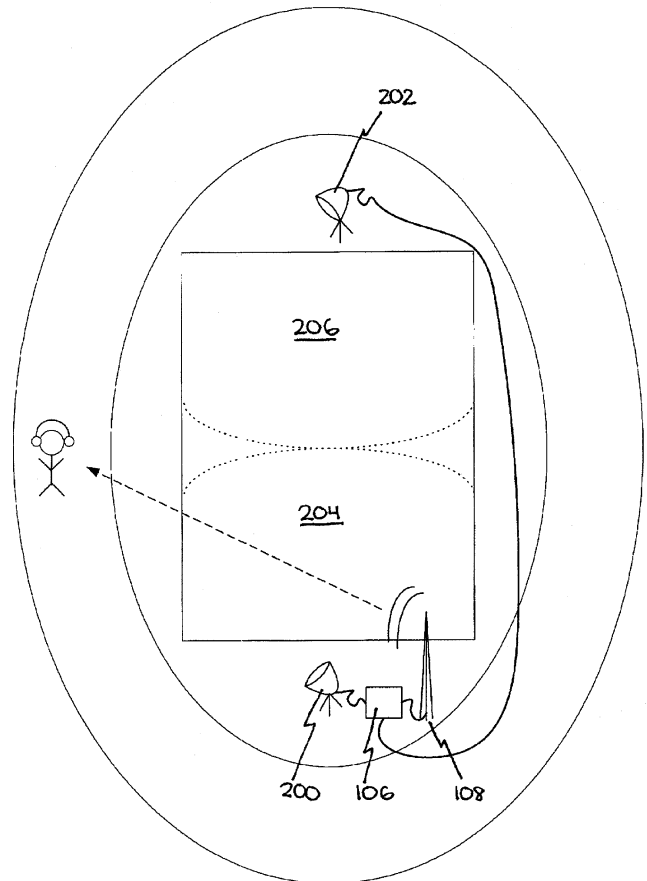
speaker 12. When the unit is opened, a second microphone opening is exposed and incoming signals are switched from the loudspeaker to a small receiver. Since the same microphone element is used in both modes, only one additional transducer is required.—GLA

7,110,552

43.38.Si PERSONAL LISTENING DEVICE FOR ARENA EVENTS

Mark Saliterman, assignor to Front Row Adv
 19 September 2006 (Class 381/77); filed 20 November 2000

This is among a new crop of "personal listening devices" for transmitting sound generated at an event to those in attendance at the event. This is accomplished by a system that collects an acoustic signal generated at a particular location within a fixed space while rejecting audio signals generated from outside the specific location. Finally, the system transmits the



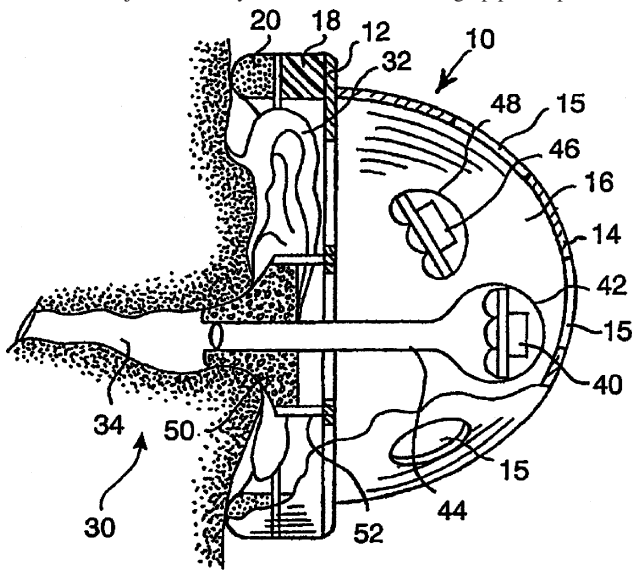
conditioned audio signal to receivers worn by individuals within the aforementioned fixed space. This is said to be particularly useful in settings within which a performance takes place, such as a football stadium, a basketball court, a hockey arena, a baseball stadium, a soccer field, a boxing ring, or a performance area in a restaurant or cruise ship.—DRR

7,110,563

43.38.Si HEADPHONE

Julien Laurent Bergere, assignor to Koninklijke Philips Electronics N.V.
19 September 2006 (Class 381/380); filed in 6 March 2002

The "Description of the Invention" section of this short patent consists mostly of self-congratulatory puffery, which is not unusual. In addition, the explanatory information makes no sense, which is unusual for this assignee. The basic idea is to combine an insert and a noninsert type headphone into an integrated two-way design. Low frequencies are reproduced by transducer 40 and injected directly into the ear canal through pipe 44, positioned



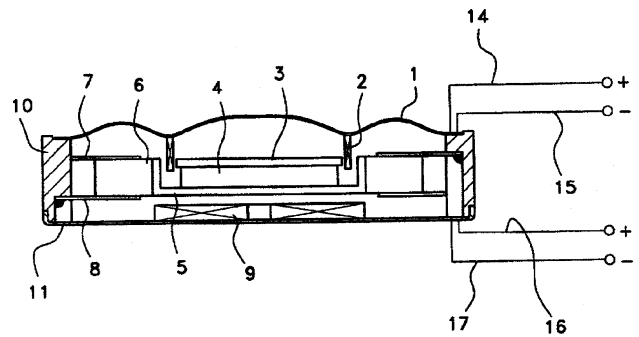
in the ear by a doughnut of open-cell foam 50. High frequencies are reproduced by transducer 46, in suitable proximity to the foam doughnut. The entire assembly is housed in cup 10, which can be completely enclosed or provided with apertures 15. According to the patent, the foam doughnut is transparent to high frequencies, yet, "at the same time provides through its sealing property, a pressure to be built up inside the auditory canal for the sound which is brought inside it through the tube." Well, you can't have it both ways; if the foam has appreciable acoustic resistance, then high frequencies will be transmitted to the inner ear through an acoustic lowpass filter. It is possible that this arrangement may offer certain benefits in performance, but they cannot be deduced from the patent document.—GLA

7,110,564

43.38.Si MULTI-FUNCTION ACTUATOR

Yeon Ho Son, assignor to Samsung Electro-Mechanics Company, Limited
19 September 2006 (Class 381/396); filed in South Korea 22 August 2001

This is a curious patent. Cellular telephones usually make use of a single, dual-function transducer to provide voice reproduction, as well as a vibrating (tactile) function. The design shown here has a diaphragm 1 driven by voice coil 2 operating in a magnetic gap energized by magnet 4—in other words, a conventional miniature loudspeaker. However, the entire magnetic assembly is supported by springs 7, 8 and can be driven at resonance by



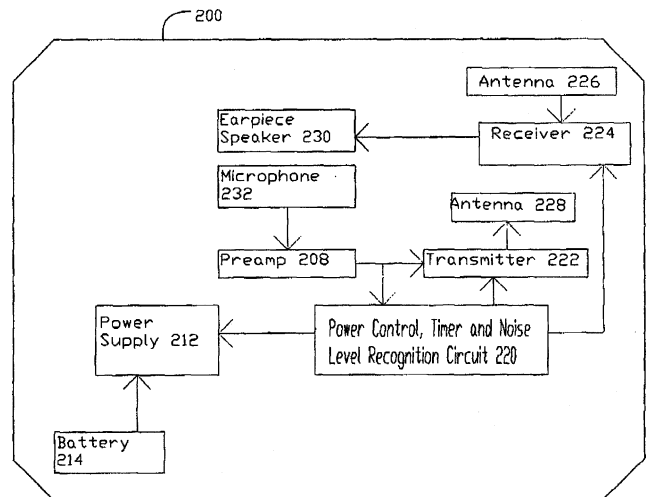
second coil 9 reacting with leakage flux below the magnet. It turns out that the novel feature is a notch filter or highpass filter (not shown) in the circuit driving voice coil 2. Thus, speech signals cannot excite the vibrating frequency. In a well-designed cellular phone one would expect such a highpass function to be included in the amplifier circuitry, obviating the need for a bulky, external filter. Yet, the patent claims are mostly taken up with descriptions of common L-R-C filter circuits.—GLA

7,110,801

43.38.Si VOICE ACTIVATED WIRELESS PHONE HEADSET

Shary Nassimi, Vancouver, Washington
19 September 2006 (Class 455/575.2); filed 17 October 2003

Safe usage of a cellular telephone or two-way radio in a motor vehicle requires hands-free operation. Because the noise level is high, a headset must be worn. However, if the user must operate a switch to receive an incoming call, safety has once again been compromised. Voice recognition has been proposed, but this patent argues that voice recognition circuitry would be too large and would use too much power for use in a wireless



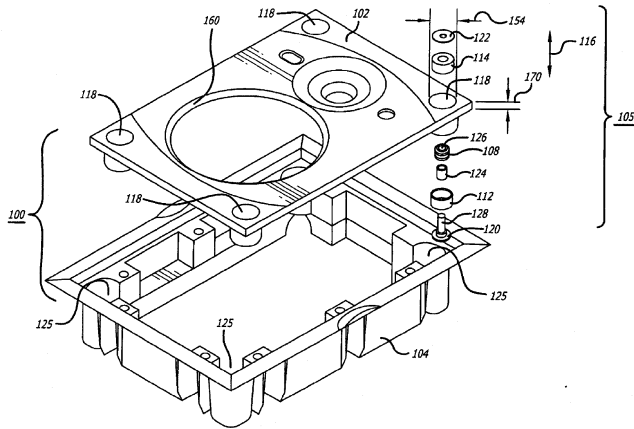
headset, which is the only type covered by the patent claims. A straightforward alternative is disclosed: If the sound level picked up by the microphone rises significantly above the ambient level within a few seconds of a ring signal (indicating that the user is speaking), then the system is energized for two-way communication. Well and good, but without voice recognition, how will a cell phone user initiate an outgoing call? Details, details.—GLA

7,073,624

43.38.Tj LOUSPEAKER BAFFLE ISOLATION SYSTEM

William Andrew Decanio *et al.*, assignors to Harman International Industries, Incorporated
11 July 2006 (Class 181/150); filed 31 July 2003

A means of isolating the baffle 102 and the loudspeakers mounted



therein from back box 104 is disclosed.—NAS

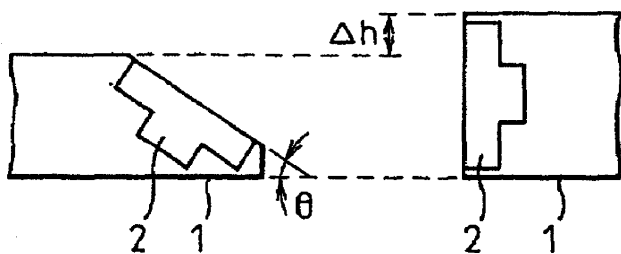
7,072,484

43.38.Vk INFORMATION PROCESSING APPARATUS HAVING SPEAKERS

Takahiro Shin and Masuo Ohnishi, assignors to Fujitsu Limited
4 July 2006 (Class 381/388); filed in Japan 24 October 1996

This is in, especially for "information processing apparatus," colloquially known as a laptop or portable computer. How to be thin and accommodate the transducers needed for sound is partially resolved by angling

(INVENTION) (PRIOR ART)



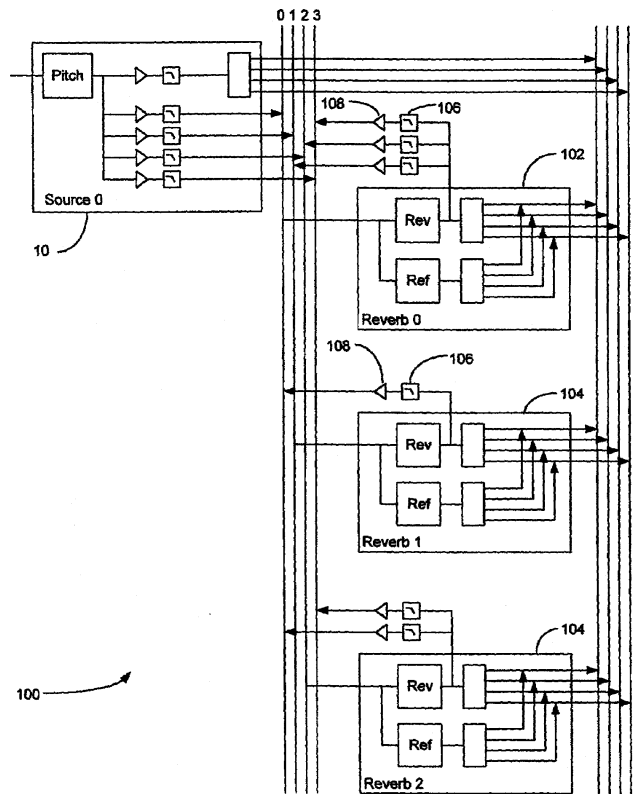
loudspeaker 2 by angle θ with respect to the top surface of the computer 1. A reduction in height equal to Δh can be realized.—NAS

7,099,482

43.38.Vk METHOD AND APPARATUS FOR THE SIMULATION OF COMPLEX AUDIO ENVIRONMENTS

Jean-Marc Jot and Samuel C. Dicker, assignors to Creative Technology Limited
29 August 2006 (Class 381/61); filed 8 March 2002

The patent deals with the generation of virtual acoustical environments for use in video games, electronic music, and the like. The basic building



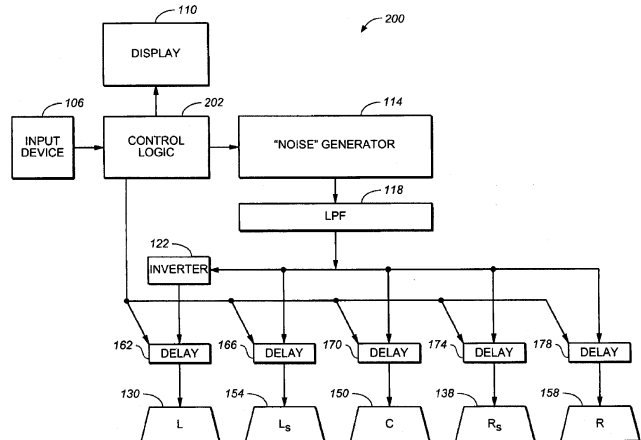
blocks are delay and reverberation generators, which can be efficiently assigned and adjusted to the scene at hand.—JME

7,103,187

43.38.Vk AUDIO CALIBRATION SYSTEM

Darren D. Neuman, assignor to LSI Logic Corporation
5 September 2006 (Class 381/59); filed 30 March 1999

The preferred setup for domestic 5.1 surround sound utilizes five identical loudspeakers equidistant from a single listening location. In a narrow room, however, left and right surround speakers may have to be located nearer to the listening location than the main speakers. To cope with this problem, some installations allow adjustment of level and delay so that

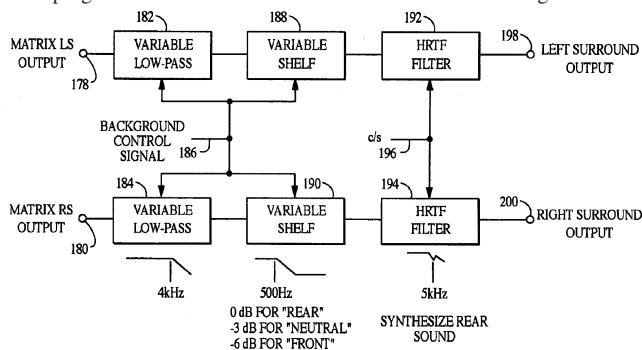


nearly-identical acoustic signals can be received from all five speakers. This patent suggests that adjustment of delay can be simplified and automated by choosing one speaker as a reference and then sequentially matching its output to that of the remaining four speakers. The novel element is that the signal to the speaker under test is inverted and the correct delay is then established by tuning for a null at the listening location.—GLA

43.38.Vk 5-2-5 MATRIX ENCODER AND DECODER SYSTEM

David H. Griesinger, assignor to Harman International Industries, Incorporated
 12 September 2006 (Class 704/228); filed 17 October 2003

The inventor is a well-regarded expert in multichannel matrixing technology and has been awarded a number of patents in the field. This latest patent discloses additional improvements to a universal matrix scheme that is compatible with other encoding systems and can also expand two-channel stereo program material into full surround sound. Performance goals include

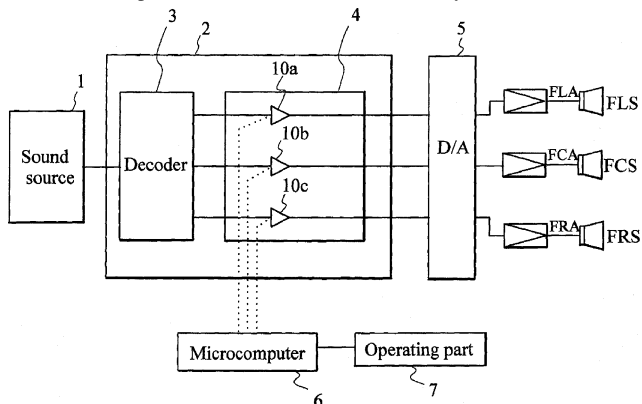


robust placement of front-left and front-right sources under a variety of conditions, and making “the sound of a five channel version closer to that of a seven channel version.” The patent does a good job of explaining a complicated subject and the 46 claims are commendably free of legal jargon.—GLA

43.38.Vk SOUND SYSTEM

Akira Motojima *et al.*, assignors to Fujitsu Ten Limited
 19 September 2006 (Class 381/27); filed in Japan 17 March 2000

The traditional left and right channels in stereo can be adequately adjusted by an off-center listener using two controls: left-right balance and overall level. In surround sound systems there are three frontal channels, and the conventional dual controls cannot properly adjust the individual listening levels. Case in point: When the listener moves, say, to the left, the two

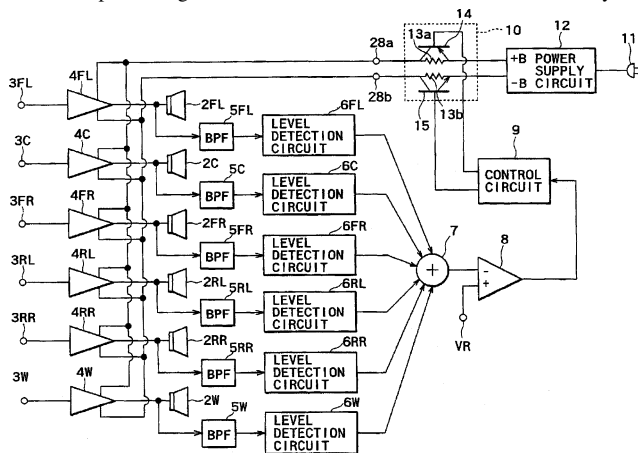


outside channels (left and right) may be appropriately adjusted with a simple balance control. But what about the center channel? Its level will probably be too loud. The patent suggests a solution for this problem, which individually adjusts all levels to produce a uniform loudness contribution from all active sources.—JME

43.38.Vk SOUND SIGNAL REPRODUCING APPARATUS

Hideaki Shiobara, assignor to Sony Corporation
 19 September 2006 (Class 381/96); filed in Japan 7 August 2001

For increased efficiency, many audio power amplifier designs now include intelligent power supplies. The power supply voltage is adjusted dynamically in response to the varying signal level. However, in a multi-channel amplifier, a single power supply normally services all of the channels. This patent argues that the loudness of a surround sound system

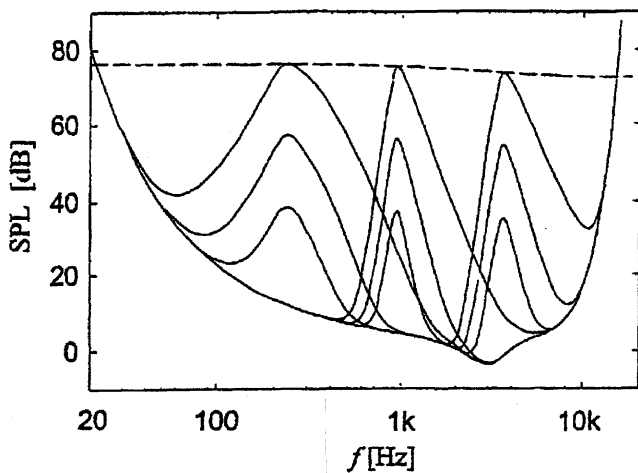


depends upon the total power delivered to the loudspeakers, no matter how it is distributed between individual channels. Therefore, the levels of all five channels can be added and used to derive a single control signal for adjustment of power supply voltage. There is at least one weak point in this reasoning, but the scheme should work well enough in practice.—GLA

43.38.WI NOISE REDUCTION IN A STEREO RECEIVER

Jens Wildhagen, assignor to Sony Deutschland GmbH
 19 September 2006 (Class 381/13); filed in 8 November 2000

In conventional FM broadcasting technology, most reception noise problems arise in the demodulation of the L-R subcarrier signal. Historically, a variety of gating and signal compression techniques have been used to reduce noise in the difference channel. This concise and well-written patent introduces the use of psychoacoustic masking thresholds to this half-century-old art. As the patent states: “The method to denoise a stereo sum



signal and a stereo-difference signal according to the present invention is characterized by a frequency-selective stereo-to-mono blending system based on the masking effect of the human auditory system.”—JME

7,104,055

43.38.Yn PRESSURE VIBRATION GENERATOR

Yoichi Matsubara *et al.*, assignors to Japan Aerospace Exploration Agency
12 September 2006 (Class 60/520); filed in Japan 19 June 2002

The patent describes an acoustic generator/amplifier driven by a heat exchange process alone without any electro-mechanical conversion. The device comprises a network of connected tuned-vibration resonators coupled into a heat exchanger. A pressure wave, originating at one of the resonators, propagates through a heat exchanger in the direction opposite to the heat flow. According to the authors, this causes an amplification of the input pressure wave “because a flow of heat ... is converted into a reverse work flow.” No further explanation is given, but the concept looks intriguing.—DMD

7,102,467

43.40.Cw METHOD FOR ADJUSTING THE FREQUENCY OF A MEMS RESONATOR

Markus Lutz and Aaron Partridge, assignors to Robert Bosch GmbH
5 September 2006 (Class 333/186); filed 28 April 2004

This patent describes several ways that electrical heating can be used to change the resonant frequency of resonant beam structures, for instance by ablation of material from the beam. The abstract promises a lot, but mostly this is just replacing the lasers used by industry with built-in electric heaters. The ideas are not new.—JAH

7,104,595

43.40.Rj AUTOMOTIVE FLOOR PANEL STRUCTURE

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

A floor panel according to this patent is designed so that it radiates little noise into the vehicle interior at a frequency at which there exists a particular external noise—for example, that from tires and wheel-well resonances. The panel is made to have low radiation efficiency by configuring it so that it has a resonant mode with small volume velocity at this frequency due to adjacent areas vibrating in opposite phase. The panel is divided into suitable areas by the addition of stiffening beads, ridges, or lines of damping material.—EEU

7,100,434

43.40.Tm DEVICES AND METHODS FOR SIMULATING TIRE NON-UNIFORMITY FORCES FOR VEHICLE VIBRATION SENSITIVITY MEASUREMENTS AND TUNING

Perry Gu *et al.*, assignors to Ford Global Technologies, LLC
5 September 2006 (Class 73/146); filed 7 October 2004

Unbalanced weights are attached to one or more rotating elements in a gearbox that is fastened to a wheel of an automotive vehicle and driven by the wheel’s rotation. This rotation may be generated by an external device,

such as a dynamometer, or by movement of the vehicle along a road. The magnitudes, rotational radii, and rotational directions of the weights are chosen to simulate the effects of tire nonuniformities.—EEU

7,102,474

43.40.Tm ADAPTABLE VIBRATION ABSORBER EMPLOYING A MAGNETORHEOLOGICAL ELASTOMER WITH VARIABLE GAP LENGTH AND METHODS AND SYSTEMS THEREFOR

Anne-Marie Albanese Lerner and Kenneth A. Cunefare, assignors to Georgia Tech Research Corporation
5 September 2006 (Class 335/229); filed 26 January 2004

Unlike a conventional vibration absorber or “tuned damper,” which consists of a mass supported on a spring of constant stiffness, an absorber according to this patent employs a mass supported on springs consisting of a magnetorheological elastomer. The effective stiffness and damping of the latter are affected by the flux from a magnet placed near the absorber’s mass, and thus, change as the mass vibration amplitude changes. The stiffness and damping of the elastomer elements also may be changed actively by controlling the magnetic field.—EEU

7,108,111

43.40.Tm SEMI-ACTIVE ISOLATOR

Gerald J. Spyche, Jr. and Kenichi Tomita, assignors to Enidine Incorporated
19 September 2006 (Class 188/378); filed 25 October 2004

This shock and vibration isolation system includes a double-acting mechanical spring assembly that acts in mechanical parallel with a fluid spring assembly. The mechanical spring assembly provides one spring rate for displacements, up to a certain displacement, and a second spring rate for larger displacements. The fluid spring is connected to an accumulator via a valve that is under control of a microprocessor, and is activated on the basis of a signal from a sensor that measures the pressure in one of the chambers of the fluid spring. The control system acts to release stored energy to protect the payload from excessive external accelerations.—EEU

7,107,127

43.40.Vn COMPUTATIONALLY EFFICIENT MEANS FOR OPTIMAL CONTROL WITH CONTROL CONSTRAINTS

Robert Karl Goodman, assignor to Sikorsky Aircraft Corporation
12 September 2006 (Class 700/280); filed 26 February 2002

Although this patent addresses the control of noise and vibration in helicopter interiors, its principle is likely to have greater applicability. The control arrangement described here deals with situations where the control requirement would exceed the capabilities of parts of the system—e.g., the force capability of an actuator. The control algorithm generates a first command that requires a component (such as an actuator) to produce a certain output. If this output exceeds a maximum allowable value, this output is

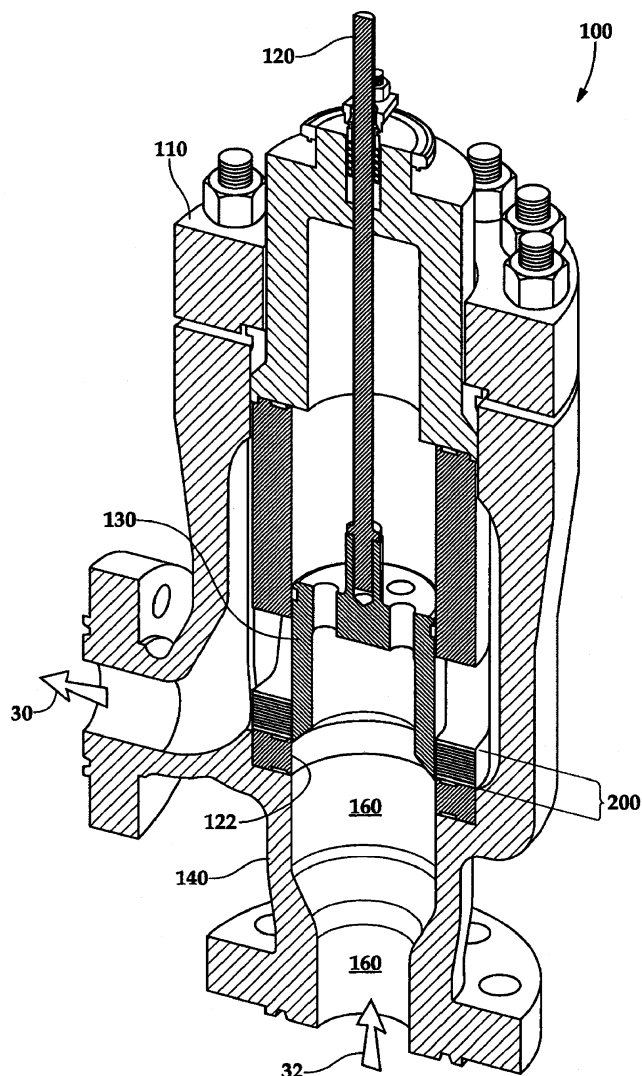
scaled down and the residual vibration with that scaled-down output is calculated. Then a second control signal is generated, ignoring the aforementioned actuator, and the effect of the resulting command on another actuator is evaluated in analogy to the first one. This process is repeated for other actuators before the command signal is sent to all of the actuators.—EEU

7,109,679

43.40.Vn DAMPING FOR ELECTROMECHANICAL ACTUATORS

Ralph D. Edson and M. Robert Mock, assignors to HR Textron, Incorporated
19 September 2006 (Class 318/611); filed 9 March 2004

Damping in systems that use electric motors for moving and positioning a load may be accomplished by use of a feedback arrangement where the actuator functions in a normal operating mode of the motor. A feedback signal may be obtained from a force, torque, or acceleration sensor or from measurement of the motor current. Damping may also be achieved passively, for example by short circuiting the motor purposely through a number of resistors. A motor with a smaller than usual stator-to-rotor diameter ratio has relatively low inertia, which is preferable for control purposes.—EEU



7,104,133

43.40.Yq TORSIONAL VIBRATION MEASURING INSTRUMENT

Hidehiko Kuroda *et al.*, assignors to Kabushiki Kaisha Toshiba
12 September 2006 (Class 73/650); filed in Japan 28 March 2002

Reflecting elements are attached at intervals along the surface of an elongated rotating element whose torsional vibrations are to be measured. Repetitive light pulses, as from a laser, are directed at these elements and the reflections are detected and processed. The processing is devised so as to permit the extraction of torsional vibration data in the presence of axial and lateral vibrations.—EEU

waves in the main flow path, so as to attenuate the noise in the main flow path. The H-Q arrangement can be fabricated into a cartridge 200, now patented, for use in valves and other such devices.—NAS

7,070,513

43.50.Gf GOLF CLUB

Hitoshi Takeda *et al.*, assignors to K.K. Endo Siesakusho
4 July 2006 (Class 473/329); filed in Japan 13 November 2003

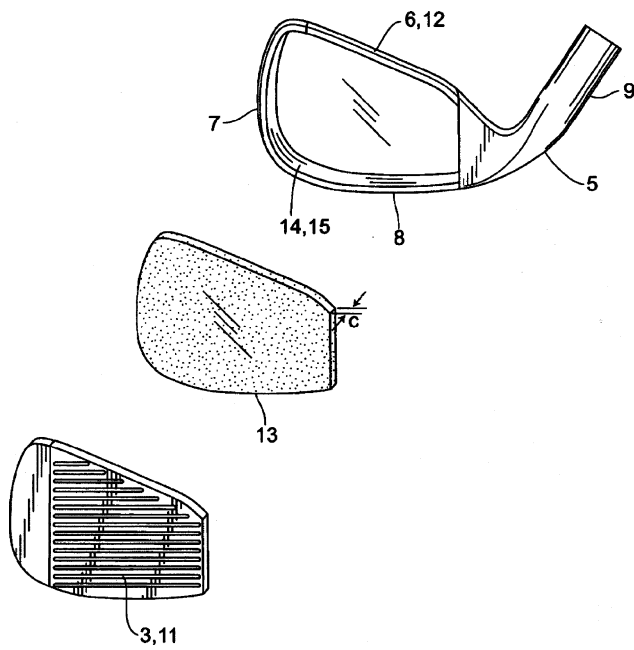
If one subscribes to the proposition that high-tech materials and design in golf club heads cause too much noise and that the exposure to same can be deleterious to a golfer's hearing, as well as being an unpleasant noise that could detract from the pleasure of the game (the reviewer can think of some more pressing unpleasanties relating to player skill, performance, foursome makeup, etc.), then this is for you. In the various embodiments, basically, a gel pad 13 is placed between the face member 11 and the head body 12. The gel pad combines both elasticity and hardness (preferably with a specific gravity=0.9–1.1, a hardness penetration ratio of ASKER C=30–55, an

7,069,950

43.50.Gf NOISE ABATEMENT MODULE USING HERSCHEL-QUINCKE TUBES

Christopher Bittner, assignor to Dresser, Incorporated
4 July 2006 (Class 137/625.37); filed 30 March 2005

Take a fluid flow, divert part of it into a secondary path, a prior art device called a Herschel-Quincke (H-Q) tube, of length, L, after which the secondary path recombines with the primary path. Make sure that L is such that any acoustic waves in the H-Q tube are out of phase with the acoustic



elongation percentage of 400%–750%, $E=30\text{--}150\text{ kPa}$, and a compression set 5% or less). The patent admits that the face member can be a ferrous alloy of Mn, Ni, Cr, C, Si, Mo, and V with a pinch of P and S, between 1.0–2.0 mm in thickness.—NAS

7,100,742

43.50.Gf METHOD FOR CONTROLLING THE DELIVERY OF LUBRICANT

Christian Gunacker and Richard Gunacker, assignors to Hy-Power Flexomatic Hydraulik Handelsoges M.b.H.
5 September 2006 (Class 184/3.1); filed in Austria 11 August 2000

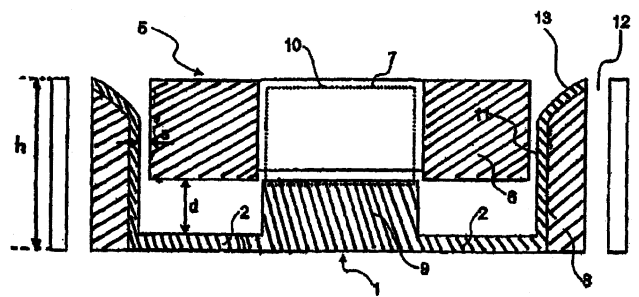
Lubricant is applied to train tracks at curves in order to reduce the squealing noise produced when rail vehicles traverse these curves. However, it is desirable to apply this lubricant only where it is needed, so as not to have it reduce braking effectiveness. The present patent describes a vehicle-mounted system that uses one or more microphones to sense the noise near the rail, filters the signal into at least two frequency bands, compares the levels in these bands, and dispenses the lubricant based on this comparison. Use of multiple bands enables the system to perform its function in the presence of variable rolling noise and extraneous noise.—EEU

7,101,149

43.50.Gf AXIAL FAN FOR COMPUTER

Magnus Huber, assignor to arctic-cooling Switzerland AG
5 September 2006 (Class 415/176); filed in Germany 22 May 2003

Axial flow fan 5–10 is supported by struts 12 located at a distance d from blades 6. Proximity of blades 6 to struts 12 normally causes blade-



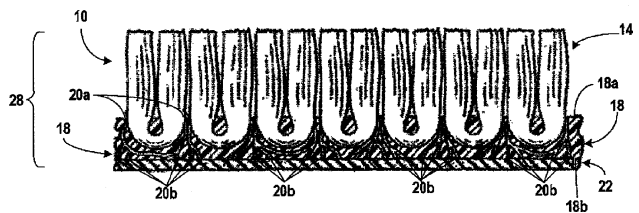
passage noise emission. The author claims a widened separation distance d to reduce blade-passage noise production.—AJC

7,105,069

43.50.Gf SOUND ABSORBING/SOUND BLOCKING AUTOMOTIVE TRIM PRODUCTS

Timothy J. Allison *et al.*, assignors to Collins & Aikman Products Company
12 September 2006 (Class 156/72); filed 21 May 2004

Sound absorbing material 28 for a vehicle passenger-compartment floor is formed from carpet 10 with tufts 14, backing 18 and thermoformable material (e.g. polyethylene) placed at 18a before heat processing. To produce the porosity required for sound absorption, presenting 150–10,000



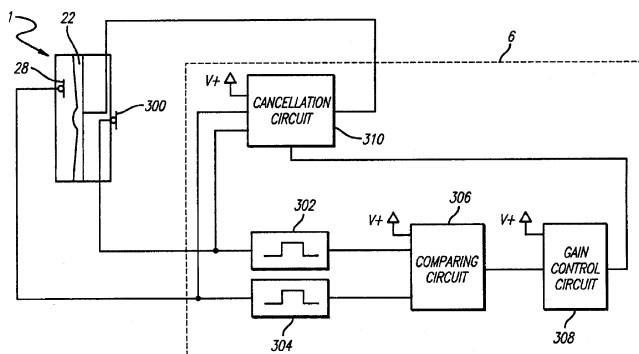
Rayls, the carpet is heated, melting added PE layer 18a and forming it into channels 20a. To stiffen this floor material, another layer of plastic 22 can be applied, which may also be rendered porous for greater flow resistance, or for added mass to serve as a sound barrier. Methods are also described for forming this material to fit floor panels.—AJC

7,103,188

43.50.Ki VARIABLE GAIN ACTIVE NOISE CANCELLING SYSTEM WITH IMPROVED RESIDUAL NOISE SENSING

Owen Jones, Alreford Colchester, United Kingdom
5 September 2006 (Class 381/71.9); filed 8 March 1999

This headphone system reduces active noise cancelling (ANC) gain when the environment is quiet and allows external signals for communications and entertainment to be mixed in with ANC energy. Three novel modes are claimed. (1) Error microphone 28 is offset a distance from the speaker center by a distance equal to the distance from the speaker center to the eardrum to effect feed-forward action. (2) To detect incomplete circumaural cushion sound seal at low frequencies, a 10 Hz infrasound signal is introduced and measured by microphone 28. This 10 Hz sound pressure level is



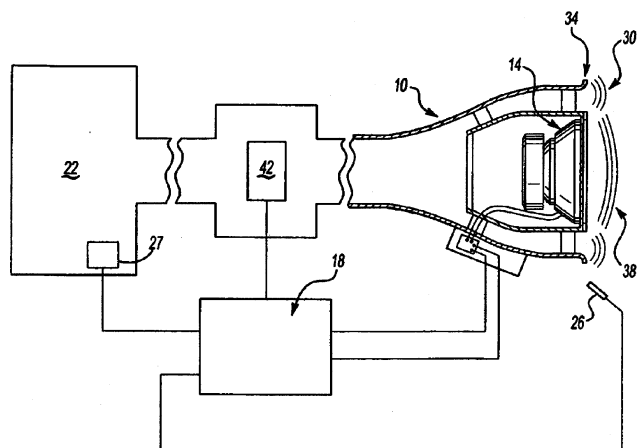
high if the seal is good and low when the seal is poor or the headset is not being worn. The absence of such a seal can destabilize the ANC system, so the ANC gain of **310** is reduced. (3) When the environment is quiet, ANC is not needed. Conditions 2 and 3 are sensed by comparison circuit **306**, that through **398** reduces the gain of the cancellation circuit **310**, thus, avoiding many conditions of instability. Another embodiment has pressure sensitive switches embedded in the circumaural seal that report when the headset is not being worn, also signalling that the gain must be reduced to avoid instability.—AJC

7,106,866

43.50.Ki ACTIVE NOISE CANCELLATION STABILITY SOLUTION

John F. Astorino *et al.*, assignors to Siemens VDo Automotive, Incorporated
12 September 2006 (Class 381/71.4); filed 6 April 2001

Instabilities in active noise cancellation (ANC) of induction noise emission **30** from engine air intake **10** can cause an audible squeal **38** under low ambient noise and engine idle conditions. The author claims control circuit **18** that senses low ambient noise via microphone **26**, low engine RPM via engine tachometer **27**, and low engine power via throttle position sensor (TPS) **42**, to determine the ratio of expected induction noise to



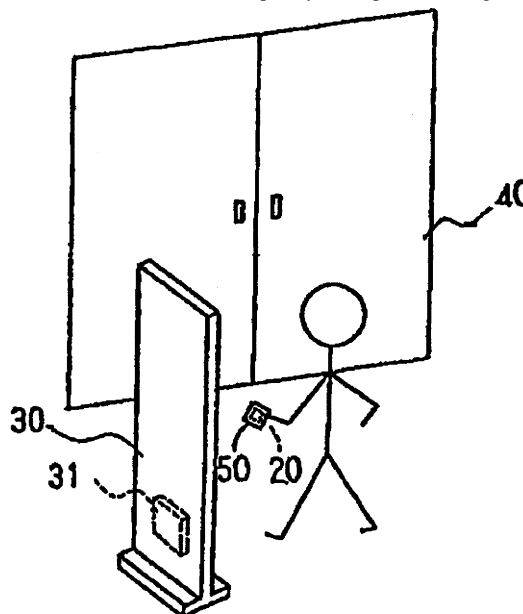
environmental noise. If microphone **26** reports a high-pitch sound while the throttle position is low, **18** shuts down the ANC control. When the engine rpm and throttle position indicate higher power, ANC is resumed. Other logic in **18** counts instances of high-frequency noise (errors). This may first reset the ANC logic, or after repeated or chronic such instances, shut down this induction noise ANC system and alert the driver that maintenance is required.—AJC

7,088,231

43.50.Rq ANTITHEFT SYSTEM

Takuya Suzuka and Hiroshi Yoshida, assignors to Sensormatic Electronics Corporation
8 August 2006 (Class 340/531); filed in Japan 12 July 2001

A burglar alarm system is described capable of accurately discriminating nearby external noises such as background music, ambient noises, etc., from reverberations of a legitimate alarm echoing from the interior of the store, etc. An alarm unit attached to an object to be protected from theft would emit an alarm of a certain frequency in response to illegal conduct.



An alarm sensor would emit an alarm signal upon sensing an alarm from the alarm unit. The alarm sensor uses two methods of evaluating signals from the alarm unit and external noises: It measures the randomness contained in the input signal and it evaluates a reflected sound caused by reverberations of the alarm.—DRR

7,110,536

43.55.Ti ACOUSTIC SEAL SYSTEM

Patrick A. Hampton *et al.*, assignors to Motorola, Incorporated
19 September 2006 (Class 379/433.02); filed 27 June 2003

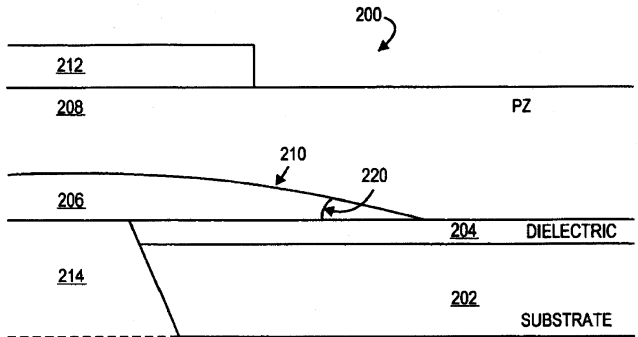
This acoustic seal system incorporates a flexible sealing element with a ridge that engages both a cover plate and another plate on the inner housing of a cellular phone or the like, in order to make an effective acoustic seal.—CJR

7,109,826

43.58.Kr TAPERED ELECTRODE IN AN ACOUSTIC RESONATOR

Eyal Ginsburg *et al.*, assignors to Intel Corporation
19 September 2006 (Class 333/187); filed 14 June 2005

This patent discloses the construction details of the electrodes in film bulk acoustic wave resonators (FBARs) and surface acoustic wave resonators (SAWs). The concept disclosed is that tapering the thickness of the



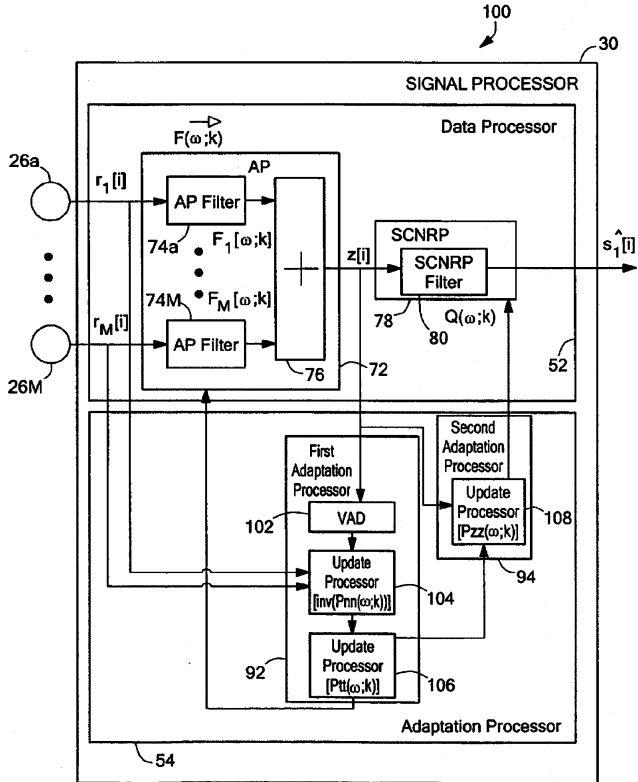
lower electrode in said structures at a "five to thirty degree angle" advantageously reduces defects in the piezoelectric layer and consequently reduces insertion loss for the resulting resonator filter shown in the figure.—JAH

7,099,822

43.60.Dh SYSTEM AND METHOD FOR NOISE REDUCTION HAVING FIRST AND SECOND ADAPTIVE FILTERS RESPONSIVE TO A STORED VECTOR

Kambiz C. Zangi, assignor to Liberato Technologies, Incorporated 29 August 2006 (Class 704/226); filed 12 August 2004

The patent describes a procedure for rapidly adapting a hands-free communication system to a specific automobile design: "A system for microphone noise reduction includes first and second filter portions in response to one of a plurality of stored vectors. Each stored vector is representative of acoustic transfer functions in accordance with a model of a vehicle and a



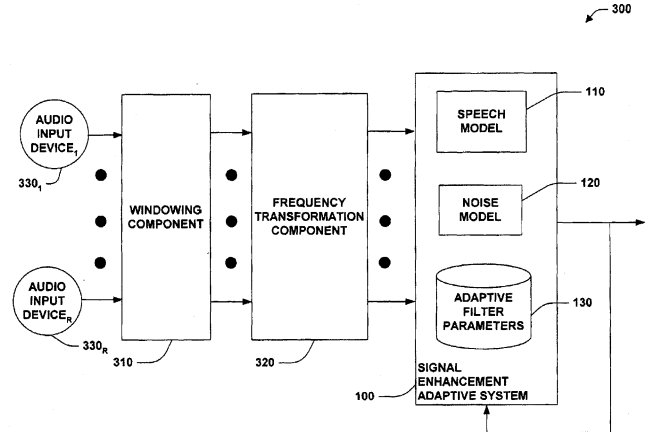
respective position within the vehicle. A method for processing microphone signals includes selecting a vehicle model, selecting positions within the vehicle model, measuring acoustic response vectors at the positions, storing the response vectors, and adapting first and second filter portions in accordance with the selected response vector."—JME

7,103,541

43.60.Dh MICROPHONE ARRAY SIGNAL ENHANCEMENT USING MIXTURE MODELS

Hagai Attias and Li Deng, assignors to Microsoft Corporation 5 September 2006 (Class 704/226); filed 27 June 2002

This math-intensive patent suggests the use of a number of adaptive parameters for enhancing speech in noisy environments. Quoting from the Abstract, "The invention includes a signal enhancement adaptive system having a speech model, a noise model and a plurality of adaptive filter



parameters ... The signal enhancement adaptive system incorporates information about the statistical structure of speech signals. The signal enhancement adaptive system can be imbedded in an overall enhancement system which also includes components of signal windowing and frequency transformation."—JME

6,996,532

43.60.Ek METHOD AND APPARATUS FOR ACCESSING A CONTENT SITE WITH A SOUND SEQUENCE

Andrew Thomas, assignor to Hewlett-Packard Development Company, L.P. 7 February 2006 (Class 704/270); filed in United Kingdom 7 December 2000

Intended for providing more convenient Internet access from a cell phone, patented here is a system for supposedly improving the representation of a web site URL as a sequence of tone codes. Prior methods using DTMF (Touch Tone™) codes are discussed (and dismissed as sounding unnatural). The method introduced here is a variation on a simple single-tone code in which the tone frequency is converted to a character code bit pattern. The cited "improvement" consists of reshuffling the coding table so that common character sequences are heard as melodic tone sequences, all to be more pleasing to the ear and thus, it is assumed, more memorable.—DLR

6,999,593

43.60.Jn SYSTEM AND PROCESS FOR ROBUST SOUND SOURCE LOCALIZATION

Yong Rui and Dinei A. Florencio, assignors to Microsoft Corporation 14 February 2006 (Class 381/92); filed 28 May 2003

This patent deals with methods of locating a signal source using multiple microphones. A prior technique which involves pairwise comparisons of the microphone cross correlations is compared to two related methods in which all microphone signals are considered simultaneously. Both new methods involve computing the Fourier transform of all mic signals, allow-

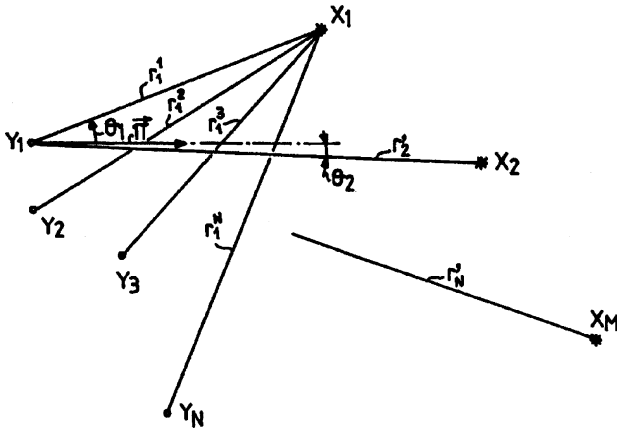
ing spectral weighting functions to be more easily included in the computations. Test results show good performance in noisy conditions (due to the use of weighting functions) in a room on the order of 20–25 ft across, using a 44.1-kHz sample rate and 1024-point transforms.—DLR

6,999,881

43.60.Jn METHOD AND APPARATUS FOR DETECTING AND LOCATING NOISE SOURCES WHETHER CORRELATED OR NOT

Alfred Permuy and Joël Millet, assignors to Metravib R.D.S.
14 February 2006 (Class 702/17); filed 17 December 2003

This patent presents a very general method for determining the locations of a small number of noise sources X_m $m=1...M$ using a small number of pickups Y_n $n=1...N$. Indicating the intended generality of the patent, the



pickups are referred to as sensors, rather than microphones. Both sound and vibration sources are discussed. Fourier transforms of the source signals may include expressions for individual sensor gains and propagation factors. Matrix equations are presented for solving the resulting system.—DLR

7,107,849

43.60.Rw VIBRATION SOURCE PROBE SYSTEM

Takeshi Sugiyama *et al.*, assignors to Chubu Electric Power Company, Limited
19 September 2006 (Class 73/592); filed in Japan 22 August 2002

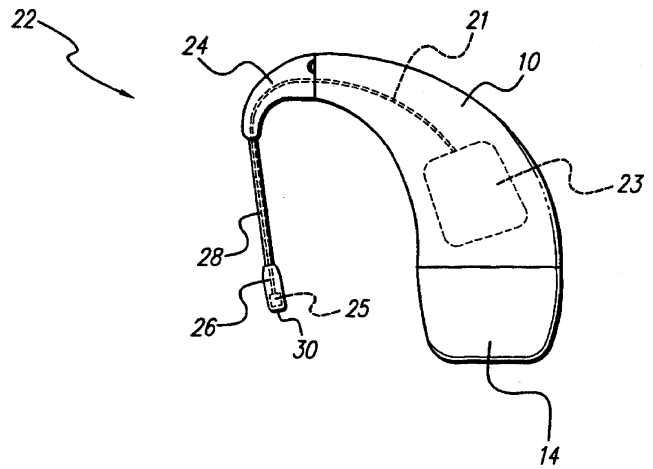
The position of a vibration source, such as a vehicle on a road or a machine on a factory floor, is estimated from the phase difference or arrival time difference of surface waves detected by an array of sensors. An image near the estimated position of the vibration source is picked up by a camera and displayed on the screen of a personal computer.—EEU

7,106,873

43.64.Me IN THE EAR AUXILIARY MICROPHONE FOR BEHIND THE EAR HEARING PROSTHETIC

William Vanbrooks Harrison *et al.*, assignors to Advanced Bionics Corporation
12 September 2006 (Class 381/330); filed 14 April 2004

Hearing aids and implants commonly incorporate a telecoil to pick up via magnetic coupling the inductive signal emanating from telephones. To improve the performance of a behind-the-ear (BTE) cochlear implant system



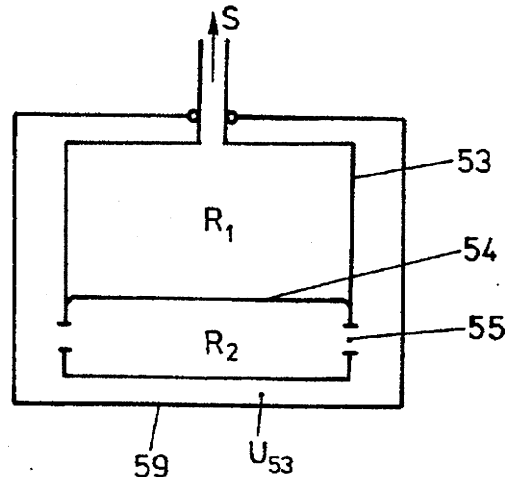
when the wearer is using a telephone that has a low magnetic output, an in-the-ear (ITE) microphone is connected to a removable earhook of the BTE component to facilitate an acoustic pickup of the telephone signal.—DAP

7,099,484

43.66.Ts BEHIND-THE-EAR HEARING AID

Andi Vonlanthen, assignor to Phonak AG
29 August 2006 (Class 381/322); filed 5 April 2002

A normally vacant space around the hearing aid receiver is used to increase the low-frequency output level of the hearing aid. The hearing aid receiver casing is vented in back of the diaphragm and is spring-mounted



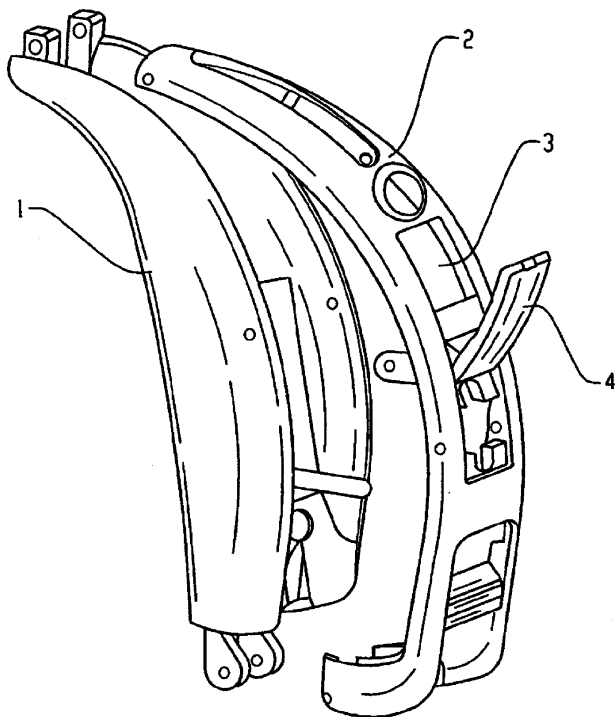
within a removable outer capsule, which also functions as a magnetic shield. The intermediate volume formed between the receiver and the outer capsule couples with the volume behind the receiver diaphragm to enhance the low-frequency acoustic output.—DAP

7,099,485

43.66.Ts HOUSING FOR HEARING DEVICES OR HEARING AIDS

Erich Dittli, assignor to Phonak AG
29 August 2006 (Class 381/322); filed 10 March 2004

A removable one-piece cap with pins protruding is described to cover an opening that provides accessibility to a programming socket. The cap is



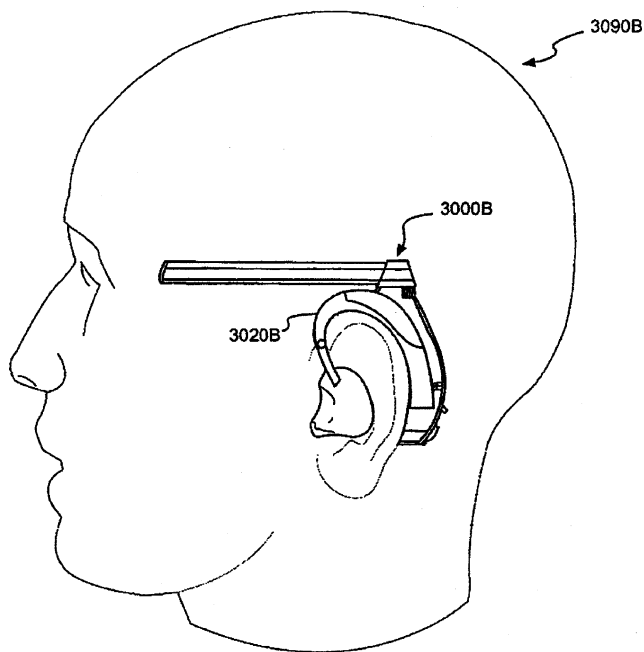
hinged via the protruding pins engaging with grooves in the housing with dead stops that help prevent the cap from falling out or being pulled out.—DAP

7,099,486

43.66.Ts MULTI-COIL COUPLING SYSTEM FOR HEARING AID APPLICATIONS

Stephen D. Julstrom *et al.*, assignors to Etymotic Research, Incorporated
29 August 2006 (Class 381/331); filed 31 January 2003

Two or more inductors are oriented to efficiently couple an audio signal from an ear-level assistive device to a behind-the-ear or an in-the-ear



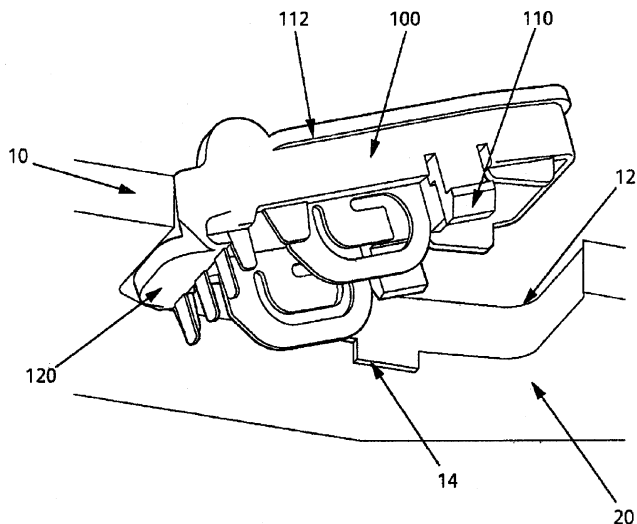
hearing aid. The inductor selected via a switch depends on which style of hearing aid is used by the wearer. The ear-level assistive device may contain a directional array of microphones.—DAP

7,106,872

43.66.Ts LOCKING MECHANISM FOR ELECTRONICS MODULE FOR HEARING INSTRUMENTS

Martin W. Masters, assignor to Siemens Hearing Instruments, Incorporated
12 September 2006 (Class 381/322); filed 27 June 2003

A battery door on an electronics module that inserts into an opening in a hearing aid housing is frequently also used as a handle to remove the hearing aid from the ear. If significant force is applied to pull the hearing aid out of the ear, latches on the module that mate with protrusions on the inside



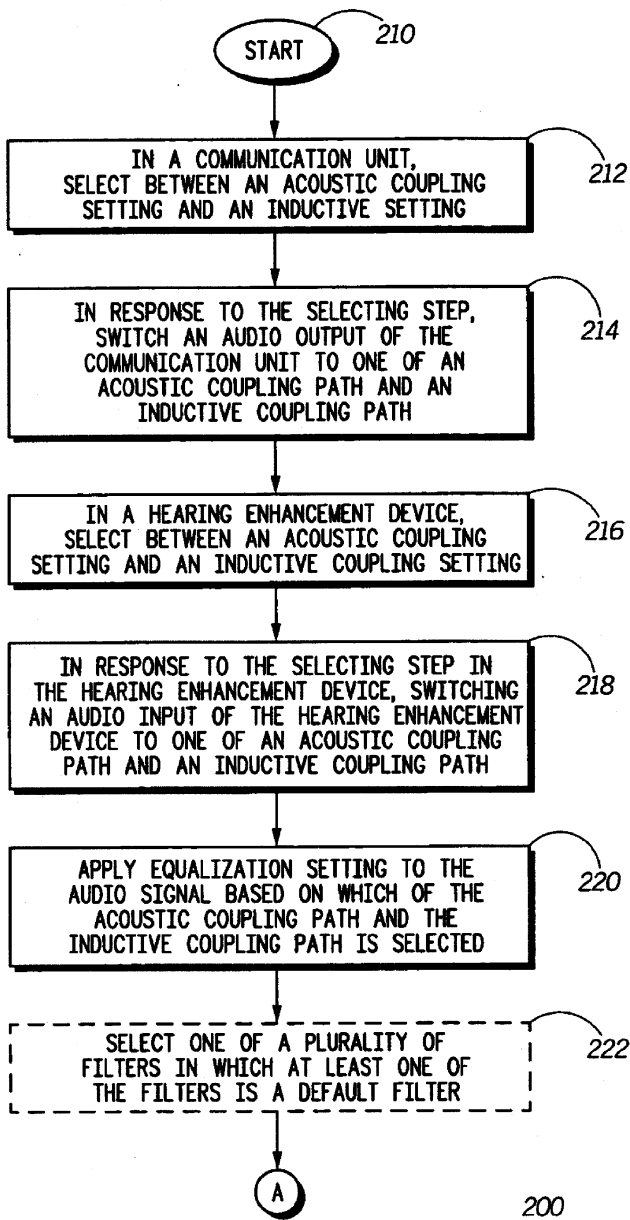
surface of the hearing aid housing may not be sufficient to prevent the module from being pulled out of the hearing aid housing. An outward-protruding tab is added near the battery door hinge on the underside of the module that slips under the inside surface of the housing to further prevent the module from being pulled out of the housing.—DAP

7,106,874

43.66.Ts METHOD AND SYSTEM FOR SELECTIVE COUPLING OF A COMMUNICATION UNIT TO A HEARING ENHANCEMENT DEVICE

Alfred B. Wiczorek and Brian L. Adair, assignors to Motorola, Incorporated
12 September 2006 (Class 381/331); filed 13 July 2004

A selection between acoustic and inductive audio-signal coupling to a hearing enhancement device is made in a communication device such as a



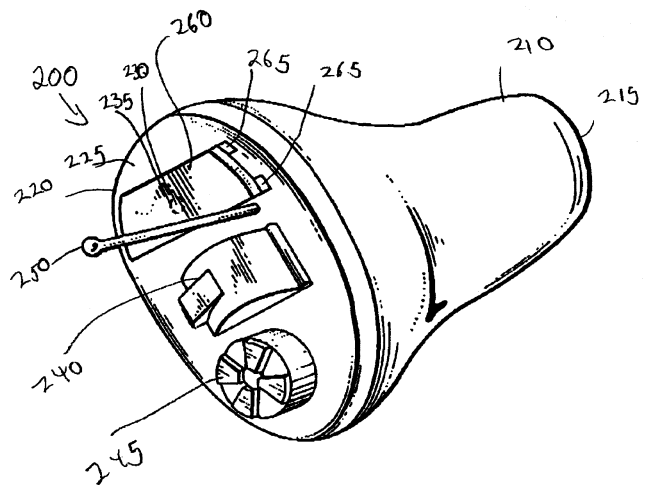
cellular telephone. Equalization is applied to the audio signal depending on which audio-signal coupling method is chosen.—DAP

7,106,875

43.66.Ts DUAL BOUNDARY PRESSURE ZONE THREE DIMENSIONAL MICROPHONE AND HEARING AID

James T. King, Silver Cliff, Colorado
12 September 2006 (Class 381/359); filed 9 September 2003

A pressure zone microphone (PZM) is created for a custom hearing aid by adding a first boundary button immediately over the microphone inlet. Another boundary is created having the microphone flush to the faceplate of



the custom hearing aid. The first boundary button is parallel to the faceplate, resulting in a pressure zone between the two boundaries. The patent claims improved frequency response, although no data are shown.—DAP

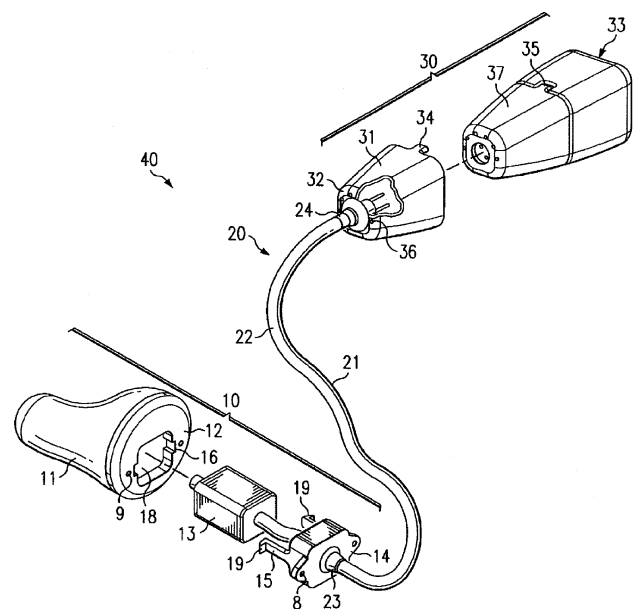
7,110,562

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

Jim Feeley and Mike Feeley, assignors to Hear-Wear Technologies, LLC

19 September 2006 (Class 381/322); filed 10 September 2002

A behind-the-ear (BTE) component containing processing circuitry and a completely-in-the-canal (CIC) component consisting of a speaker and



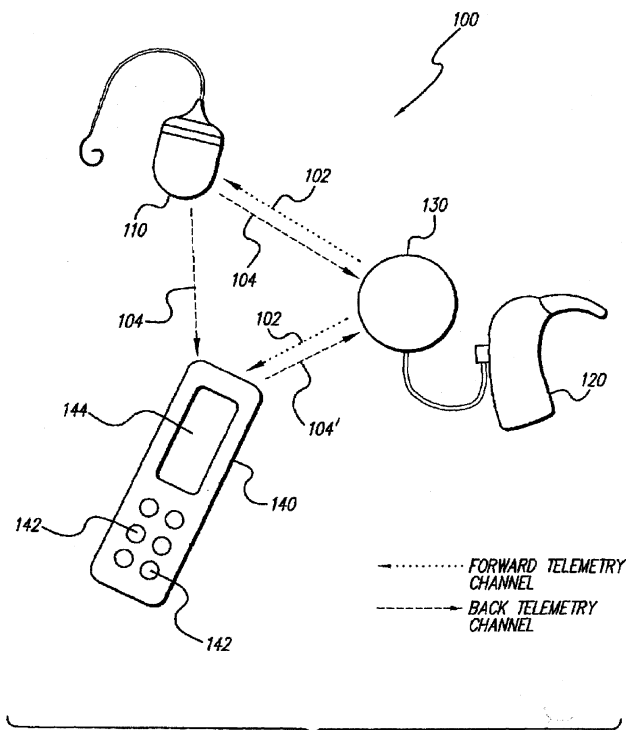
either a custom or standard-fit earmold make up a hybrid BTE/CIC auditory device. The speaker module may be detached via disposable connectors from the earmold.—DAP

7,110,822

43.66.Ts REMOTE STATUS AND CONTROL DEVICE FOR A COCHLEAR IMPLANT SYSTEM

Logan P. Palmer, assignor to Advanced Bionics Corporation
19 September 2006 (Class 607/57); filed 18 April 2003

A hand-held device is placed near the headpiece of a cochlear implant system to monitor and display status information detected from forward and back telemetry signals between the external headworn speech processor and



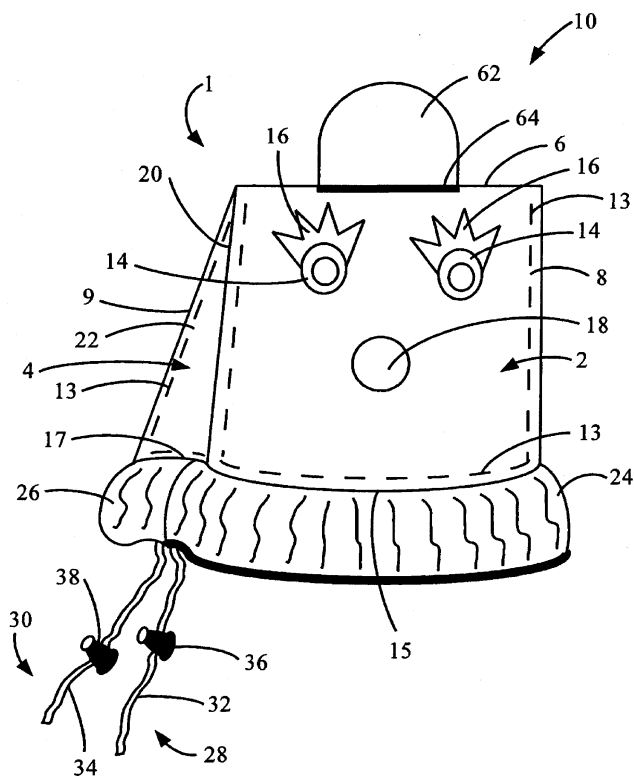
the implanted stimulator. The remote device may also send user-generated control information to the implanted device by modifying the signals for the telemetry path.—DAP

6,994,554

43.70.Dn PHONOLOGY AND ARTICULATION TOOL

Marie A. Daigle, Somers, Connecticut
7 February 2006 (Class 434/185); filed 9 February 2004

This pliable figurine is intended for use by speech therapists or physical speech trainers to demonstrate proper articulator shapes. Not only can the tongue and lips be positioned to demonstrate desired articulations, but a cartoon-like model of the palate shows how the tongue can be positioned against the palate or teeth or shaped in various ways.—DLR

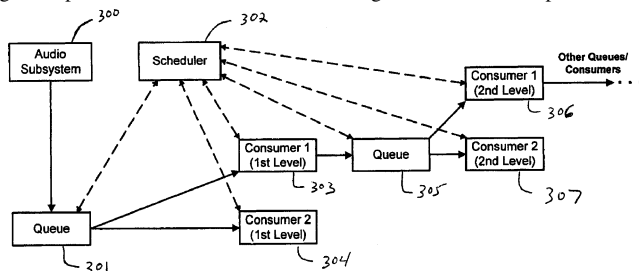


7,100,000

43.72.Gy SYSTEM AND METHODS FOR PROCESSING AUDIO USING MULTIPLE SPEECH TECHNOLOGIES

Abraham P. Ittycheriah *et al.*, assignors to International Business Machines Corporation
29 August 2006 (Class 711/147); filed 17 February 2000

Speech data associated with a single utterance are shared among multiple destinations including feature extraction engines, speech decoding engines, speaker identification/verification engines, and data-compression and



-decompression engines. A scheduler determines priorities within several queues based on registered data requirements and priority requests.—DAP

7,110,941

43.72.Gy SYSTEM AND METHOD FOR EMBEDDED AUDIO CODING WITH IMPLICIT AUDITORY MASKING

Jin Li, assignor to Microsoft Corporation
19 September 2006 (Class 704/200.1); filed 28 March 2002

To eliminate overhead in transmitting an auditory mask, masking thresholds are not sent to the decoder but are instead automatically derived from previously-coded coefficients. The masking thresholds are used to con-

trol the order that the coefficients are encoded so that coefficients for larger audio components that have a greater impact on perceived audio quality are coded first.—DAP

6,996,529

43.72.Ja SPEECH SYNTHESIS WITH PROSODIC PHRASE BOUNDARY INFORMATION

Stephen Minnis, assignor to British Telecommunications public limited company
7 February 2006 (Class 704/258); filed in United Kingdom 15 March 1999

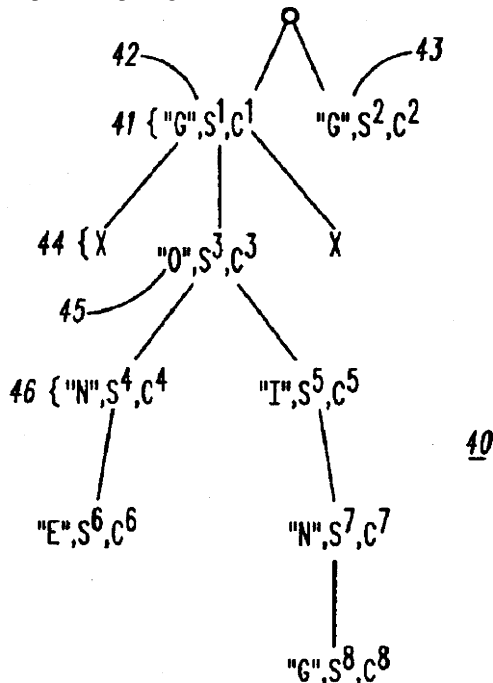
This speech synthesizer design involves a serious effort to provide natural-sounding prosodic structure for the output. The technique is, in a way, analogous to the segment-concatenation methods used by many synthesizers to generate the phonetic sequences. A database of natural utterances containing marked phrase-boundary information is stored in the synthesizer memory. A syntactic analysis of the input text is based on part-of-speech information from a lexicon. The syntactic database is then searched for the nearest matching syntactic structure, which provides a word- and phrase-boundary pattern to be used by the prosody generator.—DLR

6,999,918

43.72.Ja METHOD AND APPARATUS TO FACILITATE CORRELATING SYMBOLS TO SOUNDS

Changxue Ma and Mark Randolph, assignors to Motorola, Incorporated
14 February 2006 (Class 704/10); filed 20 September 2002

The patented object is a hierarchical structure, a tree, in which each terminal node identifies a unique sound/symbol pair. The tree is to be organized in such a way that particular nonterminal nodes could be labeled with symbol sequences so that the daughter branching structure would correspond to the phonetic spelling of a word. There is no mention of the fact that



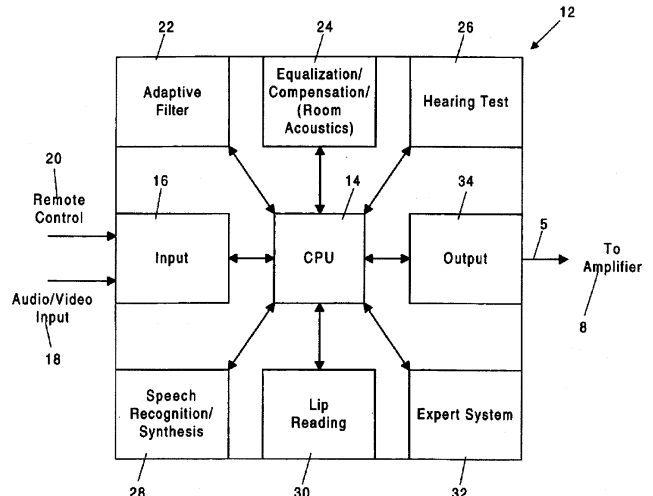
this would require a massive duplication of terminal nodes. The naive presumption is that such a structure might be useful for synthesizing speech and would do so, in a way, as would involve much less memory and processor time than the usual synthesis methods. The gargantuan shortcomings of such a scheme are beyond discussion here.—DLR

7,110,951

43.72.Ja SYSTEM AND METHOD FOR ENHANCING SPEECH INTELLIGIBILITY FOR THE HEARING IMPAIRED

Dorothy Lemelson, legal representative, Incline Village, Nevada et al.
19 September 2006 (Class 704/270); filed 3 March 2000

This system uses a combination of audio signal modification technologies integrated with hearing capability profiles, modern computer vision,



speech recognition, and expert systems for use by a hearing-impaired individual to improve speech intelligibility.—DRR

6,999,927

43.72.Ne SPEECH RECOGNITION PROGRAMMING INFORMATION RETRIEVED FROM A REMOTE SOURCE TO A SPEECH RECOGNITION SYSTEM FOR PERFORMING A SPEECH RECOGNITION METHOD

Todd E. Mozer and Forrest S. Mozer, assignors to Sensory, Incorporated
14 February 2006 (Class 704/244); filed 15 October 2003

This speech recognition system represents an attempt to achieve big-bucks speaker-independent performance with limited memory and processing power. A hierarchical vocabulary structure would allow word-by-word expansion of the context. Applicable in certain device-control situations, the idea is that at each point in the overall syntactic structure, a word or phrase from a limited vocabulary would provide the key to switch to the next limited-vocabulary domain. In this way, a series of small-vocabulary operations might perform much as a single large-vocabulary operation.—DLR

6,999,933

43.72.Ne EDITING DURING SYNCHRONOUS PLAYBACK

Dieter Hoi, assignor to Koninklijke Philips Electronics, N.V.
14 February 2006 (Class 704/278); filed in 29 March 2001

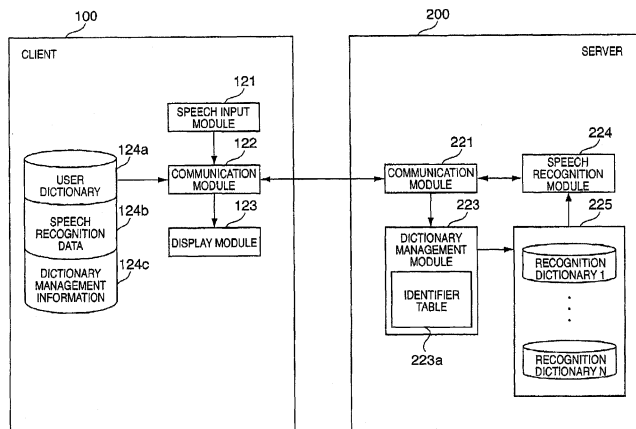
This dictation speech recognizer provides a continuous link between the recorded dictation audio stream and the recognized word sequence. At any time, a cursor may be placed at a specific word in the recognized text, causing an audio cursor to be located at the corresponding point in the audio stream. This is said to allow more rapid correction of the text, although the recommended process to be followed is rather obscured by a rambling and repetitious form of patentese.—DLR

7,099,824

43.72.Ne SPEECH RECOGNITION SYSTEM, SPEECH RECOGNITION SERVER, SPEECH RECOGNITION CLIENT, THEIR CONTROL METHOD, AND COMPUTER READABLE MEMORY

Akihiro Kushida and Tetsuo Kosaka, assignors to Canon Kabushiki Kaisha
29 August 2006 (Class 704/231); filed in Japan 27 November 2000

A client-server speech recognition system employs a stored user dictionary formed by registering target recognition words designated by the



user. The appropriate recognition dictionary in the server is selected via dictionary management information, such as the input form, that is transmitted from the client.—DAP

7,110,947

43.72.Ne FRAME ERASURE CONCEALMENT TECHNIQUE FOR A BITSTREAM-BASED FEATURE EXTRACTOR

Richard Vandervoort Cox and Hong Kook Kim, assignors to AT&T Corporation
19 September 2006 (Class 704/236); filed 5 December 2000

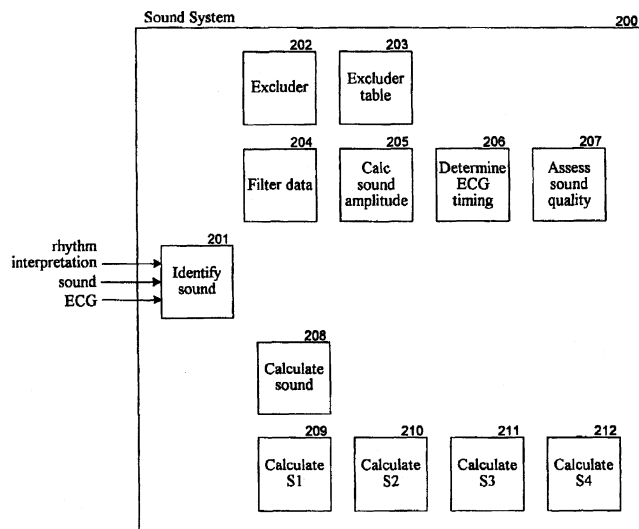
When an error is declared in a frame transmitted in a communication system, one of the frames is deleted from the bitstream and a decoding algorithm is reformulated for the hidden Markov model.—DAP

7,096,060

43.80.Qf METHOD AND SYSTEM FOR DETECTION OF HEART SOUNDS

Patricia Arand and David Lynn Burton, assignors to Innovise Medical, Incorporated
22 August 2006 (Class 600/513); filed 27 June 2003

A method and system are disclosed for automatically detecting heart sounds. A system receives sound data corresponding to beats of the heart. The sound system analyzes the sound data to detect the presence of a heart



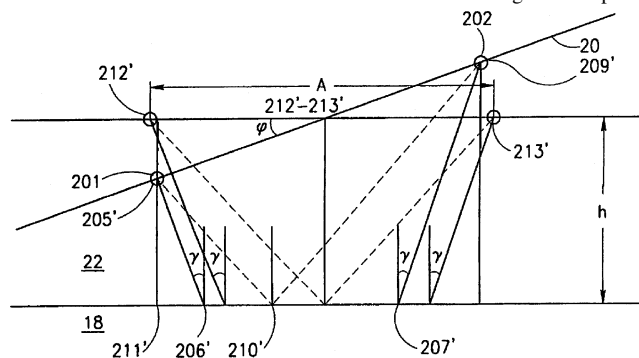
sound within the beats. It then outputs an indication of the heart sounds that were detected. The sound system may use ECG data to identify various locations (e.g., the R peak) within a beat and use those locations to assist in the detection of heart sounds.—DRR

7,112,173

43.80.Qf DETERMINATION OF ACOUSTIC VELOCITY IN BONE

Edward Kantorovich et al., assignors to Sunlight Medical Limited
26 September 2006 (Class 600/449); filed 24 June 1998

We have here a method for determining the acoustic velocity in a segment of a bone covered by a layer of soft tissue having an outer surface. The travel time of an ultrasonic wave is determined along three separate



paths from the outer surface back to the outer surface such that each path includes a different part of the bone segment. The acoustic velocity in the segment is derived from the three determined travel times.—DRR

7,101,336

43.80.Vj METHODS AND SYSTEMS FOR MOTION ADAPTIVE SPATIAL COMPOUNDING

Steven Charles Miller, assignor to General Electric Company
5 September 2006 (Class 600/443); filed 25 November 2003

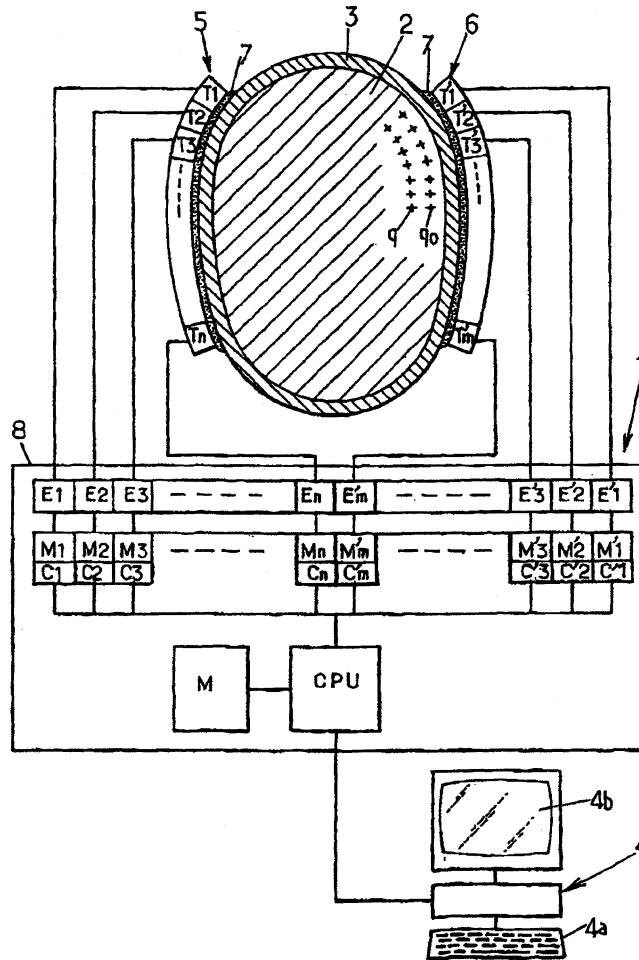
Ultrasound waves are transmitted into a volume and echoes are received for each of the transmitted waves. Sets of received echoes from a single transmitted wave are used to define a steering frame. Motion of the array transducer is detected and multiple steering frames are compounded based on the detected motion and rate of change of motion.—RCW

7,101,337

43.80.Vj METHOD AND NON-INVASIVE DEVICE FOR FOCUSING ACOUSTIC WAVES

Jean-Francois Aubry *et al.*, assignors to Centre National de la Recherche Scientifique—CNRS
5 September 2006 (Class 600/447); filed in France 20 October 2000

An imaging transducer array and a target transducer array are positioned on opposite sides of an inhomogeneous medium. Pulse responses of the medium are measured between each transducer of the imaging array and



7,103,205

43.80.Vj BREAST CANCER SCREENING WITH ULTRASOUND IMAGE OVERLAYS

Shih-Ping Wang and Fangyi Rao, assignors to U-Systems, Incorporated
5 September 2006 (Class 382/132); filed 27 November 2002

An ultrasound image representing a thick slice of breast volume parallel to an x-ray mammogram view is overlaid on the x-ray mammogram and registered manually.—RCW

7,104,956

43.80.Vj FINITE AMPLITUDE DISTORTION-BASED INHOMOGENEOUS PULSE ECHO ULTRASONIC IMAGING

Ted Christopher, assignor to Research Corporation Technologies, Incorporated
12 September 2006 (Class 600/443); filed 8 November 1996

An ultrasound signal with an amplitude that produces higher-order components is transmitted and the received signal is processed to produce an image from one of the higher-order components that are received.—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.

Comment on “The directionality of acoustic T-phase signals from small magnitude submarine earthquakes” [J. Acoust. Soc. Am. 119, 3669–3675 (2006)] (L)

Delwayne R. Bohnenstiehl^{a)}

Department of Marine, Earth and Atmospheric Sciences, North Carolina State University,
Campus Box 8208, Raleigh, North Carolina 27695-8208

(Received 19 September 2006; revised 26 December 2006; accepted 29 December 2006)

In a recent paper, Chapman and Marrett [J. Acoust. Soc. Am. 119, 3669–3675 (2006)] examined the tertiary (T-) waves associated with three subduction-related earthquakes within the South Fiji Basin. In that paper it is argued that acoustic energy is radiated into the sound channel by downslope propagation along abyssal seamounts and ridges that lie distant to the epicenter. A reexamination of the travel-time constraints indicates that this interpretation is not well supported. Rather, the propagation model that is described would require the high-amplitude T-wave components to be sourced well to the east of the region identified, along a relatively flat-lying seafloor. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2435982]

PACS number(s): 43.30.Ma, 43.30.Zk, 43.30.Wi [RAS]

Pages: 1293–1296

I. INTRODUCTION

Marine geophysical research has benefited greatly from the use of tertiary (T-) waves¹ in monitoring submarine earthquakes. The efficiency of sound propagation within the ocean, relative to the solid Earth, improves significantly our ability to detect smaller, and therefore more numerous, seismic events within remote ocean areas.² Although T-wave studies have made important contributions to our knowledge of submarine volcanism^{3,4} and fault system dynamics,^{5,6} our understanding of the influence of bathymetric features on T-wave source-radiator locations remains largely empirical in nature.

In the abyssal ridge-transform setting, T-wave-derived locations for shallow hypocenter earthquakes show a good correlation with morpho-tectonic features,^{7,8} with accuracy sufficient to lead field parties to sites of active volcanism.⁴ This suggests that T-wave source locations in this environment are well correlated with earthquake epicenters or, more accurately, moment centroids.⁹ In the subduction zone setting, longer solid-Earth paths for events inland from the trench can lead to greater complexity, with T-wave conversion points varying with azimuth to the station.^{5,10,11} Nonetheless, in this setting, the high-amplitude portion of the T-wave is sourced commonly from the ocean-facing slope between the trench and arc, at depths within the sound chan-

nel. For shallow-hypocenter events, typically the conversion point lies roughly between the epicenter and the hydrophone station.^{10,11}

The results recently presented by Chapman and Marrett,¹² however, suggest that T-waves generated from subduction-related earthquakes in the South Fiji Basin are sourced predominantly from shallow ridges and seamounts located on the abyssal plain at a distance of >100 km seaward the trench. Such findings argue T-wave-derived locations may be biased strongly by the presence of small-aperture bathymetric features even at regional distances from the epicenter. This comment reexamines the seismo-acoustic travel times that constrain the results of Chapman and Marrett.¹² Importantly, I highlight inconsistencies between the primary T-wave radiator zone identified in their manuscript and that actually computed using the travel-time model they propose. In short, I find the arrival times of the peak-amplitude T-wave components to be consistent with source regions proximal to the earthquake epicenters and show that travel-time constraints preclude the generation of these signals from the more distal topographic features identified by Chapman and Marrett.¹²

II. SOUTH FIJI BASIN EXPERIMENT

Chapman and Marrett¹² utilized data from a towed-hydrophone array deployed within the South Fiji Basin during a 12-day period in July of 1982. The physiography of the study area is dominated by the New Hebrides Trench and arc

^{a)}Electronic mail: drbohen@ncsu.edu

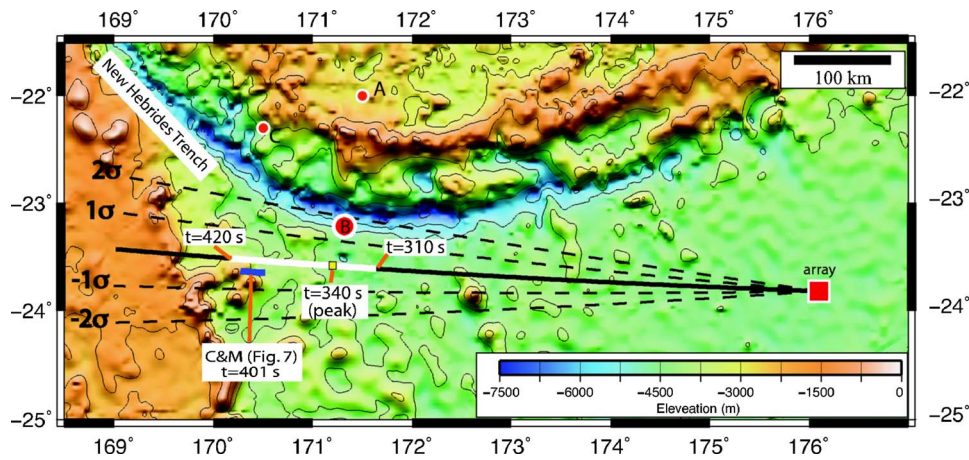


FIG. 1. (Color online) Modeled acoustic conversion points for the T-III arrival associated with earthquake event B. Catalogued epicenter is shown as a large circle. Smaller circles mark the locations of events A and C. Square marks the position of the towed array at the time event B was recorded. The P-to-T conversion point shown in Fig. 7 in Ref. 12 is labeled. A total travel time for P-wave energy to reach this point and for the converted T-wave to reach the array is 401 s (not 315 s as claimed in Table IV in Ref. 12). The solid black line shows a great circle acoustic path from the array at a bearing of 272° (Table III in Ref. 12). The white line shows the P-to-T conversion points required to explain acoustic energy arriving from that bearing within the time window of 310–420 s. Travel-time model assumes $V_p=7.5$ km/s and $V_T=1.5$ km/s. Times labeled represent total travel times relative to the event origin time. Dashed black lines show confidence limits for the azimuthal estimate ($1\sigma=3^\circ$).¹² Satellite-derived bathymetric data are shown with 1000 m contours.¹⁸ Data are displayed using a Mercator projection. Array location (23.818° S, 176.098° E) was derived from the information in Table I, in Ref. 12 assuming rhumb line bearings and distances as shown in their Fig. 7.

system, formed by the convergence of the Australia and New Hebrides Plates¹³ (Fig. 1). During the survey, the axis of the sound channel in the South Fiji Basin was located near a depth of 1250 m, with the waveguide being bottom-limited, except in the deeper portions of the trench (depths $< \sim 5200$ m).¹⁴

The hydrophone streamer utilized was 300 m long, with 32 elements. It was towed at a depth of ~ 400 m and a speed of 2–3 knots. Four depth and two compass sensors were used strictly to monitor the straightness of the array, with array heading determined relative to a 224-Hz beacon located ~ 400 km to the south of the survey area. Data were filtered in the 14–16-Hz range and then processed using 63 formed beams.

During its deployment, the system recorded three small earthquakes within the vicinity of the New Hebrides Trench at a distances of $\sim 5^\circ$ from the array (Fig. 1). The seismic locations of these events (Table 1 of Ref. 12) are constrained by six to eight defining phases, with maximum azimuthal gaps of $206\text{--}308^\circ$.¹⁵ No magnitudes were reported for events A and C, and they are therefore assumed to be smaller than $\sim 4.0 m_b$.¹² A single-station magnitude estimate of $4.7 m_b$ was reported for event B. Catalogued depths for events A and B are listed at 33 km (fixed). A depth of 59.3 km is listed event C; however, no constraining depth phases (e.g., pP, sP) are reported.¹⁵ Throughout the manuscript, Chapman and Marrett¹² use the largest, and best-recorded, earthquake (event B) as a type-example.

The arrival patterns described at the hydrophone array consist of the solid-earth P and S phases, followed by what the authors interpret as four distinct acoustic phases that persist throughout the arrival train (T-I, -II, -III, -IV). The first and smallest amplitude acoustic phase (T-I) arrives on an approximate bearing to the catalogued epicenters, but is interpreted (based on travel-time constraints) as refracted seismic energy that converts through scattering in the vicinity of

the towed array. The subsequent components all arrive at bearings oriented counter-clockwise (to the south) to the catalogued epicenters and are interpreted as downslope converted phases originating from seamounts and ridges to the southwest of the subduction boundary. Of the T-wave components described, T-III is clearly the highest amplitude and longest duration signal, with T-I, T-II, and T-IV being difficult to identify in the single-channel hydrophone data (Fig. 2). This dominant portion of the acoustic arrival (T-III) typically would be used in location analysis; therefore, its source location and mechanism of generation will be the focus of my comment.

III. TRAVEL-TIME CONSTRAINTS ON THE T-WAVE SOURCE REGION

Chapman and Marrett suggest that the dominant T-wave arrival (T-III) observed following the three earthquakes represents energy radiated into the water by downslope propagation in the vicinity of a seamount-ridge complex that lies to the west of the array (Fig. 1). This is consistent with their azimuthal constraints, which prescribe a westerly source region. They propose a simple travel-time model, with P (7.5 km/s) and S (4.5 km/s) phases propagating from the catalogued seismic epicenters to this conversion point, followed by an acoustic (1.5 km/s) path between the conversion point and hydrophone array. These solid-Earth velocities are consistent with the apparent velocity inferred from an IASPEI model at regional distances, and a speed of 1.5 km/s is a reasonable approximation for the velocity of an acoustic wave within the sound channel.

Using event B as an example, they show the region of proposed T-III conversion in the map view (Ref. 12, Fig. 7). This point is reproduced in Fig. 1 in this work. Although it is claimed this P-to-T conversion point is consistent with a total travel time of 315 s (Ref. 12, Table 4), the actual travel time

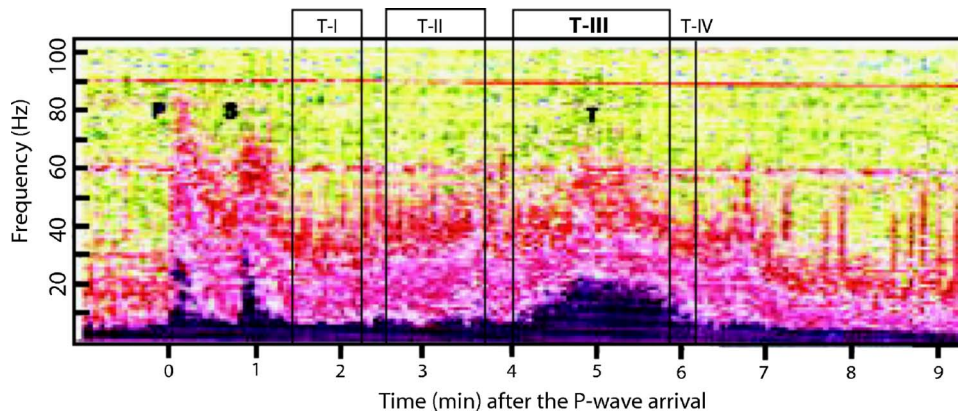


FIG. 2. (Color online) Spectrogram of a single hydrophone recording for event B (after Fig. 4 in Ref. 12). The horizontal time axis is referenced to the P-wave arrival. Note that the times given in this comment and throughout most of the original manuscript reference to the event origin time (68 s before the P-wave arrival). The figure has been modified to indicate the arrival windows for the T-wave component given in Table IV in Ref. 12. The phase labeled T-III is clearly the dominant acoustic signal and includes the peak amplitude arrival ~ 4.5 min after the P-wave. Additional arrival components were inferred from the beam intensity plot; however, the tabulated duration of T-I and T-II (as shown above) appears to exceed the time windows of significant power within these data (see Fig. 5 in Ref. 12). The spectral data are dominated by flow noise below ~ 8 Hz.¹²

using the proposed model is 401 s ($t_p=14$ s, $t_T=387$ s). This places the first arriving energy from the seamount near the tail end of the 310–420-s T-III arrival window and well after the peak amplitude is reached at 340 s. Consequently, it cannot be argued that the T-III arrival was converted from a seismic phase by downslope propagation along the seamount complex identified (i.e., following a series of reflections between the sloped seafloor and sea surface that progressively turn energy into the sound channel).

In Fig. 1, I have calculated the predicted conversion points for T-III arrivals along a great-circle path using the same velocity model, arrival-time data, and azimuth constraints given in the paper by Chapman and Marrett.¹² This modeling indicates that the first arriving T-III energy would have been sourced ~ 130 km to the east of the point identified by the authors. In fact, even if the hypocenter for event B were positioned beneath the seamount ($t_p \sim 0$), downslope-converted phases could not reach the towed array within the time required to explain the onset and peak of the T-III signal. If a S-to-T coupling is assumed, the predicted travel time for a conversion point near the seamount becomes greater and the fit to the observations even worse.

Repeating this calculation for the dominant (T-III) signal components associated with events A and C, I find the same discrepancy between modeled seismic-to-acoustic conversion points and those identified in the manuscript.¹² In these cases it also is clear that the proposed travel-time model requires T-wave conversion to the east of the region identified, on a predominantly flat-lying abyssal seafloor.

IV. AN ALTERNATIVE INTERPRETATION

Given that travel-time constraints preclude the generation of the T-III signals as described by Chapman and Marrett,¹² it is useful to consider what propagation history best satisfies both the travel-time and azimuthal observations. Ideally, any model should also be consistent with the seismo-tectonic framework of the region.

For event B, the arrival time of the T-III peak (340 s) is consistent with a purely acoustic path from a source ~ 15 km

beyond the published epicenter, at a distance within the uncertainly limits of the seismic location. Moreover, the azimuth from the array to the catalogued epicenter lies within 2σ (6°) of the center-beam azimuth published by Chapman and Marrett.¹² I therefore propose that the peak of the T-III signal originates in the near epicentral region, where scattered T-wave energy exhibits the largest amplitudes.¹⁶ Earlier arriving portions of the signal are scattered from the region between the epicenter and receiver, and the T-III coda is sourced from a seafloor more distal to the epicenter and may include energy that is reflected and/or downslope converted along the trench wall.¹⁷ The outer-rise region in the vicinity and to the west of event B is a seismically active area where shallow hypocenters with normal and oblique-slip mechanisms are common (See Fig. 10 of Ref. 13); the tectonics of this area in part reflect the transition from arc-normal to highly oblique subduction as the Hew Hebrides Trench curves.

Acoustic travel-time and azimuthal constraints for the smaller events A and C similarly are consistent with T-wave sources within this outer-rise region to the west of event B, or, alternatively, if the stated azimuthal error bounds were relaxed by a few degrees, these events could be sourced from the southward facing trench wall. Source locations along the trench wall would be more consistent with the published hypocenters beneath the arc,¹⁵ as well as the pattern typically observed from earthquake events generated along the subduction interface.^{10,11}

Regardless, my reassessment shows that the high-amplitude portions of the observed T-wave arrivals cannot be sourced from the area originally identified. The distances between the proposed point of conversion and array are simply too great. My alternative interpretation not only provides consistency with the travel-time constraints, but also avoids the complexity of efficiently converting southwest-polarized P-wave energy along an east-facing slope, as required for the model originally envisioned.¹² Moreover, it does not necessitate that the seismic-to-acoustic conversion occur preferentially in association with a specific seamount, which does not

appear to be particularly steep-sided or shallow relative to other bathymetric features at similar distances within the arc and abyssal plain.

- ¹I. Tolstoy and M. Ewing, "The T phase of shallow-focus earthquakes," *Bull. Seismol. Soc. Am.* **40**, 25–51 (1950).
- ²C. G. Fox, R. P. Dziak, H. Matsumoto, and A. E. Schreiner, "Potential for hydrophone arrays," *Mar. Technol. Soc. J.* **27**, 22–30 (1994).
- ³J. Talandier and E. A. Okal, "Monochromatic Twaves from underwater volcanoes in the Pacific Ocean: Ringing witnesses to geyser processes?" *Bull. Seismol. Soc. Am.* **86**, 1529–1544 (1996).
- ⁴R. P. Dziak and C. G. Fox, "The January 1998 Earthquake Swarm at Axial Volcano, Juan de Fuca Ridge: Evidence for submarine volcanic activity," *Geophys. Res. Lett.* **26**, 3429–3432 (1999).
- ⁵M. Tolstoy and D. R. Bohnenstiehl, "Hydroacoustic contributions to understanding the December 26th 2004 Great Sumatra-Andaman Earthquake" *Surv. Geophys.* **27**(5), 633–646 (2006).
- ⁶J. J. McGuire, M. Boettcher, and T. H. Jordan, "Foreshock sequences and earthquake predictability on East Pacific rise transform faults," *Nature (London)* **434**, 457–461 (2005).
- ⁷C. G. Fox, H. Matsumoto, and T.-K. A. Lau, "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," *J. Geophys. Res.* **106**, 4183–4206 (2001).
- ⁸D. R. Bohnenstiehl, M. Tolstoy, D. K. Smith, C. G. Fox, and R. Dziak, "The decay rate of aftershock sequences in the mid-ocean ridge environment: An analysis using hydroacoustic data," *Tectonophysics* **354**, 49–70 (2002).
- ⁹J. Pan and A. M. Dziewonski, "Comparison of mid-oceanic earthquake epicentral differences of travel time, centroid locations, and those determined by autonomous underwater hydrophone arrays," *J. Geophys. Res.* **110**, B07302 (2005).
- ¹⁰F. M. Graeber and P.-F. Piserchia, "Zones of T-wave excitation in the NE Indian Ocean mapped using variations in backazimuth over time obtained from multi-channel correlation of IMS hydrophone triplet data," *Geophys. J. Int.* **158**, 239–256 (2004).
- ¹¹C. deGroot Hedlin and J. A. Orcutt, "Excitation of T-waves by seafloor scattering," *J. Acoust. Soc. Am.* **109**, 1944–1954 (2001).
- ¹²N. R. Chapman and R. Marrett, "The directionality of acoustic T-phase signals from small magnitude submarine earthquakes," *J. Acoust. Soc. Am.* **119**, 3669–3675 (2006).
- ¹³P. Bird, "An updated digital model of plate boundaries," *Geochem., Geophys., Geosyst.* **4**, 1027 (2003).
- ¹⁴Generalized Digital Environmental Model (GDEM) Version 3. Available online <https://128.160.23.42/gdemv/gdemv.html>, viewed online: 10 September 2006.
- ¹⁵International Seismological Centre, On-line Bulletin, <http://www.isc.ac.uk/Bull.Internatl.Seis.Cent.>, Thatcham, United Kingdom, viewed online 10 September 2006.
- ¹⁶P. D. Slack, C. G. Fox, and R. P. Dziak, "P wave detection thresholds, Pn velocity estimates, and T wave location uncertainty from oceanic hydrophones," *J. Geophys. Res.* **104**, 13061–13072 (1999).
- ¹⁷R. H. Johnson, R. A. Norris, and F. K. Duennebier, "Abyssally generated T-phases," in *The Crust and Upper Mantle of the Pacific Area*, edited by L. Knopoff, C. L. Drake, and P. J. Hart, AGU Monograph **12**, pp. 70–78 (1968).
- ¹⁸W. H. F. Smith and D. T. Sandwell, "Global seafloor topography from satellite altimetry and ship depth soundings," *Science* **277**, 1957–1962 (1997).

Reply to “Comment on ‘The directionality of acoustic T-phase signals from small magnitude submarine earthquakes’ [J. Acoust. Soc. Am. 119, 3669–3675 (2006)]” (L)

N. Ross Chapman^{a)}

*School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055,
Victoria, BC V8W 3P6 Canada*

R. Marrett

Defence Technology Agency, Auckland, New Zealand

(Received 11 December 2006; accepted 21 December 2006)

D. Bohnenstiehl has raised important issues about the interpretation of the T-phase directionality data presented in the paper by Chapman and Marrett [N. R. Chapman and R. Marrett, *J. Acoust. Soc. Am.* **119**, 3669–3675 (2006)]. The central point in that Comment is that the maximum component in the T-phase signal is radiated directly from the earthquake source location. This Reply points out that the results in the original paper are not inconsistent with Bohnenstiehl’s model; moreover, the array directionality data contain significantly more information that can be useful in understanding the influence of bathymetric features in the radiation of T-phase sound by an earthquake.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2435984]

PACS number(s): 43.30.Ma, 43.30.Cq, 43.30.Wi, 43.30.Zk [RAS]

Pages: 1297–1298

D. Bohnenstiehl has raised important issues about the interpretation of the T-phase directionality data presented in our paper.¹ The central point in his Comment is that the maximum component in the T-phase signal is radiated directly from the earthquake source location.² We would like to point out in this Reply that our results are not inconsistent with Bohnenstiehl’s model; however, our data contain significantly more information that can be useful in understanding the influence of bathymetric features in the radiation of T-phase sound by an earthquake. It is not beyond reason to suggest that towed array directionality data may be inverted to refine the locations of earthquake events, as well as those of T-phase radiation sites at a distance from the earthquake source.

Our motivation for writing the paper was to point out the advantages of towed hydrophone line array data compared to single hydrophone data for studying the influence of bathymetric features on source-radiator locations of acoustic T-phases. The hydrophone array provides directional gain that can be used to resolve the directions of T-phases throughout the duration of the signal. By comparison, the only information available from single hydrophone data is the timing of the signal peak; this provides no directional information in itself but it can be used in triangulation algorithms to locate the T-phase source.

Revision of the original manuscript benefited greatly from several discussions with the author of the Comment, who acted as a reviewer of our paper. In consideration of the questions raised in his review, the original directionality data from the experiment were reexamined. Figures 6(a) and (b) in the original work presented the reexamined results for

array beam signal intensity and directionality for earthquake event B.¹ The beam directions were obtained from the array frequency-azimuth information averaged over 10 s, using T-phase spectral data over a broad frequency band from 10 to 60 Hz. The signal-to-noise ratios (SNR) of the beam-formed data were very high, varying from at least 40 dB at the low frequencies to around 20 dB at 60 Hz. This combination of high SNR and high frequency response allowed very high resolution of the T-phase directions in the frequency-azimuth data, to within a beam.

The signal direction data for the three events are presented in Table III of our paper¹ and are briefly summarized here. The initial strong component of the T-phase (identified as T-II in that paper) was received on a bearing of $\sim 260^\circ$, well south of the bearings to the event locations. This behavior was consistent for all three events. Later arrivals near the end of the T-phase signal (identified as T-IV) arrived on bearings from 260° to 265° ; this behavior was also consistent for all three events. As implied in the Comment by Bohnenstiehl, it is not possible to derive this directional information about the early and later stages of the signal from the single hydrophone data because the single-channel SNR is too low to resolve the separate components in time. Simple travel time models are consistent with T-phase radiation sites from bathymetric features in the basin south of the ridge for these components.

Bohnenstiehl’s Comment focuses on only the T-phase signal maximum and argues on the basis of the single hydrophone timing data that the maximum signal is received from the earthquake locations. As he points out, this interpretation is consistent with the behavior observed for T-phases from various other earthquakes, and it is the basis for using T-phases in studying fault rupture processes.

^{a)}Electronic mail: chapman@uvic.ca

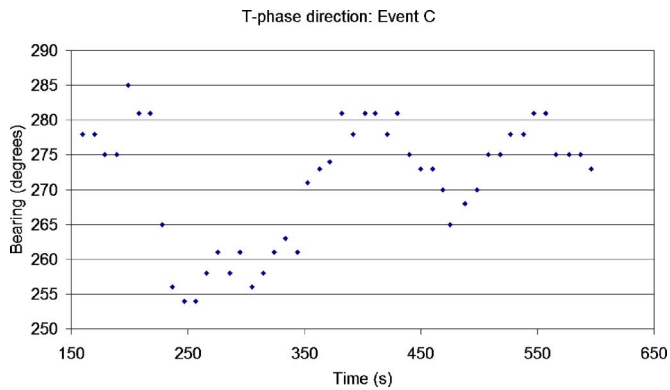


FIG. 1. T-phase directionality for event C, located beneath the ridge to the north of the basin. The bearing to the earthquake location is 284° . The origin of the time axis is the onset time of the event.

Our results are not inconsistent with this simple interpretation for the source of the maximum signal component. However, our analysis indicates that only a small fraction of the T-phase signal radiates directly from the earthquake source. For example, Fig. 1 shows the T-phase signal directionality for event C. The first part of the signal arrives from the direction of the earthquake ($\sim 284^\circ$ for event C); subsequent (and stronger) signal components are generated at sites distant to the earthquake along bearings of 260° (T-II, beginning at 240 s), $\sim 272^\circ$ (T-III, beginning at 350 s), and $\sim 265^\circ$ (T-IV, beginning at 475 s). Around 380 s the bearing shifts to $\sim 280^\circ$, close to the bearing of the earthquake source; the signal is at maximum intensity around this time. Since this time corresponds to the travel time for signal propagation from the source location at the speed of sound in water, this component is likely radiated directly from the source.

However, for event B the relationship between signal bearing and travel time near the T-phase signal maximum is more complicated. The event bearing is $\sim 280^\circ$, and Bohnenstiehl points out correctly that the direct path travel time is around 340 s^2 . From our analysis, the maximum signal is

received over a time from 338–394 s [Fig. 6(b) in Ref. 1], but the T-phase signal bearing over most of this period (from 315 to 386 s) is 272° , distinct from the direction to the earthquake source [Fig. 6(a) in Ref. 1]. After this period, the signal bearing then shifts to the direction of the earthquake source between 395 to 435 s. The onset time of the shift is much later than the direct path travel time. Consequently, the time-intensity data are not entirely consistent with the simple interpretation proposed by Bohnenstiehl for this event.

Although it may be difficult to resolve this discrepancy by arguments based on simple travel time models, we can point out some simple features about event B that differentiate it from events A and C. Since events A and C occur beneath the northern ridge, T-phase radiation directly from these locations is likely by downslope conversion. By comparison, since event B is located beneath the deep trench, the coupling with the sound channel by scattering from the abyssal basin may not be as effective. Close inspection of the map shown by Bohnenstiehl in his Comment shows small features in the abyssal region just south of the event location that could be more effective radiators.²

The array directionality data reveal a complex but similar pattern in the T-phase signal for the events observed in our experiment. Our analysis of the array data showed that a significant fraction of the T-phase signal was radiated from sites distant from the earthquake location. From those data and simple travel time arguments, the signal generation sites for prominent signal components in the beamformed data can be related to bathymetric features, some of which lie in the basin south of the event locations.

¹N. R. Chapman and R. Marrett, "The directionality of acoustic T-phase signals from small magnitude submarine earthquakes," *J. Acoust. Soc. Am.* **119**, 3669–3675 (2006).

²D. R. Bohnenstiehl, "Comment on 'The directionality of acoustic T-phase signals from small magnitude submarine earthquakes' [*J. Acoust. Soc. Am.* **119**, 3669–3675 (2006)]," *J. Acoust. Soc. Am.* **121**, 1293–1296 (2007).

A pseudo-inverse algorithm for simultaneous measurements using multiple acoustical sources (L)^{a)}

Ning Xiang^{b)} and Shu Li

Graduate Program in Architectural Acoustics, School of Architecture, Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 28 October 2006; revised 23 December 2006; accepted 27 December 2006)

Simultaneous multiple acoustical sources measurement (SMASM) has been proposed for more effective and reliable identification of acoustical systems under critical conditions [N. Xiang and M. R. Schroeder, *J. Acoust. Soc. Am.* **113**, 2754–2761 (2003); N. Xiang, J. N. Daigle, and M. Kleiner, *J. Acoust. Soc. Am.* **117**, 1889–1894 (2005)]. This paper presents a pseudo-inverse algorithm for the SMASM correlation technique as an alternative way of extracting impulse responses of acoustical channels. Simulations and room acoustics experiments are carried out and the results prove the feasibility of the proposed algorithm.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2434763]

PACS number(s): 43.60.Vx, 43.60.Ek, 43.58.Gn [EJS]

Pages: 1299–1302

I. INTRODUCTION

This paper derives a new algorithm for simultaneous acoustic measurements using multiple sound/vibration sources. The simultaneous multiple acoustic sources measurement (SMASM) uses several acoustic sources at the same time to determine impulse responses of linear time invariant systems in the same frequency range. The technique can be used in acoustic delay-time tomography as to investigate sound propagation near the ground surface in outdoor environment.^{1,2} Also in room acoustics, room-acoustic parameters are derived from room impulse responses between loudspeakers and microphones. Due to the simultaneous excitation of multiple sound sources and one or multiple sound receivers, the SMASM considers the acoustic system under test as a linear time-invariant multiple-inputs and multiple-outputs (MIMO) system.

In recent years, the maximum length sequences (MLS) technique proposed by Schroeder³ serves as an effective method in acoustical system identification. Based on MLS's excellent correlation properties, MLS technique is highly immune to extraneous noise and provides a repeatable solution with high accuracy for the single-input and single-output system identification.⁴ In the SMASM technique, the MIMO system identification requires several simultaneous excitations associated with multiple acoustic sources within the same frequency range. A previous work⁵ presented a technique using reciprocal MLS pairs as excitations in dual-channel measurements using two simultaneous acoustical sources. However, for the complex MIMO system with more sound sources, excitation signal classes with similar correlation properties, but more than two channels have to be adopted to meet the practical requirements of SMASM. The autocorrelation of one coded signal, such as Gold or Kasami

sequences,^{6–9} is a pulse-like function, while the cross correlation between any two different coded signals in these signal classes is a function with small values relative to the peak value of the autocorrelation function. With respect to this correlation property, a large number of coded signals can be easily derived for the SMASM. Among currently available MLS and MLS-related sequences as applied to a broad range,^{8,9} Kasami sequences possess the lowest cross-correlation bound value.⁵ This paper uses Kasami sequences as the excitations for the purpose of algorithm verification.

A recent work¹⁰ proposed a specialized fast cross-correlation method for extracting acoustic channel impulse responses. Based on this specialized cross-correlation method, this paper derives a novel algorithm for SMASM using coded signals, such as MLS and MLS-related classes. Different from the specialized fast cross correlation as done in Ref. 10, impulse responses of the MIMO system can be determined using the pseudo-inverse algorithm. The coded signals are widely applied in spread spectrum communications.⁹ The pseudo-inverse algorithm has not been documented in major acoustical journals. This paper along with Ref. 10 may piece together a coherent understanding of efficient algorithms for the simultaneous measurements using multiple acoustic sources, which can meet critical needs of some acoustical applications.^{1,2}

II. PSEUDO-INVERSE ALGORITHM

A. Simultaneous multiple acoustic sources measurement

Figure 1 illustrates a SMASM scheme, where the vector $\mathbf{s}=(s_1, \dots, s_n)^T$ stands for the multiple coded signals as system's excitations, and $\mathbf{r}=(r_1, \dots, r_p)^T$ denotes the system's responses to these excitations, with $()^T$ standing for matrix transpose; n is the number of simultaneous sources, while p is the number of receivers.

With purposely selected excitations, the system identification task is to determine the impulse response matrix \mathbf{H}

^{a)}Aspects of this work have been presented at the 151st ASA Meeting, *J. Acoust. Soc. Am.* **119**, 3272 (2006) (A).

^{b)}Author to whom correspondence should be addressed. Electronic mail: xiangn@rpi.edu

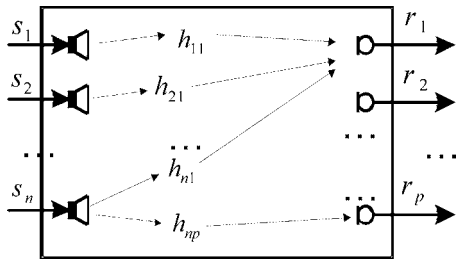


FIG. 1. Simultaneous multiple acoustic sources measurement (SMASM). In SMASM, multiple excitations with acoustical sources s_1, s_2, \dots, s_n are applied to drive the system under test at the same time; several receivers r_1, r_2, \dots, r_p are receiving the responses. Acoustical impulse responses h_{ij} are defined between this multiple-input and multiple-output system.

$= [h_{ij}]$ for $1 \leq i \leq n, 1 \leq j \leq p$. The MIMO system under simultaneous excitations of multiple sources is represented as

$$\mathbf{r} = \mathbf{H} * \mathbf{s} \quad (1)$$

with $*$ denoting linear convolution.

B. Pseudo-inverse algorithm

The periodic autocorrelation function of binary coded signal is a pulse-like function and their periodic cross-correlation functions (PCCFs) are of small values.¹¹ A straightforward way is to correlate \mathbf{s} at both sides of Eq. (1). When $\mathbf{s} \otimes \mathbf{s} = \mathbf{\Delta}$, matrix \mathbf{H} can be determined as the cross-correlation between responses and excitations, where $\mathbf{\Delta}$ is a diagonal matrix with a unit sample sequence $\delta(t)$ being each of its diagonal elements, \otimes denotes the periodic cross-correlation.

However, this solution met two major difficulties in practice. One is the direct correlation for SMASM is very time-consuming with n sources and p receivers. A Hadamard transform based MLS transform^{4,5} is no more applicable for coded signals such as Gold or Kasami sequences. For this reason, the recent work proposed a specialized correlation algorithm,¹⁰ in order to exploit the fast Fourier transform (FFT) for efficient correlation processing of coded signals of a length being not suitable for FFT. Inspired by the specialized FFT deployment in Ref. 10, this paper derives a pseudo-inverse approach.

The other difficulty lies in the error caused by the assumption of $\mathbf{s} \otimes \mathbf{s} = \mathbf{\Delta}$. Small PCCF bound values of the coded signals,^{7,8} imply $\mathbf{s} \otimes \mathbf{s} \approx \mathbf{\Delta}$, introducing inaccuracy into impulse response extraction using the direct cross-correlation method.⁷

The MLS and MLS-related coded signals are of length $2^N - 1$, with N being a positive integer. To exploit the FFT based on the specialized algorithm,¹⁰ original responses and excitations are both appended with $2^N + 1$ zeros, so, s_i and r_j are both augmented to form new sequences s'_i, r'_j whose lengths are $L' = 2^{(N+1)}$, being suitable for FFT_(N+1), the subscript $N+1$ of the FFT, and later of the IFFT, explicitly denotes the length of $L' = 2^{(N+1)}$. With the complex signals $\mathbf{R}' = [R'_1, \dots, R'_p]$ and $\mathbf{S}' = [S'_1, \dots, S'_n]$ in frequency domain after FFT_(N+1), Eq. (1) is expressed as

$$\mathbf{R}' = \mathbf{H}' \cdot \mathbf{S}' \quad (2)$$

Multiplying \mathbf{S}'^* , the conjugated form of \mathbf{S}' , on both sides of Eq. (2) yields

with $\mathbf{\Phi} = \mathbf{S}' \cdot \mathbf{S}'^*$, $\mathbf{\Phi}$ is a n by n by L' matrix. Each element of matrix $\mathbf{\Phi}$ is a vector with L' items. Let $\mathbf{\Phi}^\#$ be a pseudo-inverse matrix of $\mathbf{\Phi}$, matrix \mathbf{H}' is approximated by

$$\mathbf{H}' \approx \mathbf{R}' \cdot \mathbf{S}'^* \cdot \mathbf{\Phi}^\# \quad (4)$$

To avoid possible singularities, singular value decomposition (SVD) is used to calculate the pseudo-inverse matrix. Since the excitation signals are known prior to the acoustical measurement, matrix

$$\mathbf{K} = \mathbf{S}'^* \cdot \mathbf{\Phi}^\# = \mathbf{S}'^* \cdot (\mathbf{S}' \cdot \mathbf{S}'^*)^\# \quad (5)$$

needs to be calculated only once in advance and saved in memory, so that

$$\mathbf{H}' \approx \mathbf{R}' \cdot \mathbf{K} \quad (6)$$

Applying the inverse FFT of length L' to each element in \mathbf{H}' yields $h'_{ij} = \text{IFFT}_{(N+1)}(H'_{ij})$ with $i = 1, \dots, n; j = 1, \dots, p$.

According to the specialized correlation algorithm proposed in a recent paper,¹⁰ a point-wise addition of the first L points with the last L points of h'_{ij} :

$$h_{ij}(k) = h'_{ij}(k) + h'_{ij}(L' - k), \quad 0 \leq k \leq L \quad (7)$$

yields the impulse response h_{ij} in time domain for the channel between i th source and j th receiver. Impulse responses can be obtained by convoluting the responses of the system under test with the matrix \mathbf{K} using Eq. (6) and Eq. (7). Since matrix \mathbf{K} is prepared in advance and saved in memory, this algorithm is equally efficient as that of the specialized correlation algorithm in Ref. 10 in terms of computational loads when a large number of measurements using the same multiple excitation signals is done.

III. DIGITAL SIMULATIONS

In order to verify the feasibility of the pseudo-inverse algorithm in the SMASM application, digital simulations are carried out to mimic practical measurements. Eighth-order low-pass Chebyshev filters with a cut-off frequency of 18 kHz are adopted to simulate the acoustic channels at a sampling frequency of 50 kHz. Each excitation is shifted for a certain, but different delay time to simulate the different distances from acoustical sources to microphones. Each sequence is filtered by the low-pass filter, and finally the filtered signals are summed up as a collected signal for one receiver. In this way, the simulated, multiple responses of SMASM are obtained. The pseudo-inverse algorithm is used to extract the impulse responses in simulation (similar to Ref. 7). Figure 2 illustrates a set of simulation results employing Gold sequences as excitations, by changing the length of Gold sequences and the number of simultaneous excitations, and impulse responses are obtained. The peak to noise ratio (PNR) of impulse responses as defined in Ref. 10 are illustrated in Fig. 2.

In Fig. 2, degree 12, e.g., denotes Gold sequences of length $2^{12} - 1$. Obviously, the longer the coded signals, the better the PNR of the impulse responses will be. Also, the more simultaneous channels are adopted, the lower the PNR

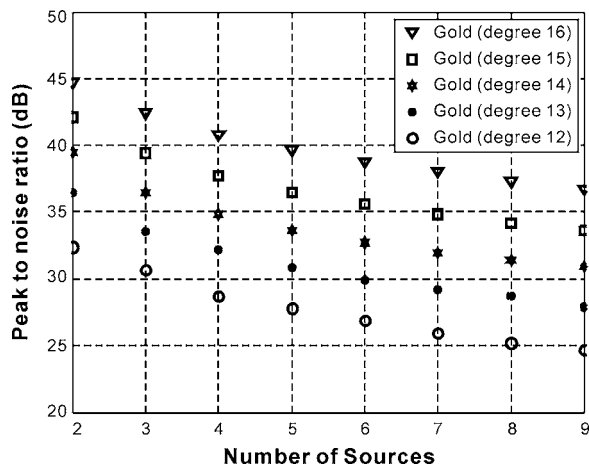


FIG. 2. Peak-to-noise ratio of impulse responses as functions of number of simultaneous sources using Gold sequences of length from $2^{12}-1$ to $2^{16}-1$. Digital low-pass filters are used to simulate acoustical channels.

becomes, due to channel cross-talk noise. Similar simulations are also undertaken with Kasami sequences of different lengths; the PNR of corresponding impulse responses with respect to number of simultaneous excitations are illustrated in Fig. 3. Using Kasami sequences, the SAMSM is able to reach a higher PNR in impulse responses compared with Gold sequences with the same degree.

IV. EXPERIMENTAL RESULTS

The pseudo-inverse algorithm derived above has been used in practical acoustical measurements for verification. Coded signals are directly fed into sources at an update rate of 50 kHz to cover the frequency range of interest. The room acoustical measurement of room impulse responses is carried out in San Patrick Church, Watervliet, New York. In the measurements, four loudspeakers are used as the acoustical sources driven by four Kasami sequences of degree 22, their corresponding matrix \mathbf{K} in Eq. (6) was prepared in advance. At the receiving end, a binaural artificial-head system with two microphones (left, right ears) is used to capture the

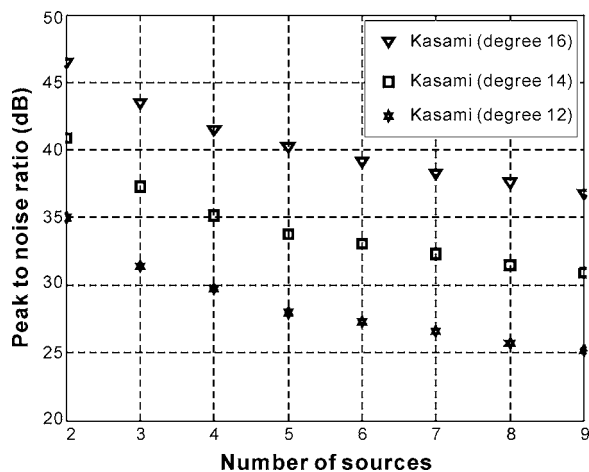


FIG. 3. Peak-to-noise ratio of impulse responses as functions of number of simultaneous sources using Kasami sequences of length $2^{12}-1$, $2^{14}-1$ and $2^{16}-1$. Digital low-pass filters are used to simulate acoustical channels.

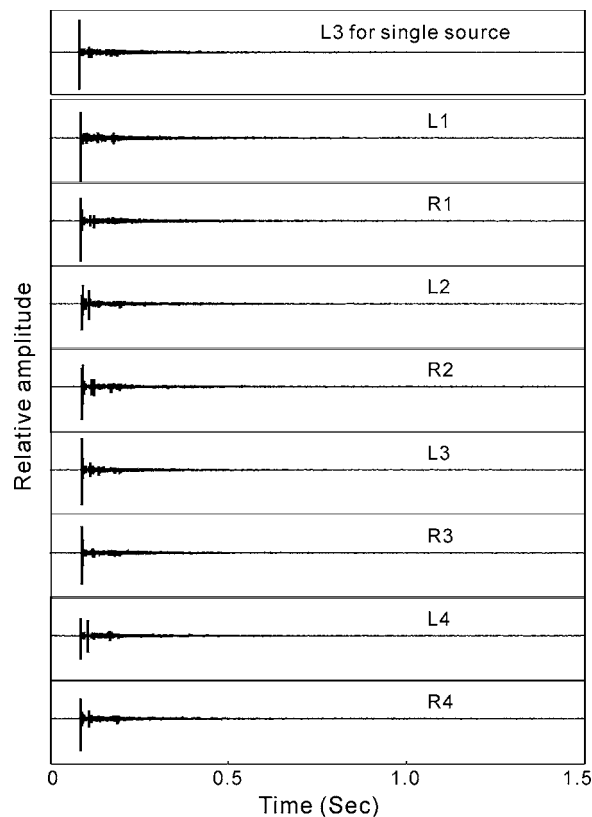


FIG. 4. First 1.5 s of room impulse responses determined by the pseudo-inverse algorithm. Kasami sequences of degree 22 serve as excitations. The figure at the top shows room impulse response of conventional single source measurement corresponding to L3 channel. The other part shows eight room impulse responses of MIMO system with four loudspeakers and two ear microphones; L_i denotes, e.g., the impulse response between source 1 and the left-ear microphone.

room's steady-state responses to the excitations. The sampling rate of the measurement platform is set to 50 kHz.

After the measurement platform records the whole steady-state responses of the SMASM, the pre-prepared matrix \mathbf{K} is correlated with the acoustic responses yielding room impulse responses using the pseudo-inverse algorithm [Eqs. (5)–(7)]. Figure 4 illustrates eight room impulse responses (first 1.5 s segments) for channels between the four sound sources and the left and right ear microphones. A single-source impulse response, corresponding to L3, but achieved from one loudspeaker-microphone system is also plotted at the top of Fig. 4 for comparison.

The PNR of impulse responses is primarily determined by the PCCF bounds.⁷ It is also influenced by the available source power, the sensitivity and bandwidth of receiver, the attenuation of the propagation channel, etc.^{7,11} Nonlinear and time-variant components within the system under test and the background noise will also derogate the measurement accuracy. In this experiment with four simultaneous acoustical sources, the PNR of all room impulse responses can be higher than 45 dB.

V. CONCLUSIONS

This paper proposes an alternative algorithm to extract impulse responses of multiple measurements using coded

signals to excite simultaneous sources. Pseudo-inverse algorithm provides a solution for the simultaneous measurement of multiple acoustical channels and implements the extraction of all impulse responses. The proposed algorithm can effectively separate the impulse response of source-receiver pair and characterize this channel even when each single receiver captures the compound responses to several simultaneous sources. The effectiveness of the proposed algorithm is proved by experimental room-acoustical measurements. The pseudo-inverse algorithm proposed in this paper relies on domain transform and the singular value decomposition (SVD); its effectiveness in acoustical measurements for multiple-input and multiple-output systems has been demonstrated. In contrast to the direct cross correlation, it will promise a chance of improving the signal-to-noise ratio using more stable SVD or other numerical techniques inside the current algorithm.

¹A. Ziemann, K. Arnold, and A. Raabe, "Acoustic tomography as a method to identify small-scale land surface characteristics," *Acta. Acust. Acust.* **87**, 731–737 (2001).

²D. K. Wilson, A. Ziemann, V. E. Ostashev *et al.*, "An overview of acoustic travel-time tomography in the atmosphere and its potential applications," *Acta. Acust. Acust.* **87**, 721–730 (2001).

³M. R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Am.* **66**, 497–500 (1979).

⁴J. Borish and J. B. Angell, "An efficient algorithm measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc.* **31**, 478–488 (1983).

⁵N. Xiang and M. R. Schroeder, "Reciprocal maximum-length sequence pairs for acoustic dual source measurements," *J. Acoust. Soc. Am.* **113**, 2754–2761 (2003).

⁶R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions," *IEEE Trans. Inf. Theory* **14**, 154–156 (1968).

⁷N. Xiang, J. N. Daigle, and M. Kleiner, "Simultaneous acoustic channel measurement via maximal-length-related sequences," *J. Acoust. Soc. Am.* **117**, 1889–1894 (2005).

⁸D. V. Sarwate and M. B. Pursley, "Properties of pseudorandom and related sequences," *Proc. IEEE* **68**, 593–619 (1980).

⁹S. W. Golomb and G. Gong, *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar* (Cambridge University Press, Cambridge, 2005).

¹⁰J. N. Daigle and N. Xiang, "A specialized fast cross correlation for acoustical measurements using coded sequences," *J. Acoust. Soc. Am.* **119**, 300–335 (2006).

¹¹T. G. Birdsall and K. Metzger, "Factor inverse matched filtering," *J. Acoust. Soc. Am.* **79**, 91–99 (1986).

Characteristics of penalty mode scattering by rigid splices in lined ducts

WenPing Bi,^{a)} Vincent Pagneux, Denis Lafarge, and Yves Aurégan

Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Av. O. Messiaen, 72085 Le Mans Cedex 9, France

(Received 19 July 2006; revised 6 December 2006; accepted 16 December 2006)

In lined ducts, incident modes are scattered by axially and circumferentially nonuniform impedance. Experiments and numerical calculations have proved that this mode scattering can reduce the liner performance in some cases. This paper is devoted to the characterization of the penalty mode scattering excited by hard-walled splices which often exist in lined ducts. It is shown that, in the range of small splice angles, the transmission loss may decrease sharply with increasing splice angle when one mode, which is near cut-off or has high azimuthal order, is incident. When the incident sound field is composed of several acoustical modes, the phase interferences of incident modes are important for the penalty mode scattering. The effects of other parameters, e.g., liner length, mode quasiresonance on the penalty mode scattering are also presented. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434242]

PACS number(s): 43.20.Fn, 43.20.Hq, 43.20.Mv [LLT]

Pages: 1303–1312

I. INTRODUCTION

Using mode scattering to increase the noise attenuation in lined ducts has been studied extensively. Selected references are, e.g., Refs. 1–5. On the other hand, the penalty effects of mode scattering on noise attenuation have received little attention. In this paper, we study the penalty mode scattering effects for the case of lining nonuniformity in the form of rigid splices. Typically, in lined duct, the liner is usually manufactured in sections which each cover part of the duct's circumference. The sections are joined together by longitudinal splices. The splices are acoustically hard. Sarin and Rademaker⁶ pointed out—by measuring the acoustic modes in the inlet of Fokker 100 aircraft—that the sound pressure level is modulated clearly by the rigid splices (hard-walled stripes) at 70% engine speed (approach). The azimuthal mode m' at 1 blade passage frequency (BPF) is scattered to lower azimuthal modes m . Rademaker *et al.*⁷ measured, by using a mode detection array, the scattered sound field in a model turbofan intake duct. They pointed out that the levels of scattered modes are significantly higher than the one produced by rotor-alone tone when the fan tip speed was supersonic and less than 80% of the maximum design rpm. These scattering effects cause a significant reduction in liner performance. Rademaker *et al.*⁸ then tried to quantify the effects of splices by experiments. They investigated 12 kinds of configurations including one to four splices and no splices, respectively. The incident azimuthal modes are $m'=0, 1, 2, 3$, respectively. They concluded that the m -mode spectra of the transmitted fields are clearly modulated by the rigid splices, while this scattering appears to have only a small influence on the total in-duct transmitted acoustic energy flux.

Regan *et al.*⁹ used the FEM method to study this problem. One calculation example for $K=20$, lining impedance $Z/\rho c=2-j$, and splice angle 0.15 rad, is presented, where K

is the dimensionless wave number, $K=kR$, $k=2\pi f/c$, f the frequency, ρ and c the density and sound velocity in air, and R the radius of ducts. The m -mode spectrum is clearly modulated by the rigid splices for the incident mode (17,0) which is the last cut-on mode for this frequency. McAlpine *et al.*^{10,11} used finite/boundary element method to try to assess the effects of the rigid splices for the supersonic rotor alone tones source in turbofan aeroengines. Because of the limitation of the finite element method (a realistic limit for the spliced-liner simulations is about $K < 25$)¹⁰ a “model problem” with reduced number of fan blades and reduced BPF was designed to simulate the full-scale problem. The m -mode spectra which are modulated by the rigid splices are shown clearly. The conclusions of the “model problem” are that the effects of splices will be significant when a high cut-off ratio mode is incident or a high circumferential order mode is incident. The FEM may be impractical to evaluate systematically the above-mentioned mode scattering effects, due to the significant time required to solve this three-dimensional (3D) problem at realistic frequencies. Furthermore, when the splice angles tend to be small, the nodes per wavelength used in FEM have to increase. This will increase significantly the calculation memory and time.

Tester *et al.*¹² validated a perturbation method to evaluate the effects of small splices. This method agrees with a commercially available finite element code, at small splice angles. Wright¹³ developed a hybrid analytical/numerical method to study this problem.

A relevant problem to study the above-mentioned mode scattering by nonuniform impedance in lined ducts is that of sound propagation in a hard-walled duct of circular cross section, fitted with a region of nonuniform liner. Given the complex modal input amplitudes at one end of the hard-walled duct, the problem is to compute the complex modal output amplitudes at the other end. The computational difficulties appear for relatively high dimensionless wave number K , ($K > 30$) and completely 3D configuration. Significant

^{a)}Electronic mail: wenping.bi@univ-lemans.fr

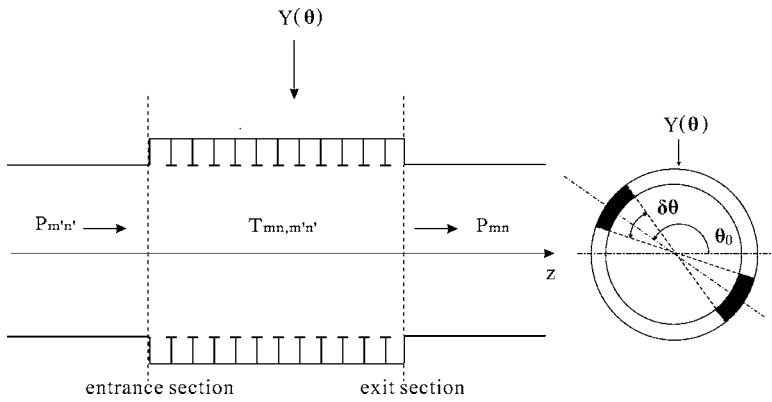


FIG. 1. Mode scattered by lining impedance.

CPU time and memory are required. Systematic assessments of the effects of mode scattering by splices are then difficult and time consuming.

In this paper, the multimodal propagation method (MPM), developed by the authors¹⁴ is used, to study the penalty mode scattering in an infinite cylindrical rigid duct lined with one axial segment liner without flow, as shown in Fig. 1. The penalty mode scattering means, in this paper, that incident modes are scattered into other modes by the rigid splices to reduce the performance of liner. The lining impedance is axially uniform, with two axial, diametrically opposed, rigid splices. The variations of the penalty mode scattering with splice angles, lining lengths, and incident mode content are focused on. The penalty effects of mode scattering on the liner mode quiresonance are also studied.

The plan of this paper is as follows. In Sec. II, we outline the MPM which is used in this paper. Then, the penalty effects of mode scattering for one mode incidence are shown in Sec. III. In Sec. IV, we present the penalty effects of mode scattering for multimode incidence. Finally, in Sec. V, we summarize the calculation results.

II. MPM AND MODE SCATTERING

In Ref. 14, we developed a MPM to study sound propagation in nonuniform lined ducts. It is an application of the multimodal method, which was first proposed by Pagneux *et al.*¹⁵ to study sound propagation in rigid waveguide with varying cross section, in the study of sound propagation in lined ducts with lining impedance being variable in both axial and circumferential directions.

The points of the multimodal method, first proposed in Ref. 15, are as follows: The two first-order differential equations, the equation of mass conservation combined with the equation of state and the equation of momentum conservation, are decomposed in the transverse direction over an orthogonal and complete set of functions (rigid duct modes in Ref. 15) which are known *a priori*; by defining a modal impedance matrix the decomposed coefficients of the two above-mentioned equations are then reformulated as a Riccati equation in the axial direction, which is stable to solve numerically. The process of transverse decomposition and axial reformulation was then applied to solve the wave propagation in rigid bends.^{16,17} This process was also applied in uniform ducts lining nonuniform impedance,¹⁴ uniform bends lining uniform impedance,¹⁸ and varying cross-section

ducts lining nonuniform impedance,¹⁹ and was extended to include mean flow in ducts.²⁰ The objective of Refs. 14 and 19 is to uncover the ability of the method to deal with the nonuniform impedance boundary conditions in relatively high frequency ($K=30-50$), which is important in intakes of aeroengine and may be difficult by pure numerical method, e.g., FEM. In Ref. 21, this method was improved by introducing another matrix variable, with which the decomposition coefficients can be solved by marching algorithm in axial direction. This significantly improves the practical implementation of this method for the problem needed large transverse decomposition (e.g., in aeroengine intakes¹⁹). This method was also improved by efficient transverse decomposition.^{22,23} By this improvement, the noise propagation in nonuniform lined intakes of aeroengine can be calculated up to $K=80$. This may be an important improvement for the industry needs.

As compared to FEM codes which involve spatial discretization in both the transverse and longitudinal directions, the MPM involves discretization only in the transverse directions. This is a significant economy which enables one to apply the method at quite high reduced wave numbers. Coefficients of reflection and transmission are calculated directly for any types of sources.

The MPM was validated in two dimension by comparison with FEM.¹⁴ It was validated in the 3D case with multimodal incidence by comparison with experiment.²⁴ In the 3D case, when one mode is incident, we will show in this paper that the MPM results are the same as Tester's results.¹²

In Ref. 14, the whole duct is decomposed into three contiguous regions with respect to axial coordinate z : left and right semi-infinite uniform rigid ducts connected by an arbitrary lined region (see Fig. 1). The sound pressure p and axial velocity v_z are decomposed using the rigid duct modes

$$p(r, \theta, z) = \sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} P_{mn}(z) \Psi_{mn}(r, \theta), \quad (1)$$

$$v_z(r, \theta, z) = \sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} V_{mn}(z) \Psi_{mn}(r, \theta), \quad (2)$$

where P_{mn} and V_{mn} are the decomposed coefficients (amplitudes). The basis functions Ψ_{mn}

$$\Psi_{mn} = \frac{1}{\sqrt{\pi\Lambda_{mn}}} e^{-jm\theta} \frac{J_m(\alpha_{mn}r)}{J_m(\alpha_{mn})}, \quad (3)$$

are the eigenfunctions of the hard-walled cylindrical duct, where Λ_{mn} are normalized coefficients of eigenfunctions Ψ_{mn} , J_m are m order Bessel functions, and α_{mn} are eigenvalues of rigid modes. They are known *a priori* and correspond to the exact, physical decomposition in the rigid regions, while providing only a coupled, mathematical representation of the total wave field in the lined region.

Because the duct is uniform, the decomposed coefficient \mathbf{P} obeys a second-order differential equation¹⁴

$$\mathbf{P}'' + \mathbf{A}\mathbf{P} = 0, \quad (4)$$

where \mathbf{P} is a vector, its elements are P_{mn} , the double prime refers to the second derivative with respect to z , the matrix \mathbf{A}

is constant along z when the lining impedance is unchanged with z , and expressed as

$$\mathbf{A} = (K^2\mathbf{I} - \mathbf{L}^2) + \mathbf{C}, \quad (5)$$

where \mathbf{I} is an identity matrix and \mathbf{L} is a diagonal matrix with eigenvalues of rigid modes on the diagonal. Matrix \mathbf{C} describes clearly the mode scattering excited by the circumferentially nonuniform impedance. It is shown that the mode scattering excited by circumferentially nonuniform impedance is decided by the Fourier transformation of the lining admittance function with respect to θ . If the nonuniformity of lining admittance is due to rigid splices, and the rigid splices are uniformly distributed in circumference, the lining admittance function may be expressed as

$$\beta(\theta) = \begin{cases} 0 & \text{when } \frac{2p\pi}{N} + \theta_0 - \frac{\delta\theta}{2} < \theta < \frac{2p\pi}{N} + \theta_0 + \frac{\delta\theta}{2}; p = 0, \dots, N-1 \\ \beta_0 & \text{otherwise,} \end{cases} \quad (6)$$

where $\delta\theta$ is the angle of splices, θ_0 is an angle related to the angular positioning of the overall configuration, as indicated in Fig. 1, and N refers to the number of splices. The elements of matrix \mathbf{C} may then be written in analytical form as

$$C_{mn,m'n'} = \frac{1}{\pi\sqrt{\Lambda_{mn}\Lambda_{m'n'}}} \int_0^{2\pi} Y(\theta) e^{-j(m'-m)\theta} d\theta = \begin{cases} -\frac{2Y_0}{\sqrt{\Lambda_{mn}\Lambda_{m'n'}}} \frac{N\delta\theta}{2\pi} \text{sinc}\left[(m'-m)\frac{\delta\theta}{2}\right] e^{-j(m'-m)\theta_0} & \text{when } |m-m'| = C \times N \\ \frac{2Y_0}{\sqrt{\Lambda_{m'n'}\Lambda_{m'n'}}} \left(1 - \frac{N\delta\theta}{2\pi}\right) & \text{when } m = m' \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where $Y(\theta) = -jK\beta(\theta)$. Equation (7) means that only the modes which satisfy the formula $|m-m'| = C \times N$, where C is an arbitrary integer, and $m=m'$ can be excited by the splices. This formula was also obtained in analogy with a similar expression of the rotor-stator interaction.^{9,10}

Equation (4) is a matrix constant coefficient ordinary differential equation. Its solutions can be written explicitly. By using the continuity of pressure and velocity at the interfaces between lined and rigid parts, the coefficients of transmission and reflection can be obtained directly, by some matrix calculations. (For details, see Ref. 14).

Therefore, the mode scattering phenomena are well captured by the MPM in nonuniformly lined ducts. The transmission coefficients obtained by MPM are defined by

$$P_{mn} = \sum_{m'n'} T_{mn,m'n'} P_{m'n'}, \quad (8)$$

where P_{mn} and $P_{m'n'}$ refer to the modal coefficients of sound pressure in the rigid duct, at the exit and entrance sections (see Fig. 1).

A typical example of mode scattering in lined duct with rigid splices is shown in Fig. 2. The calculation frequency is $K=20$, the lining impedance is $Z/\rho c = 1-j$, and two acoustically rigid splices with angles of 0.15 rad are placed in opposite directions. The modulus of the above-mentioned pressure transmission coefficients $T_{mn,m'n'}$ are plotted. The indices m', n' , correspond to the incident modes, while m, n , correspond to the output modes. In Fig. 2, we present only the first five radial modes ($n(n')=0-4$) for every $m(m')$. The clusters in the diagonal correspond to output modes (m, n) having $m=m'$ ($n=n'$ or $n \neq n'$). These clusters would be present (with slightly modified amplitudes) in the absence of the splices, and correspond to the scattering associated with the axial nonuniformity. On the contrary, for the transmission coefficients of modes excited by circumferential nonuniformity, we have $m \neq m'$ and their values are reported significantly off the diagonal, out of the mentioned clusters. It is immediately clear from Fig. 2 that the corresponding amplitudes are small compared to those in the clusters. The amplitudes of the modes excited by the circumferential nonuniformity

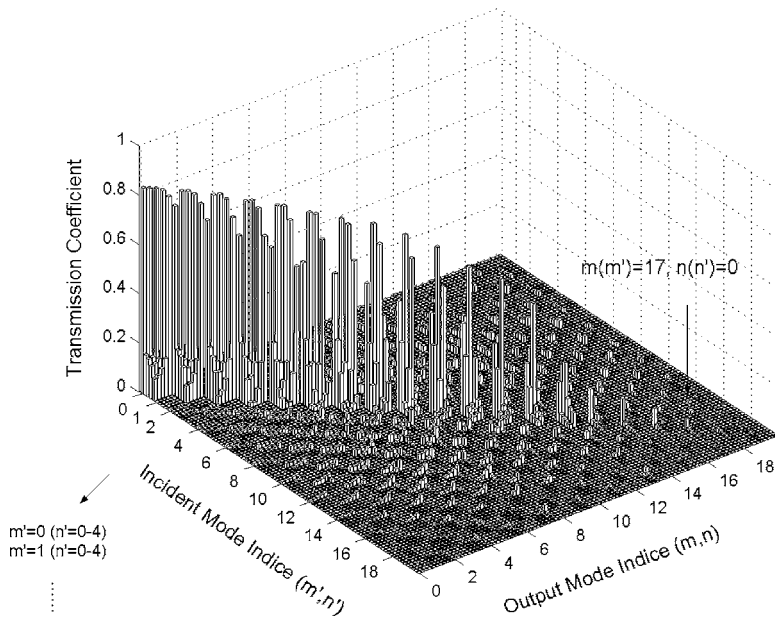


FIG. 2. A typical figure of mode transmission coefficients, $K=20$, $Z/\rho c=1-j$, two splices with angles of 0.15 rad, $L/R=0.5882$. Indices (m, n) and (m', n') are arranged along the two axis in the following order: $(0,0), (0,1), (0,2), (0,3), (0,4), (1,0), (1,1), (1,2), (1,3), (1,4), \dots, (18,0), (18,1), (18,2), (18,3), (18,4)$.

mity are much less than those of the modes excited by the axial nonuniformity. One exception is for the incident mode $(17,0)$, which is near cut-off. Its self-transmission coefficient $T_{m=17, n=0, m'=17, n'=0}$ is very small. The diagonal cluster tends to disappear and the modes excited by the circumferential nonuniformity, which have relatively larger transmission coefficients $T_{mn, m'=17, n'=0} (m \neq 17)$, dominate the output sound field. The m orders of those modes are less than $m'=17$. This is one case in which there exists significant penalty scattering due to the splices. This penalty effect can be seen if the mode $(17,0)$ is strongly excited by the source.

III. PENALTY MODE SCATTERING BY RIGID SPLICES WHEN ONE MODE IS INCIDENT

In this section, we consider the penalty mode scattering when one mode is incident. We will study the influence of splice angle and lining length on the penalty mode scattering. The penalty effects of mode scattering on the liner mode quasiresonance are also studied.

The penalty effects are shown by the reduction of transmission loss (TL). It is defined as

$$TL(\text{dB}) = -10 \log_{10} \frac{W_o}{W_i}, \quad (9)$$

where W_o and W_i are the output and incident sound power at the exit and entrance sections, respectively (see Fig. 1). The sound power is defined as

$$W = \frac{1}{2} \text{Re} \left(\int_S p(r, \theta, z) v_z^*(r, \theta, z) dS \right), \quad (10)$$

where S is the surface of duct cross section and the asterisk indicates the complex conjugate. By using Eqs. (10), (1), and (2), we obtain the total sound power

$$W = \frac{1}{2} \text{Re}(\mathbf{P}^\dagger \mathbf{V}), \quad (11)$$

where the dagger refers to the adjoint operator, \mathbf{P} and \mathbf{V} are vectors, their elements are P_{mn} and V_{mn} defined in Eqs. (1) and (2).

A. The influence of splice angle

As shown in Refs. 6, 7, 10, and 11, when one mode is incident in lined ducts with rigid splices, the transmission loss may decrease, compared with that in lined ducts without splices. This decrease may be generated by penalty mode scattering and reduction of lining surface because of rigid splices. Two cases are important: one is that the incident mode is near cut-off, the other is that the incident modes have high azimuthal orders.

1. The incident mode is the last cut-on mode and is near cut-off

When the incident mode is the last cut-on mode and is near cut-off, the sound power in the duct without splices will be well attenuated. If there exist rigid splices, this last cut-on mode will be scattered to the liner modes of lower m orders which are difficult to attenuate. The penalty effects of splices will be significant. An example is for $K=20$, the incident mode $(17,0)$ (the last cut-on mode) determines a field which is rapidly attenuated in the lining part. The transmission loss is shown in Fig. 3 versus the splice angle. For well understanding the behavior of the penalty mode scattering, a reference curve, $C - 10 \log_{10} \delta\theta$, is also plotted in the Fig. 3, where C is a constant. It is shown that there exists a critical angle, $\delta\theta \approx 0.2$ rad in this example. At small splice angles, less than the critical angle, which is in the range of realistic values, the sound transmission loss decreases sharply with increasing of the splice angle. The decreasing rate is greater than the reference curve $-10 \log_{10} \delta\theta$. The performance of the liner reduces significantly. When the splice angles increase, greater than the critical angle, the transmission loss decreases slowly, approaching $-10 \log_{10} \delta\theta$.

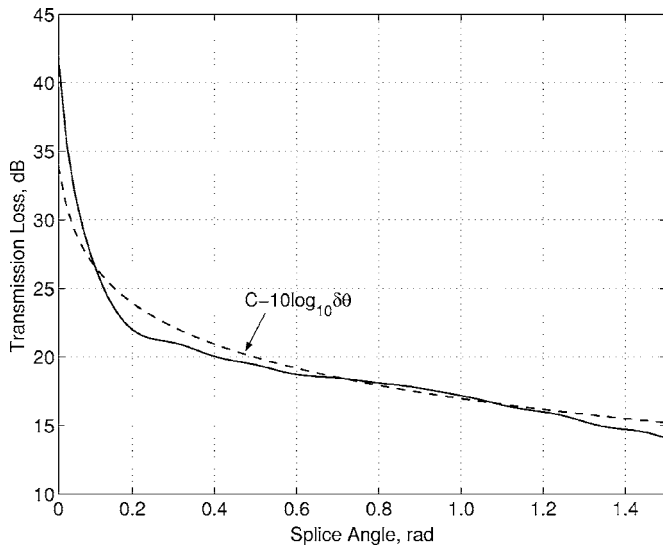


FIG. 3. Effects of splices vs $\delta\theta$ for incident mode (17,0), the last cut-on mode, and near cut-off. $K=20$, two opposite splices of angle $\delta\theta$, lining impedance $Z/\rho c=1-j$, and lining length $L/R=1$.

2. The high azimuthal order mode is incident

We have shown in Fig. 2 that the effects of rigid splices on sound power transmission are important only when the sound power of the incident mode and the modes scattered by axial nonuniformity are well attenuated. The most important example of this is that the incident mode is the last cut-on mode. Another important example¹² which takes place often in turbofan intakes is: The m' order of the incident mode is relatively high, e.g., $m'=26$ at BPF. To estimate the penalty mode scattering, the variation of transmission loss versus splice angle is shown in Fig. 4. For comparison, we plot also the variation of transmission loss with three splices; the variation of transmission loss when mode (20,0) is incident. In Fig. 4, the reference curves $C_{1,2}-10\log_{10}\delta\theta$ and $C_3-20\delta\theta/\pi$ are also plotted, where C_1 , C_2 , and C_3 are constants, respectively. It is shown that the transmission loss approaching $-10\log_{10}\delta\theta$ is still observed with increasing splice angle, while being initially less rapid.

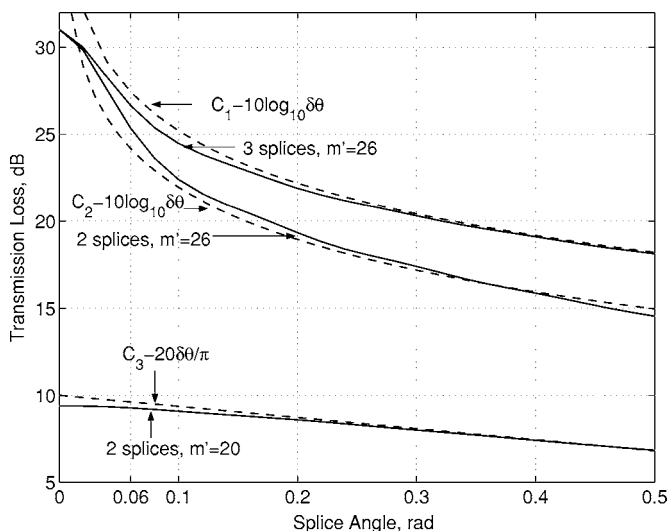


FIG. 4. Effects of splices vs $\delta\theta$. $K=31.26$, two splices, lining impedance is $Z/\rho c=2-j$, and lining length $L/R=0.48$.

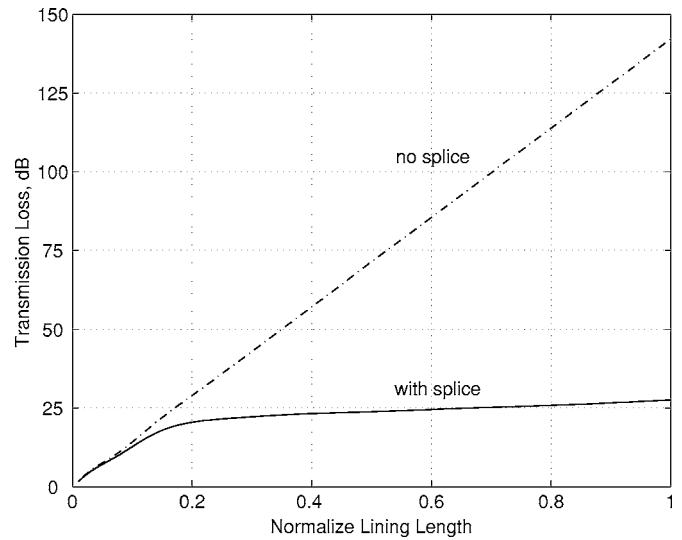


FIG. 5. Effects of the splices vs lining length L/R . Mode (17,0) is incident, $K=20$, two opposite splices of angles $\delta\theta=0.15$, lining impedance $Z/\rho c=1-j$.

On the other hand, when mode (20,0) is incident, the transmission loss keeps nearly unchanged at small splice angles, but decreases linearly with the increasing of splice angle.

At $\delta\theta=0.06$, it is shown that there is about 5 dB less attenuation with two splices than without splices when mode (26, 0) is incident, which is the same result as that obtained by Tester *et al.*¹²

B. The influence of lining length

To see the effects of the splices in another manner, we fix $\delta\theta=0.15$ rad and plot in Fig. 5 the sound transmission loss versus the lining length. We compare the value with that with splice suppressed. It is clearly shown that when $L/R > 0.2$ the unfavorable scattering, i.e., the penalty of splices, is significant. It may be important to point out that although in this case the penalty effects of splices are significant, the total transmission loss is already about 25 dB. Therefore this penalty mode scattering is to be considered only when the levels of the source are very high.

The influence of lining length on the penalty mode scattering can be seen more clearly by the sound pressure amplitude distributions over different cross section as shown in Fig. 6. The parameters are the same as in Fig. 5, two splices are at 0 and π , except that we keep the lining length $L/R=100$ unchanged. It is found that at the beginning of liner, $z/R=0.01$ and $z/R=0.1$, the splices only perturb the sound pressure near the splices, the sound pressure amplitude distributions are similar to that without splices. When the distance increases from the beginning of liner, at $z/R=0.5$, $z/R=5$, $z/R=10$, and $z/R=20$, the sound pressures are decided by all the modes scattered by the splices. Finally, when the length of liner is sufficiently long, at $z/R=30$, $z/R=40$, and $z/R=90$, the higher liner modes are well attenuated. The mode which has the lowest attenuation decides the sound pressure distribution.

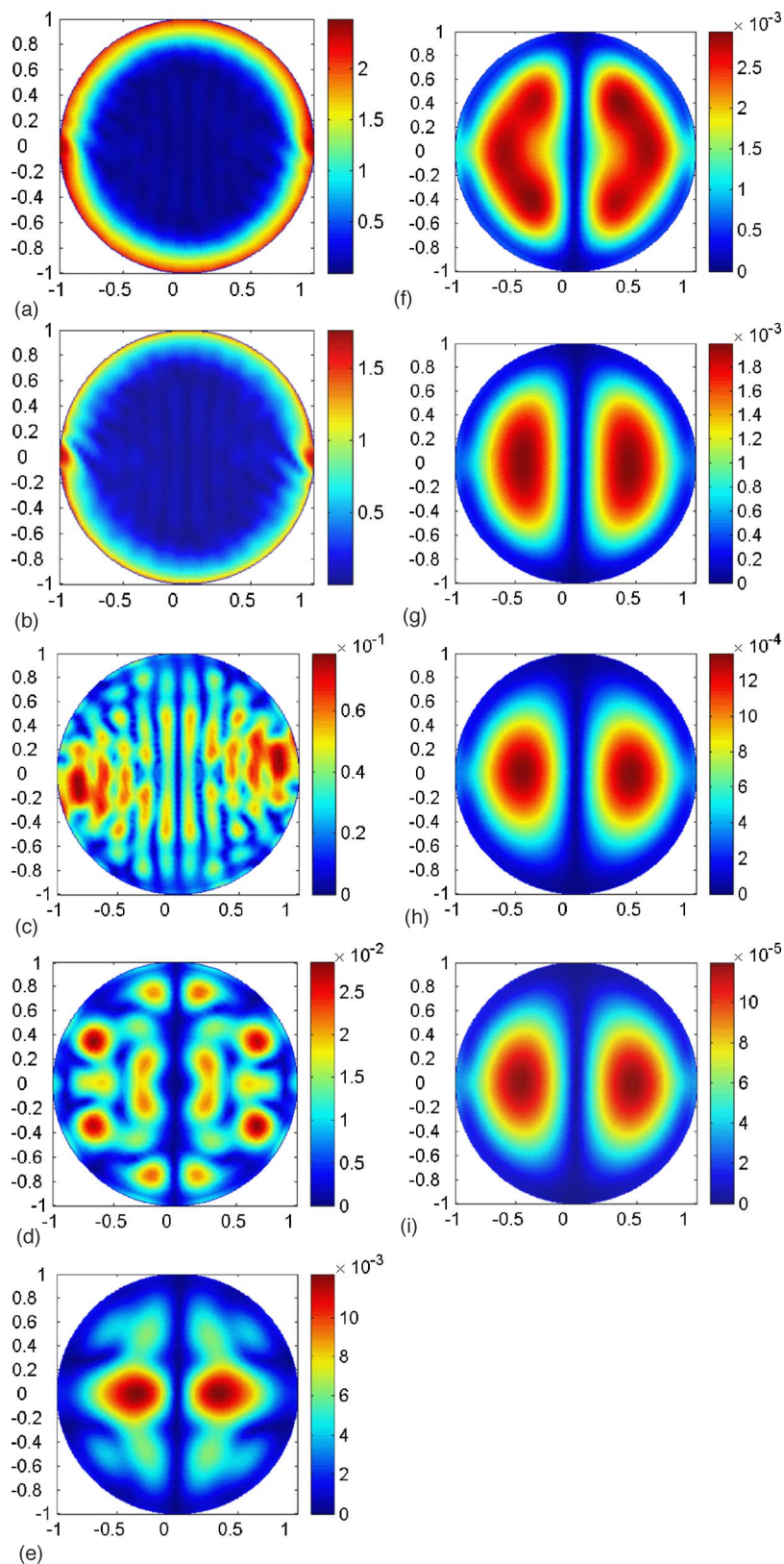


FIG. 6. (Color online) Sound pressure amplitude distributions over the duct cross section, for (a) $z/R=0.01$, (b) $z/R=0.1$, (c) $z/R=0.5$, (d) $z/R=5$, (e) $z/R=10$, (f) $z/R=20$, (g) $z/R=30$, (h) $z/R=40$, and (i) $z/R=90$, keeping the lining length $L/R=100$. The parameters are: $K=20$, $Z/\rho c=1-j$, two splices with angle 0.15 rad, respectively, incident mode (17,0).

C. Quasiresonance of the *liner* modes excited by axial nonuniformity

We now pass to a significantly different case, in which the penalty effects of the splices can be seen. In some special region of the reduced frequency K , resonances can take place

and produce attenuation peaks in the absence of splice. Those peaks may be strongly reduced by the presence of splices.

An example is shown in Fig. 7 for incident mode (9,0). The impedance for this case is $Z/\rho c=1+j[0.01K/R-\cot(0.03K/R)]$, where ρc is the wave impedance of air. This

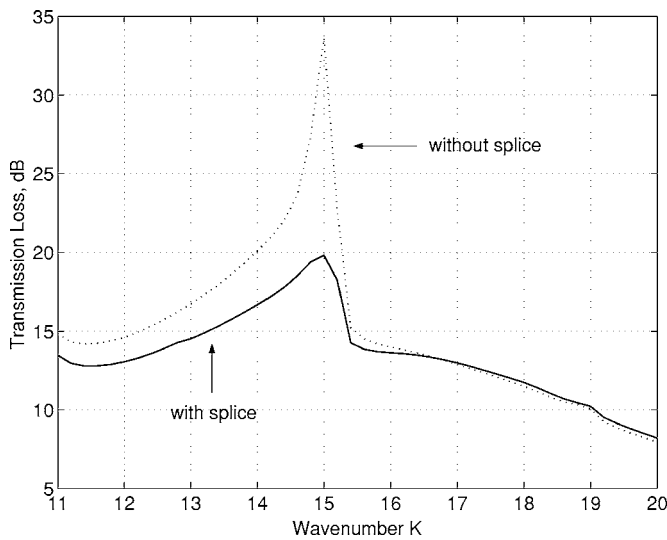


FIG. 7. The penalty effects of rigid splices on the quaresonance. Mode (9,0) incident, $L/R=1$, $\delta\theta=0.15$. $Z/\rho c=1+j[0.01K/R-\text{ctg}(0.03K/R)]$, where ρc is the wave impedance of air.

lining impedance corresponds to the liners made with honeycombs covered by face sheet resistance material. The face sheet resistance is 1, the mass reactance is $0.01K/R$, and $-\text{cot}(0.03K/R)$ corresponds to the cavity resonance. The splice angle is $\delta\theta=0.15$ rad, and $L/R=1$, $R=0.85$. The sharp peak of attenuation, at about $K=15$, without splices is significantly reduced by the presence of splices. The penalty effects of splices are significant in this frequency region.

At $K=15$ without splices, two liner modes (9,0) and (9,1) are cut-on in the lining part as shown in Fig. 8. (“The liner modes are cut-on” means the imaginary parts of axial wave numbers turn to be significantly small.) Mode (9,1) is excited by the axial nonuniformity, its real part of axial propagating constant K_z is 3.37, which satisfies approximately the resonance condition $K_z L \approx \pi$, where L is the dimensionless length of liner. This resonance is not a true resonance of a trapped mode,²⁵ it is called quaresonance. This quaresonance determines a significant peak on the trans-

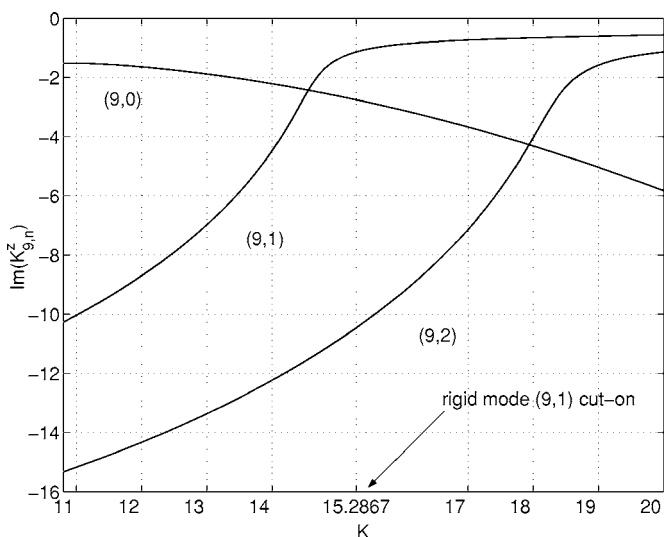


FIG. 8. Axial wave number of liner modes (9,0), (9,1), and (9,2), in absence of splice.

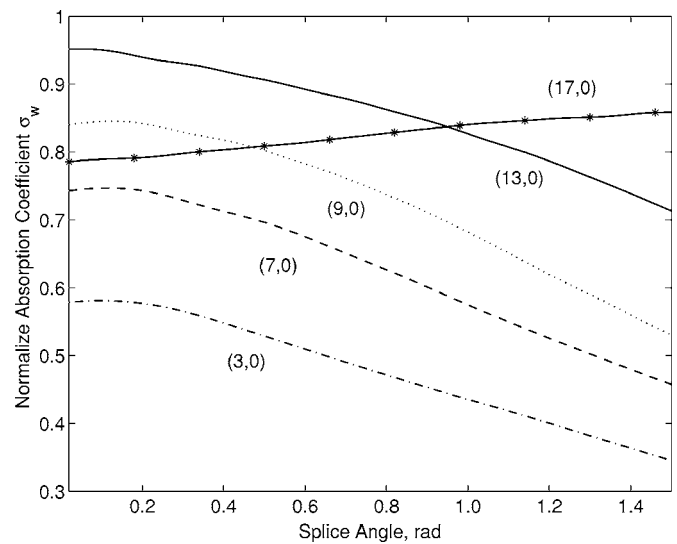


FIG. 9. The coefficient of sound absorption σ_w vs $\delta\theta$ for $K=20$, $L/R=1$, and $Z/\rho c=1-j$.

mission loss curve. When the splices are introduced in the lining part, the quaresonance is partially destroyed because the liner modes excited by splices do not satisfy the resonance condition. The transmission loss decreases, as compared to the situation without splices.

D. The effects of mode scattering on sound absorption

The sound absorption may be sometimes interesting to consider. We define the coefficient of sound absorption as

$$\sigma_w = (W_i - W_r - W_o)/W_i, \quad (12)$$

where W_i , W_o , and W_r are the incident, output, and reflected sound power, respectively. The coefficients of sound absorption are shown in Fig. 9 for the different incident modes (3,0), (7,0), (9,0), (13,0), and (17,0) (parameters given in Fig. 9). It shows that the coefficients of sound absorption decrease globally with the angle of splices for incident modes (3,0), (7,0), (9,0), and (13,0). But in the region of $\delta\theta < 0.2$ rad the coefficients of sound absorption are nearly stationary. This range of angle corresponds to the practical angles of splices in turbofan inlet liners. It may be concluded that the splices have no penalty effects on the coefficient of sound absorption for practical splice angles, in the limit of the cases we have studied. It is interesting to point out that for the incident mode (17,0) the coefficient of sound absorption increases in all the range of $\delta\theta$, while the transmitted sound power decreases (see Fig. 3). In fact for this mode the reflected sound power is much greater than the output sound power (all modes near cut-off have relatively high reflection coefficients). The increasing of the angle of splices reduces the reflected sound power and more sound energy enters the duct and is absorbed. There is no penalty scattering effect when this mode is incident if we consider only the coefficients of sound absorption. However it is the transmission coefficient which is meaningful to characterize the penalty effects of rigid splices.

IV. PENALTY MODE SCATTERING BY RIGID SPLICES WHEN MULTIMODES ARE INCIDENT

It may be interesting to assess the effects of rigid spllices on the liner performance when multimodes are incident. When multimodes are incident, the transmission loss TL is rewritten as¹⁸

$$\begin{aligned} \text{TL(dB)} &= -10 \log_{10} \frac{W_o}{W_i} \\ &= -10 \log_{10} \frac{\text{Re}(\sum_{\alpha\beta} (\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\beta} P_\alpha^{i*} P_\beta^i)}{\text{Re}(\sum_{\alpha} Y_{\alpha\alpha}^c P_\alpha^{i*} P_\alpha^i)}, \end{aligned} \quad (13)$$

where the superscript “i” refers to incidence, P_α^i, P_β^i refer to incident modal coefficients as indicated in Eq. (8), here we use one index α or β to replace indices ($m'n'$), \mathbf{T} is the transmitted matrix, its elements are transmission coefficients $T_{mn,m'n'}$ as defined in Eq. (8), the dagger refers to the adjoint operator (conjugate transpose), \mathbf{Y}^c is a diagonal matrix, its elements are axial dimensionless wave numbers of rigid modes K_α^z/K in the diagonal $Y_{\alpha\alpha}^c$. In Eq. (13), we have used the Eq. (8), and $\mathbf{V}^i = \mathbf{Y}^c \mathbf{P}^i$, where \mathbf{V} and \mathbf{P} are vectors, their elements V_{mn} (or V_α) and P_{mn} (or P_α) are defined in Eqs. (1) and (2).

In general, the detailed knowledge of the sources is not known, they are often assumed that all propagating modes are incident, and the total incident sound power is distributed uniformly among those propagating modes. It means that

$$|P_\alpha^i| = \sqrt{\frac{2W_i}{N_{\text{mode}} Y_{\alpha\alpha}^c}}, \quad (14)$$

where W_i is the total incident sound power and N_{mode} is the number of propagating modes in rigid ducts. The incident modes are then expressed as

$$P_\alpha^i = \sqrt{\frac{2W_i}{N_{\text{mode}} Y_{\alpha\alpha}^c}} e^{-j(K_\alpha^z z + f_\alpha)}, \quad (15)$$

where f_α is the initial phase of incident mode α . The transmission loss Eq. (13) is then rewritten as

$$\text{TL(dB)} = -10 \log_{10} \text{Re} \left(\frac{1}{N_{\text{mode}}} \sum_{\alpha\beta} \frac{(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\beta}}{\sqrt{Y_{\alpha\alpha}^c Y_{\beta\beta}^c}} e^{j(f_\alpha - f_\beta)} \right), \quad (16)$$

where we have assumed $z=0$ for incident modes, and it does not change the generality of the problem.

In absence of detailed knowledge of the distribution of initial phases, we first consider the two following simple cases. A more general case is shown in Sec. IV C.

A. Phases of incident modes are random and incoherent

When the phases of incident modes are random and incoherent, it means

$$\langle e^{j(f_\alpha - f_\beta)} \rangle = \delta_{\alpha\beta}, \quad (17)$$

where $\langle \rangle$ refers to the ensemble average over realizations, δ Kronecker delta.

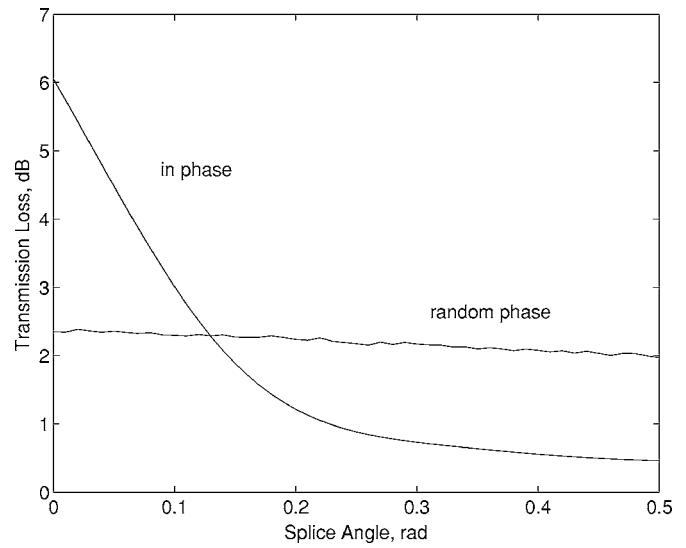


FIG. 10. Effects of spllices vs $\delta\theta$ when all propagating modes are incident. $K=31.26$, two opposite spllices of angle $\delta\theta$, lining impedance $Z/\rho c=2-j$, and lining length $L/R=0.48$.

The transmission loss is

$$\begin{aligned} \text{TL(dB)} &= -10 \log_{10} \frac{\langle W_o \rangle}{\langle W_i \rangle} \\ &= -10 \log_{10} \text{Re} \left(\frac{1}{N_{\text{mode}}} \sum_{\alpha\alpha} \frac{(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\alpha}}{Y_{\alpha\alpha}^c} \right). \end{aligned} \quad (18)$$

This is equivalent to saying that every propagating mode is incident one by one. The phase interferences of incident modes are not included.

B. Phases of incident modes are coherent and in phase

The other simple case appears when the phases of incident modes are coherent and in phase, the transmission loss is rewritten as

$$\text{TL(dB)} = -10 \log_{10} \text{Re} \left(\frac{1}{N_{\text{mode}}} \sum_{\alpha\beta} \frac{(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\beta}}{\sqrt{Y_{\alpha\alpha}^c Y_{\beta\beta}^c}} \right). \quad (19)$$

In this case, the phase interferences of incident modes are included explicitly.

Figure 10 shows the effects of spllices versus $\delta\theta$ when all propagating modes are incident. All incident modes have the same sound power, i.e., equipartition. The phases of incident modes are assumed as shown in Fig. 10: All incident propagating modes are in phase for one curve, and random incoherence for the other curve. Parameters are presented in Fig. 10. When the incident modes are in phase, the effect of spllices is obvious. The penalty effects of spllices significantly depend on the splice angle, being especially important at small splice angles. On the other hand, for incident modes with random incoherent phases, the effect of spllices is negligible. This means the phase interferences play a leading role in the penalty mode scattering. This will be shown more clearly in the following section.

In Fig. 10, the curve of random phase has some fluctuations. Those are from the ensemble average of finite realizations. Equation (18) corresponds to the mean value of this curve.

C. Phase difference between modes $\chi=f_\alpha-f_\beta$ fulfills a Gauss distribution

In practice, the phases of incident modes may be neither in phase, nor incoherent, they may fulfill some distribution. The penalty mode scattering effects will depend on this distribution. In this section, we assume the phase difference between modes $\chi=f_\alpha-f_\beta$ fulfills a Gauss distribution, the transmission loss, after an ensemble average, is

$$\begin{aligned} \text{TL(dB)} &= -10 \log_{10} \text{Re} \left(\frac{1}{N_{\text{mode}}} \sum_{\alpha\beta} \frac{(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\beta}}{\sqrt{Y_{\alpha\alpha}^c Y_{\beta\beta}^c}} \frac{1}{\sqrt{2\pi}\Delta} \right. \\ &\quad \left. \times \int_{-\infty}^{\infty} d\chi \exp \left(j\chi - \frac{\chi^2}{2\Delta^2} \right) \right) \\ &= -10 \log_{10} \frac{1}{N_{\text{mode}}} \left[\text{Re} \left(\sum_{\alpha\alpha} \frac{(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\alpha}}{Y_{\alpha\alpha}^c} \right) \right. \\ &\quad \left. + \text{Re} \left(\sum_{\alpha\beta, \alpha \neq \beta} \frac{(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\beta}}{\sqrt{Y_{\alpha\alpha}^c Y_{\beta\beta}^c}} e^{-\Delta^2/2} \right) \right], \quad (20) \end{aligned}$$

where Δ is standard deviation and we have assumed the mean value is zero. Equation (20) shows clearly that the sound transmission in lined ducts with arbitrary axial and circumferential nonuniformity is the sum of two terms. The first term $\sum_{\alpha\alpha} (\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\alpha} / Y_{\alpha\alpha}^c$ describes the self-scattering (transmission) of modes in the presence of axial and circumferential nonuniformity. It depends only on the liner properties. The second term $\sum_{\alpha\beta, \alpha \neq \beta} [(\mathbf{T}^\dagger \mathbf{Y}^c \mathbf{T})_{\alpha\beta} / \sqrt{Y_{\alpha\alpha}^c Y_{\beta\beta}^c}] e^{-\Delta^2/2}$ describes the cross-scattering (transmission) of modes in the presence of axial and circumferential nonuniformity. It depends on not only the liner properties, but also the configurations of the sources. The effects of sources is from $e^{-\Delta^2/2}$, where the standard deviation Δ is between $0 \sim 2\pi$. When Δ equals to zero, the factor $e^{-\Delta^2/2}$ equals to 1, Eq. (19) is recovered, it corresponds to the extremal case when all the incident modes are in phase. When $\Delta > \pi$, $e^{-\Delta^2/2}$ tends to zero, Eq. (18) is recovered. In this case, the performance of liner may be described totally by the liner properties and is independent of the sources.

V. CONCLUSION

We have studied the penalty modes scattering excited by rigid splices in lined ducts without flow. The main conclusions are as follows.

When one mode is incident, which is near cut-off, there exists a critical angle for the splices such that below this critical angle the transmission loss decreases sharply with increasing splice angle, and above this critical angle the transmission loss decreases more slowly, like $-10 \log_{10} \delta\theta$ with increasing splice angle. When one mode is incident which is of high azimuthal order, the transmission loss approaching $-10 \log_{10} \delta\theta$ is still observed with increasing splice angle, while being initially less rapid.

When multimodes are incident, phase interferences of incident modes play a leading role in the penalty mode scattering. When all incident modes are in phase, the penalty effects of splices significantly depend on the splice angle, being especially important at small splice angles. On the other hand, when the phases of incident modes are random and incoherent, the effects of splices are negligible.

Another case in which the rigid splices may produce penalty mode scattering effects is that of a quasiresonance (in absence of splices) of the lined section.

ACKNOWLEDGMENT

The authors gratefully acknowledge support provided by a grant from SNECMA MOTEURS, France.

- ¹C. R. Fuller, "Propagation and radiation of sound from flanged circular ducts with circumferentially varying wall admittances. I. Semi-infinite ducts," *J. Sound Vib.* **93**, 321–340 (1984).
- ²C. R. Fuller, "Propagation and radiation of sound from flanged circular ducts with circumferentially varying wall admittances. II. Finite ducts with sources," *J. Sound Vib.* **93**, 341–351 (1984).
- ³R. Mani, "Acoustic duct with peripherally segmented acoustic treatment," United States Patent No. 3,937,590 (1976).
- ⁴M. S. Howe, "The attenuation of sound in a randomly lined duct," *J. Sound Vib.* **87**, 83–103 (1983).
- ⁵W. R. Watson, "Circumferentially segmented duct liners optimized for axi-symmetric and standing-wave sources," NASA report 2057 (1982).
- ⁶S. L. Sarin and E. R. Rademaker, "In flight acoustic mode measurements in the turbofan engine inlet of Fokker 100 aircraft," AIAA Pap. 93-4414, 1993.
- ⁷E. R. Rademaker, P. Sijtsma, and B. Tester, "Mode detection with an optimized array in a model turbofan engine intake at varying shaft speeds," AIAA Pap. 2001-2181, 2001.
- ⁸E. R. Rademaker, S. L. Sarin, and C. A. Parente, "Experimental investigation on the influence of liner non-uniformities on prevailing modes," AIAA Pap. 96-1682, 1996.
- ⁹B. Regan and J. Eaton, "Modeling the influence of acoustic liner non-uniformities on duct modes," *J. Sound Vib.* **219**, 859–879 (1999).
- ¹⁰A. McAlpine, M. C. M. Wright, H. Batard, and S. Thezelais, "Finite/ boundary element assessment of a turbofan splices intake liner at supersonic fan operating conditions," AIAA Pap. 2003-3305, 2003.
- ¹¹A. McAlpine and M. C. M. Wright, "Acoustic scattering by a spliced turbofan inlet liner at supersonic fan speeds," *J. Sound Vib.* **292**, 911–934 (2006).
- ¹²B. Tester, N. Baker, A. Kempton, and M. Wright, "Validation of an analytical model for scattering by intake liner splices," AIAA Pap. 2004-2906, 2004.
- ¹³M. C. M. Wright, "Hybrid analytical/numerical method for mode scattering in azimuthally nonuniform ducts," *J. Sound Vib.* **292**, 583–594 (2006).
- ¹⁴W. P. Bi, V. Pagneux, D. Lafarge, and Y. Aurégan, "Modeling of sound propagation in a non-uniform lined duct using a multi-modal propagation method," *J. Sound Vib.* **289**, 1091–1111 (2006).
- ¹⁵V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. i. Theory and validation," *J. Acoust. Soc. Am.* **100**, 2034–2048 (1996).
- ¹⁶S. Felix and V. Pagneux, "Sound propagation in rigid bends: A multimodal approach," *J. Acoust. Soc. Am.* **110**, 1329–1337 (2001).
- ¹⁷S. Felix and V. Pagneux, "Multi-modal analysis of acoustic propagation in three-dimensional bends," *Wave Motion* **36**, 157–168 (2002).
- ¹⁸S. Felix and V. Pagneux, "Sound attenuation in lined bends," *J. Acoust. Soc. Am.* **116**, 1921–1931 (2004).
- ¹⁹W. P. Bi, V. Pagneux, and D. Lafarge, "Sound propagation in varying cross section ducts lined with non-uniform impedance by multi-mode propagation method," AIAA Pap. 2005-3066, 2005.
- ²⁰Y. Aurégan, M. Leroux, and V. Pagneux, "Measurement of liner impedance with flow by an inverse method," AIAA Pap. 2004-2838, 2004.
- ²¹V. Pagneux and A. Maurel, "Lamb wave propagation in inhomogeneous elastic waveguide," *Proc. R. Soc. London, Ser. A* **458**, 1913–1930 (2002).
- ²²C. Hazard and V. Pagneux, "Improved multimodal approach in

waveguides with varying cross-section," *Proceedings of the International Congress on Acoustics*, Rome, 2001, Vol. 25, pp. 3–4.

²³W. P. Bi, V. Pagneux, D. Lafarge, and Y. Aurégan, "An improved multimodal method for sound propagation in non-uniform lined duct," *J. Acoust. Soc. Am.* (submitted).

²⁴W. P. Bi, "Sound propagation in nonuniform lined waveguides," Doctoral thesis, Université du Maine, Le mans, France, 2004.

²⁵V. Pagneux, "Sound propagation in varying cross section waveguides and the effects of flow," Doctoral thesis, Université du Maine, Le Mans, France, 1996.

Finite element analysis of guided waves in fluid-filled corrugated pipes

Matthias Maess,^{a)} Jan Herrmann, and Lothar Gaul

Institute of Applied and Experimental Mechanics, University of Stuttgart, Pfaffenwaldring 9, 70550 Stuttgart, Germany

(Received 24 April 2006; revised 28 December 2006; accepted 4 January 2007)

Free wave propagation in fluid-filled corrugated pipes is analyzed using finite element methods in combination with a wave-based approach. By combining discretized models with a wave-based approach, complex mechanism of wave motion in the three-dimensional waveguide is fully included. The pipes are treated as waveguides having periodic properties in the direction of wave propagation. The analysis of these guided waves leads to dispersion curves which show the strong frequency-dependency of the different wave modes. The method also allows the inclusion of coupling between fluid-borne and structure-borne wave modes which occur at the acoustic-structure interface. Phase and group velocities of the wave modes are derived in postprocessing steps. Additionally, the energy ratio of the fluid-domain and solid-domain vibrational energies is computed. Finally, linear damping models are included in order to explore wave mode attenuation. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436711]

PACS number(s): 43.20.Mv [DSB]

Pages: 1313–1323

I. INTRODUCTION

Amongst others, corrugated pipes are part of the engine fuel supply in automobiles and are located in the fuel supply module between the electric fuel pump and the fuel line connector, as illustrated in Fig. 1. The fuel module provides a sufficient fuel volume flow to the engine. However, periodic sound is generated by the pump and transmits sound power to car components through the corrugated pipe. As a result, undesired noise levels in the interior of the vehicle may occur. To obtain a complete understanding about sound transmission phenomena, the analysis of the acoustic field in the low kilohertz range is of practical relevance.

Harmonic wave propagation has been widely investigated in pipe engineering, since pumps and valves represent acoustic sources transmitting sound power into the pipe. Excessive noise levels may result, and may even lead to fatigue-induced deterioration of pipe components or attached structural components. From an acoustic point of view, pipes are longitudinal waveguides, and energy is transmitted in one direction over very large distances. This characteristic behavior is very important in the case of low-frequency noise in water-filled metal pipes. Here, the structure and the fluid provide very low viscosity or material damping, thus leading to almost nondecaying wave propagation. In order to avoid improper pipe design from an acoustic standpoint, harmonic wave propagation, and in particular sound speeds, need to be analyzed. The knowledge of sound speeds is also important for setting up transfer matrix methods for one-dimensional¹ or three-dimensional² pipe components. Wave propagation in pipes shows strong dispersive phenomena.³ Furthermore, several wave modes with frequency dependent wave numbers coexist. As a consequence, group and phase velocities also depend on frequency and wave number. The prediction

of wave dispersion in elastic pipes has been widely researched in the past decades.^{4–6} Fuller and Fahy⁷ describe an analytical model for an acoustic fluid embedded in a thin elastic cylindrical shell including bending stiffnesses. However, the use of analytical models is restricted to relatively simple geometries. To overcome this limitation, discretization methods may be employed. Special spectral or semianalytical finite elements have been developed to determine dispersion in uniform waveguides. These techniques have been applied for various solid waveguides such as plate structures, railways, sandwich panels,⁸ as well as for fluid-filled pipes including acoustic-structure interaction.^{9,10} However, special system matrices need to be developed as a function of the wave number.

Mace *et al.*^{11,12} have introduced an alternative method for analyzing waves in structural waveguides. This method combines the discretization capabilities of standard finite element (FE) code, such as that used for structural analysis, with a wave-based approach. This method can be applied to any waveguide with uniform properties in the direction of wave propagation. This technique has been complemented by Maess *et al.*^{13,14} for dispersion problems in uniform fluid-filled pipes including acoustic-structure interaction.^{15,16} The main advantage of this approach is that dispersion curves for two-field waveguides with arbitrary cross sections can be efficiently computed using a standard FE package.

Applications for periodic structural waveguides have been reported as well. To realize the analysis of these waveguides, it is necessary to perform a dynamic condensation of the interior degrees of freedom. In this paper, the method is applied to fluid-filled corrugated pipes with periodic properties in the direction of wave propagation in order to analyze guided wave properties.

^{a)}Electronic mail: maess@iam.uni-stuttgart.de

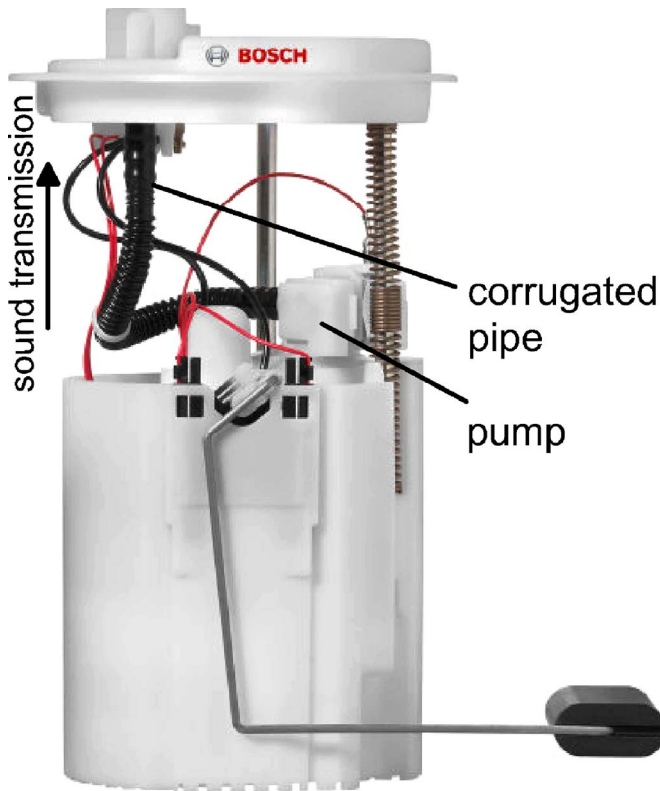


FIG. 1. (Color online) In-tank fuel supply module with pump and corrugated pipe (courtesy of Robert Bosch GmbH).

II. MODELING OF THE CORRUGATED PIPE

Sound propagation in fluid-filled pipes is characterized by a relatively strong interaction between structure-borne sound and fluid-borne sound at the acoustic–structure interface. Fully coupled models are therefore required for accurate prediction of the sound propagation. The acoustic model is restricted to the linear elastic domain for small wave amplitudes. Furthermore, the fluid is assumed to be at rest. In reality, a superimposed mean flow exists. However, the corresponding Mach numbers are less than 1% in the present fuel supply module, so that the influence of fluid flow on the acoustic field is negligible.¹⁷ It should also be mentioned that the corrugated pipe is not necessarily assembled in a straight fashion. However, this investigation is performed for a straight pipe configuration. Although the fuel supply module is located in the fuel tank (Fig. 1), the pipe is on top of the module and is therefore embedded in air for most fuel levels. The interaction with the light air medium is negligible due to the impedance mismatch, and for this reason, no exterior acoustic field is modeled. This assumption leads to an interior acoustic–structure problem.

A. Acoustic–structure interaction problem

In the present work, acoustic problems in interior domains Ω_f are considered. One possible choice for the field variable is the acoustic excess pressure p . The wave motion for a fluid at rest is governed by the hyperbolic linear wave equation¹⁸

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0. \quad (1)$$

Here, the free wave speed in the fluid is given by $c = \sqrt{K_f/\rho}$, where K_f is the compression modulus, and ρ is the mean fluid mass density.

The deformation \vec{u} of a solid is governed by the Lamé–Navier equations or equivalent plate or shell models. A generalized representation is given by

$$\mathcal{K}(\vec{u}) + \rho_s \frac{\partial^2 \vec{u}}{\partial t^2} = \vec{f}_s + \vec{p}_\Gamma, \quad (2)$$

where $\mathcal{K}(\cdot)$ is an operator governing equilibrium in the structure, often chosen as the divergence of the Cauchy stress tensor $\mathcal{K}(\vec{u}) = \nabla \cdot \vec{\sigma}$, $\vec{u} \in \mathbb{R}^3$. The solid mass density is given by ρ_s , volume forces are represented by \vec{f}_s , and \vec{p}_Γ is the distribution of surface pressure in the normal direction.

Acoustic fluid–structure coupling is imposed on the acoustic–structure interface Γ_i . The continuity of particle displacements and velocities in the normal direction \vec{n} are expressed in terms of the Euler equation as the Neumann coupling condition,

$$\rho \vec{n} \cdot \ddot{\vec{u}} = -\nabla p \cdot \vec{n} \quad \text{on } \Gamma_i. \quad (3)$$

The reaction force axiom

$$\vec{\sigma} \cdot \vec{n} = -p \vec{n} \quad \text{on } \Gamma_i, \quad (4)$$

here expressed in terms of the Cauchy stress tensor $\vec{\sigma}$ in the solid, is an additional Dirichlet coupling condition. Equations (1)–(4) describe the fully coupled acoustic–structure interaction problem.¹⁹

B. Finite element modeling

The finite element method provides a discrete but high-order model of the fully coupled pipe. D’Alembert’s principle is then applied to the structural field equations, and the weak form of the wave equation in the fluid is formulated. This leads to a coupled finite element formulation²⁰ in terms of nodal displacements \mathbf{u} and nodal acoustic pressures \mathbf{p} ,

$$\underbrace{\begin{bmatrix} \mathbf{M}_s & \mathbf{0} \\ \rho \mathbf{C}^T & \mathbf{M}_f \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} \ddot{\mathbf{u}} \\ \ddot{\mathbf{p}} \end{bmatrix}}_{\ddot{\mathbf{x}}} + \underbrace{\begin{bmatrix} \mathbf{D}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_f \end{bmatrix}}_{\mathbf{D}} \underbrace{\begin{bmatrix} \dot{\mathbf{u}} \\ \dot{\mathbf{p}} \end{bmatrix}}_{\dot{\mathbf{x}}} + \underbrace{\begin{bmatrix} \mathbf{K}_s & -\mathbf{C} \\ \mathbf{0} & \mathbf{K}_f \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} \mathbf{f}_s(t) \\ \mathbf{f}_f(t) \end{bmatrix}. \quad (5)$$

The system matrices of the acoustic partition in the fluid domain Ω_f are defined as

$$\mathbf{M}_f = \int_{\Omega_f} \frac{1}{c^2} \mathbf{R}^T \mathbf{R} d\Omega, \quad \mathbf{K}_f = \int_{\Omega_f} (\mathcal{D}_f \mathbf{R})^T \mathcal{D}_f \mathbf{R} d\Omega. \quad (6)$$

Here, c is the free sound speed in the fluid, ρ is the mean fluid mass density, \mathbf{R} are interpolation functions in matrix notation, and $\mathcal{D}_f = \nabla(\cdot)$ is a differential operator used for building spatial gradients. The structural mass matrix and stiffness matrix, \mathbf{M}_s and \mathbf{K}_s , respectively, represent standard finite element matrices, whereas matrix \mathbf{C} introduces

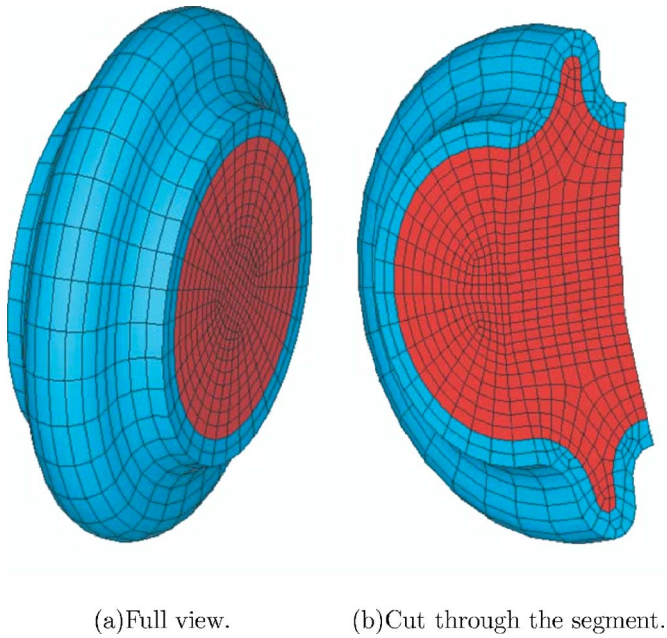


FIG. 2. (Color online) Model of a corrugated pipe segment with 10 873 DOFs including structural and acoustic finite elements.

acoustic–structure coupling on the interface in the normal direction. External nodal structural forces and fluxes are arranged in \mathbf{f}_s and \mathbf{f}_f , respectively. Linear damping in both the solid and in the fluid are integrated into the system equations (5) using damping matrices \mathbf{D}_s and \mathbf{D}_f . Damping effects are investigated in Sec. IV C.

A finite element model of one periodic segment having 10.873 degrees of freedom (DOFs) is depicted in Fig. 2.

III. WAVE ANALYSIS IN THE PIPE SEGMENT

According to the method of Mace *et al.*,^{12,21} one “segment” model (also called cell) which periodically repeats in the waveguide needs to be generated. Standard FE code can be applied to discretize the segment model. In this study, ANSYS²² was used. The corresponding finite element model of one fluid-filled pipe segment, as depicted in Fig. 3, on the left side of the pipe, is generated. Linear 8-node FLUID30 and SOLID45 elements are used in the acoustic and solid partitions, respectively. In Fig. 3, the geometry is continued periodically to illustrate the periodicity of the problem, but

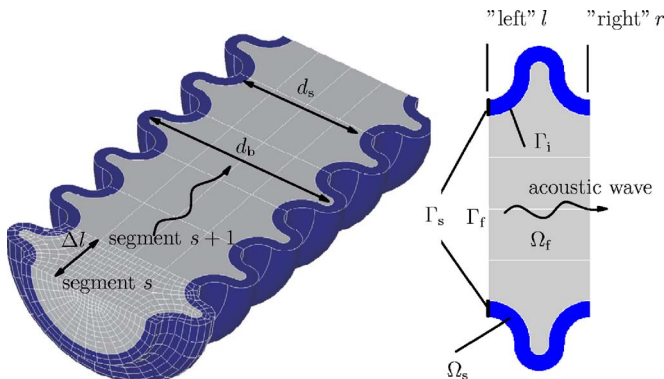


FIG. 3. (Color online) Waveguide FE model within the pipe and dimensions (left). Domain (Ω) and boundary (Γ) definitions (right).

no further FE mesh needs to be introduced on other segments. The displacement and pressure nodal degrees of freedom are formally arranged as $\mathbf{x}=[\mathbf{u}^T \mathbf{p}^T]^T$, and the forces and fluxes as $\mathbf{f}=[\mathbf{f}_s^T \mathbf{f}_f^T]^T$. Assuming only time-harmonic system behavior of the form $e^{i\omega t}$ in the waveguide, the model equations can be represented by the dynamic stiffness matrix

$$\tilde{\mathbf{H}}(\omega) = \mathbf{K} + i\omega\mathbf{D} - \omega^2\mathbf{M}, \quad (7)$$

where $\tilde{\mathbf{H}}(\omega)\mathbf{x}=\mathbf{f}$. In the next step, all DOFs are partitioned into interior, left, and right portions with respect to their positions on the pipe segment, i.e., $\mathbf{x}=[\mathbf{x}_i^T \mathbf{x}_l^T \mathbf{x}_r^T]^T$ and $\mathbf{f}=[\mathbf{f}_i^T \mathbf{f}_l^T \mathbf{f}_r^T]^T$ as shown in Fig. 3 (right). The dynamic stiffness matrix and the equations of motion are then partitioned accordingly,

$$\begin{bmatrix} \tilde{\mathbf{H}}_{ii} & \tilde{\mathbf{H}}_{il} & \tilde{\mathbf{H}}_{ir} \\ \tilde{\mathbf{H}}_{li} & \tilde{\mathbf{H}}_{ll} & \tilde{\mathbf{H}}_{lr} \\ \tilde{\mathbf{H}}_{ri} & \tilde{\mathbf{H}}_{rl} & \tilde{\mathbf{H}}_{rr} \end{bmatrix} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_l \\ \mathbf{x}_r \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_l \\ \mathbf{f}_r \end{bmatrix}. \quad (8)$$

It is assumed that there are no external loads applied to the interior degrees of freedom. Considering the first line of Eq. (8), the interior DOFs are expressed by

$$\mathbf{x}_i = -\tilde{\mathbf{H}}_{ii}^{-1}(\tilde{\mathbf{H}}_{il}\mathbf{x}_l + \tilde{\mathbf{H}}_{ir}\mathbf{x}_r). \quad (9)$$

This allows an exact dynamic condensation of the interior DOFs, and leads to the system representation

$$\begin{bmatrix} \mathbf{H}_{ll} & \mathbf{H}_{lr} \\ \mathbf{H}_{rl} & \mathbf{H}_{rr} \end{bmatrix} \begin{bmatrix} \mathbf{x}_l \\ \mathbf{x}_r \end{bmatrix} = \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_r \end{bmatrix}, \quad (10)$$

where the single partitions of the condensed dynamic stiffness matrix are defined as

$$\begin{aligned} \mathbf{H}_{ll} &= \tilde{\mathbf{H}}_{ll} - \tilde{\mathbf{H}}_{li}\tilde{\mathbf{H}}_{ii}^{-1}\tilde{\mathbf{H}}_{il}, & \mathbf{H}_{lr} &= \tilde{\mathbf{H}}_{lr} - \tilde{\mathbf{H}}_{li}\tilde{\mathbf{H}}_{ii}^{-1}\tilde{\mathbf{H}}_{ir}, \\ \mathbf{H}_{rl} &= \tilde{\mathbf{H}}_{rl} - \tilde{\mathbf{H}}_{ri}\tilde{\mathbf{H}}_{ii}^{-1}\tilde{\mathbf{H}}_{il}, & \mathbf{H}_{rr} &= \tilde{\mathbf{H}}_{rr} - \tilde{\mathbf{H}}_{ri}\tilde{\mathbf{H}}_{ii}^{-1}\tilde{\mathbf{H}}_{ir}. \end{aligned} \quad (11)$$

It should be noted that the computation of the inverse $\tilde{\mathbf{H}}_{ii}^{-1} \in \mathbb{C}^{n \times n}$ becomes expensive for large dimensions n (in this paper $n \rightarrow 10\,000$). Instead, the expressions $\tilde{\mathbf{H}}_{ii}^{-1}\tilde{\mathbf{H}}_{il}$ and $\tilde{\mathbf{H}}_{ii}^{-1}\tilde{\mathbf{H}}_{ir}$ are directly computed, since the nonquadratic matrices $\tilde{\mathbf{H}}_{il}, \tilde{\mathbf{H}}_{ir} \in \mathbb{C}^{n \times m}$ reduce the dimension of the matrix product. In these investigations, m is about one-tenth of n . In addition, it is worth pointing out that the node locations on the “right” interface DOFs are made to coincide with the node locations of the “left” DOFs. The system of equations (10) is now rearranged to a slightly modified form, which relates the Dirichlet and Neumann data on the left segment interface with the corresponding properties on the right interface,

$$\begin{bmatrix} \mathbf{H}_{ll} & -\mathbf{I} \\ \mathbf{H}_{rl} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_l \\ \hat{\mathbf{f}}_l \end{bmatrix} = \begin{bmatrix} -\mathbf{H}_{lr} & \mathbf{0} \\ -\mathbf{H}_{rr} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_r \\ \hat{\mathbf{f}}_r \end{bmatrix}. \quad (12)$$

The order of the left DOFs \mathbf{x}_l and the right DOFs \mathbf{x}_r must match, so that continuity and periodicity conditions between the data on the right side of the s th segment and the data on the left side of the virtual $s+1$ th segment (Fig. 3) may be applied. These conditions are

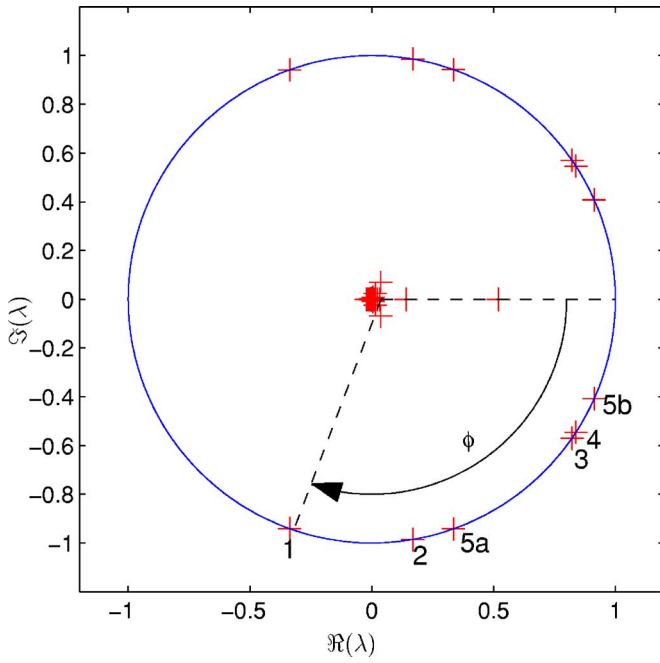


FIG. 4. (Color online) Eigenvalues λ around the unit circle at 4.53 kHz.

$$\mathbf{x}_l^{s+1} = \mathbf{x}_r^s \quad (13)$$

for the nodal Dirichlet data and

$$\mathbf{f}_l^{s+1} = -\mathbf{f}_r^s \quad (14)$$

for the nodal Neumann data, where the reaction force axiom is applied. This leads to the frequency-dependent transfer matrix system representation

$$\begin{bmatrix} \mathbf{H}_{ll} & -\mathbf{I} \\ \mathbf{H}_{rl} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_l^s \\ \hat{\mathbf{f}}_l^s \end{bmatrix} = \begin{bmatrix} -\mathbf{H}_{lr} & \mathbf{0} \\ -\mathbf{H}_{rr} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_l^{s+1} \\ \hat{\mathbf{f}}_l^{s+1} \end{bmatrix}. \quad (15)$$

Furthermore, harmonic wave propagation assumptions for free wave modes hold in the z direction,

$$\begin{aligned} \mathbf{x}_l^{s+1} &= e^{\kappa \Delta \ell} \mathbf{x}_l^s, \\ \mathbf{f}_r^{s+1} &= e^{\kappa \Delta \ell} \mathbf{f}_l^s, \end{aligned} \quad (16)$$

where $\Delta \ell$ is the segment length, and κ is the complex circular wave number. Combining Eqs. (15) and (16) gives a generalized eigenvalue problem

$$\begin{bmatrix} \mathbf{H}_{ll} & -\mathbf{I} \\ \mathbf{H}_{rl} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_l \\ \hat{\mathbf{f}}_l \end{bmatrix} = \lambda \begin{bmatrix} -\mathbf{H}_{lr} & \mathbf{0} \\ -\mathbf{H}_{rr} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_l \\ \hat{\mathbf{f}}_l \end{bmatrix} \quad (17)$$

with the eigenvalues $\lambda = e^{\kappa \Delta \ell}$. Alternatively, an eigenvalue problem is obtained by multiplying the matrix equation (17) by the inverse of the matrix on the right side of the equation. Although this formulation of the eigenvalue problem has been favored by most authors, direct solution of the generalized eigenvalue problem according to Eq. (17) is preferred in this study, since the matrices are better conditioned. Thus, shorter computing times result if a QZ-solver is applied for the eigenvalue problem.²³ The spectrum in the vicinity of the unit circle for a given pipe configuration is depicted in Fig. 4. The physical complex wave numbers κ are then computed from the eigenvalues λ by

$$\kappa = \frac{\text{Ln}(\lambda)}{\Delta \ell}. \quad (18)$$

For a fixed frequency ω , the number of numerically computed solutions of κ equals the dimension of the transfer matrix system. The logarithm of the complex eigenvalue

$$\lambda = |\lambda| e^{i(\phi + 2\pi m)}, \quad m = 0, 1, 2, 3, \dots \quad (19)$$

is

$$\text{Ln}(\lambda) = \ln(|\lambda|) + i(\phi + 2\pi m). \quad (20)$$

The spatial variation of the wave amplitude is given by the real part of the complex wave number, whereas the imaginary part describes wave propagation. Furthermore, it can be seen that the spectrum contains eigenvalues in pairs of λ_j and $1/\lambda_j$ representing left-propagating and right-propagating wave modes. Although no mathematical proof for this property is performed for the nonsymmetric system matrices as found for symmetric system representation in Mace *et al.*¹² so far, the underlying physics require this property. For example, the use of an alternative symmetric system representation for the acoustic–structure problem must result in the same eigenvalues, which are independent of the system formulation. Furthermore, it is obvious that only one set of them—in this case right propagating wave modes—needs to be considered further. According to the definition in Eq. (16), they are associated with negative phase angles ϕ in the complex plane. For practical reasons, only a very limited number of the remaining eigenvalues represent solutions of interest. Right-propagating wave modes with $|\lambda| > 1$ increase their amplitude according to Eq. (16) while propagating, and are considered nonphysical. Furthermore, wave modes associated with eigenvalues $|\lambda| \ll 1$ decay exponentially very quickly, so that they do not propagate over far distances. Hence, these eigenvalues are excluded from further consideration. For these reasons, only wave modes with negative phase angles and nonincreasing amplitudes are analyzed in the consecutive chapters. The associated eigenvalues λ are therefore in the lower half of the complex plane on or immediately inside the unit circle. They are labeled in Fig. 4 according to the wave branches which will be discussed in the following sections. An exhaustive discussion of relevant wave modes has been given by Mace *et al.*¹² and Maess *et al.*¹⁴

The wave analysis is restricted to modes whose wave lengths fulfill the condition

$$\frac{\pi}{\kappa} > \Delta \ell, \quad (21)$$

or in other words, the half wavelength should be longer than the segment length $\Delta \ell$. This wave number results from uniqueness considerations similar to Nyquist's sampling theorem.¹² For an increasing wave number and $\pi/\kappa \leq \Delta \ell$, the eigenvalue λ in the complex plane associated with right-propagating waves would have a negative phase angle $\phi < -\pi$. The eigenvalue would appear in the upper half of the complex plane, and would therefore be ignored according to the eigenvalue selection criteria. This behavior is also reflected by Eq. (19), where the logarithm for a complex ei-

TABLE I. Material properties of the corrugated pipe and the fuel.

Material	Mass density $\rho(\text{kg/m}^3)$	Young's modulus $E(\text{MPa})$	Poisson ratio ν	Free sound speed $c(\text{m/s})$
Shell: Polymer Rilsan	1050	330	0.35	...
Fluid: Engine fuel	800	1250

genvalue is not uniquely defined by m , although $m=0$ holds for wave modes starting with a zero wave number. For wave numbers above the upper limit, an additional criterion would be required to track the modes to a higher wave number regime. For the corrugated pipe and its given segment length $\Delta \ell = 4$ mm, an upper bound of $\kappa_{\text{max}} = 785.4 \text{ m}^{-1}$ exists. Investigations in Sec. IV show that all wave modes are correctly represented in a frequency band up to 6.7 kHz, which is sufficient for the underlying vibro-acoustic problem.

IV. DISPERSIVE WAVE MOTION IN FLUID-FILLED CORRUGATED PIPES

The dispersion characteristics of free wave motion for the fuel-filled corrugated pipe as part of the in-tank fuel assembly are numerically computed. To generate dispersion curves, the wave analysis from the previous section is applied to all frequencies of interest, i.e., the eigenvalue problem is solved numerically for each prescribed frequency.

The material properties of the corrugated pipe components are given in Table I, and dimensions are listed in Table II. The corrugated shape consists of quarter model slices, which are rotated around the symmetry axis.

First, the undamped case is considered, for which the damping matrices in Eq. (5) vanish. The propagation characteristics of the relevant wave modes can be visualized by computing the dispersion curves. Modal energy velocities, power flows, and energy densities are obtained by postprocessing the different branches of the dispersion diagram. The dispersion curves are obtained after solving the eigenvalue problem (17) associated with the transfer matrix representation. From the eigenvalues, wave numbers are recovered, whereas the eigenvectors describe the lateral shapes of the wave modes. The dispersion diagram of the corrugated pipe is shown in Fig. 5 for only the propagating acoustic wave modes. Four different branches are visible in the complete frequency domain. The bending wave mode (branch 1) has properties of a Bernoulli beam-type wave and undergoes strong dispersive behavior, especially at low frequencies. The compressional–extensional shell mode (branch 2) is less dispersive in the computed frequency domain. The dispersion curves for the plane–wave fluid mode (branch 3) and for

TABLE II. Dimensions of the corrugated pipe in (mm) (Fig. 3).

Smallest diameter d_s	Largest diameter d_b	Wall thickness t	Segment length Δl
7.20	11.40	0.65	4.00

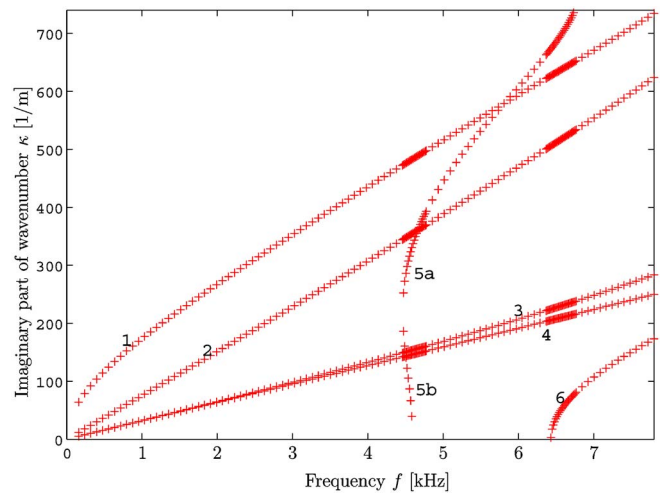


FIG. 5. (Color online) Dispersion curves.

the torsional mode (branch 4) are close to each other, they have smaller slopes as compared to the bending and the compressional–extensional modes. Moreover, two cut-on modes (branch 5 and branch 6) start propagating acoustically at 4.46 and 6.40 kHz, respectively. Branch 5 is a so-called double-valued wave mode, with subbranch 5a having a positive slope, and with subbranch 5b having a slightly negative slope after cutting on. The subbranch 5b mode cuts off again at 4.60 kHz.

An animation tool is used to visualize the different wave modes. The FE mesh of 11 periodic corrugated pipe segments is continued in the z direction periodically, and the wave forms of the free wave modes are displayed by combining the segment results periodically in the direction of wave propagation. It should be noticed that this step is only performed for illustration purposes, and that no analysis is done on these additional segments.

The plane–wave fluid-type mode is illustrated in Fig. 6(c). Clearly, consecutive positive and negative pressure pulsations are observed to propagate through the pipe. For this mode, the acoustic field strongly interacts with the flexible shell. A complex mechanism of expansion and compression of the pipe shell and superimposed bending over the segment radii is observed. A positive acoustic pressure causes an overall axisymmetric widening of the pipe, whereas the pipe contracts for negative acoustic pressure zones.

The longitudinal compressional–extensional shell mode is characterized by the axisymmetric deformations of the pipe segments in longitudinal directions [Fig. 6(d)]. In contrast to the plane–wave fluid mode, the radial extension of the pipe shell is less dominant. The interior acoustic field is also less pronounced.

Figures 6(a) and 6(b) illustrate the bending mode together with the associated acoustic field. The wave form is dominated by a tilting of the corrugated pipe segments and by an acoustic field “hugging” the pipe, where opposite sides of the corrugated pipe segments show pressure maxima. The overall structural view of the bending mode illustrates the bending of the segments with respect to each other.

The animation tool helps to interpret the shapes of the cut-on modes. The first cut-on mode (branch 5a) turns out to

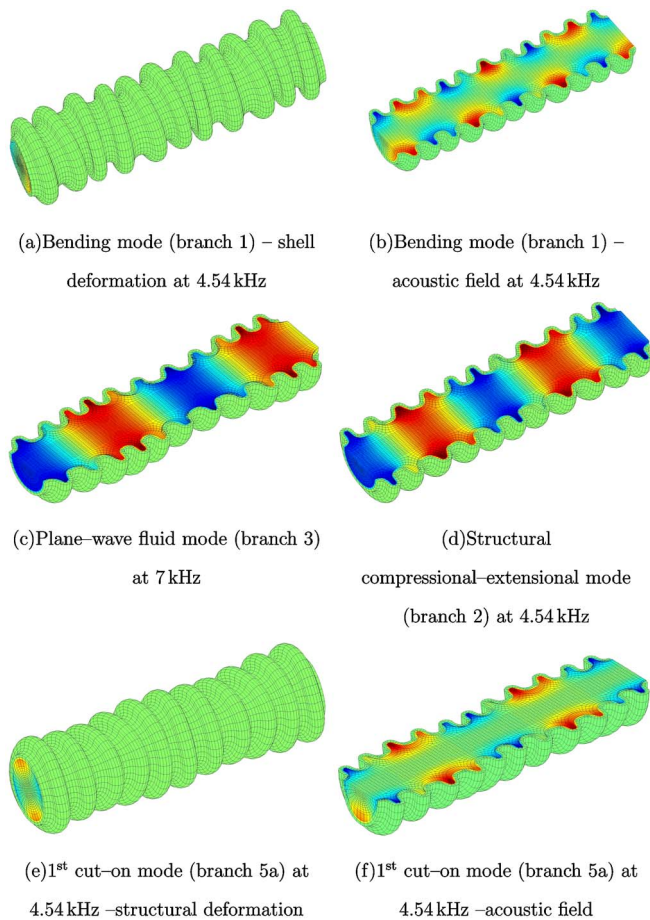


FIG. 6. (Color online) Illustration of the free wave modes.

be a higher-order circumferential bending-type mode (shell bending), as shown in Figs. 6(e) and 6(f). Hereby, the shell cross section undergoes elliptical deformation. The second cut-on mode (branch 6) has an acoustic field similar to the one of the first acoustic cut-on mode in a rigid wall acoustic duct. Its wave shape is characterized by a nodal line through the middle of the cross section. The shell undergoes some motion reminding of bending, where the segments are tilted against each other, but here they are not displaced largely from the center axis. The analysis of energy ratios in Sec. IV B will show that most energy is located in the shell, therefore the term “acoustic mode” might be misleading.

Computational errors result from the discretized treatment of the corrugated pipe segment. Maess *et al.*¹⁴ have shown the convergence of the computed dispersion cases for refined meshes in the case of uniform fluid-filled pipes. Figure 7 shows the dispersion curves for different mesh sizes in the case of the corrugated pipe model. It is obvious that the dispersion curves converge for accurate meshing. While the wave numbers of a roughly meshed pipe segment (1677 DOFs) are shifted to higher frequencies, the refined meshes (6500 and 10873 DOFs, respectively) show asymptotic behavior. The solution of the largest model is considered to be sufficiently accurate, and it is therefore applied throughout this paper.

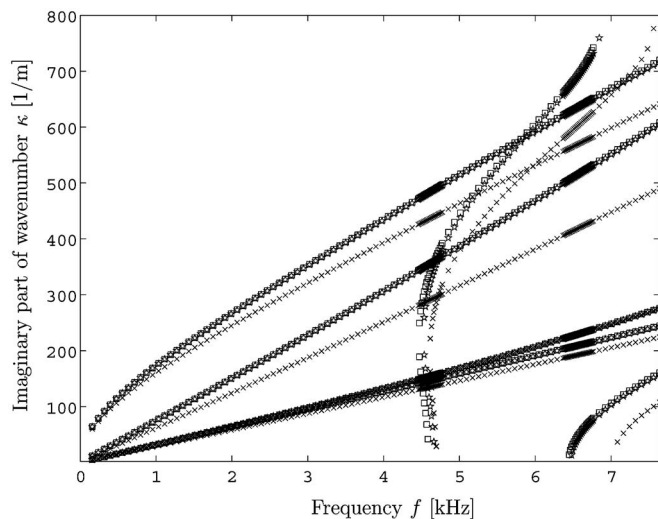


FIG. 7. Influence of the mesh size on the wave numbers. 1677 DOFs (×); 6500 DOFs (*); 10873 DOFs (□).

A. Phase and energy velocities

The dispersion curves contain information about the free wave motion. For engineering purposes, it is often more convenient to express dispersion in terms of wave speeds, namely the phase velocity and the energy velocity. The phase velocity is the velocity at which equal phases travel, and it is computed using the relation

$$c_p = \frac{\omega}{\kappa}. \quad (22)$$

According to the previous section, only right-propagating wave modes in terms of the phases are considered. Therefore, all wave modes have positive phase speeds. Phase velocities are shown in Fig. 8 as a function of the frequency. Typically, the Bernoulli beam-type bending mode (branch 1) starts with zero phase velocity and approaches the phase velocities of the longitudinal compressional-extensional shell mode (branch 2) for higher frequencies. As expected from the dispersion curve, the longitudinal mode propagates with an almost constant phase velocity. The plane-wave fluid and torsional modes (branch 3 and branch 4, respectively) show higher phase velocities as compared to the bending and longitudinal modes.

The phase velocity of the torsional mode is nearly constant, whereas the velocity of the plane-wave fluid mode

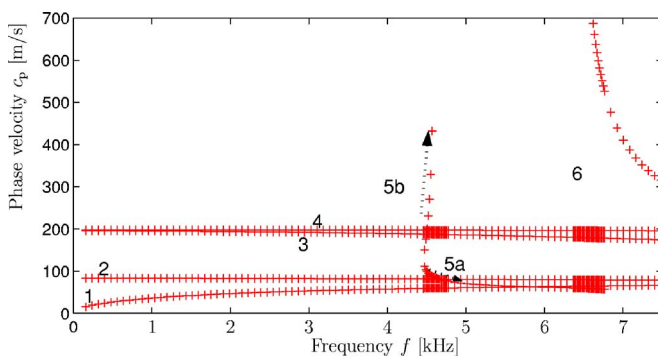


FIG. 8. (Color online) Phase velocities.

decreases for higher frequencies. At low frequencies, the wave speed in the fluid in uniform flexible pipes or hoses for plane waves is usually approximated by the nondispersive Korteweg-Moens correction,²⁴

$$c^k = \sqrt{\frac{c}{1 + \frac{2c^2 r}{Et}}}, \quad (23)$$

which assumes concentric quasistatic compressibility of the pipe shell. Here, t is the shell thickness and r is the mean radius. The minimum and maximum radii can be applied to find an upper and lower bound for reference, i.e., $d_s/2+t/2$ and $d_b/2+t/2$ from Table II. Computations yield a lower bound of $c_b^k=149 \text{ m s}^{-1}$, and an upper bound of $c_s^k=184 \text{ m s}^{-1}$. The three-dimensional finite element model analysis predicts a numerical phase velocity of $c_p=195 \text{ m s}^{-1}$ at low frequencies. This means in reality that the corrugated shell is stiffer in the radial direction as compared to a uniform pipe, thus generating a faster phase speed. This is plausible, since the segmentation ribs act as stiffeners.

The first cut-on mode (double-valued branch 5 at 4.46 kHz) cuts in at a phase velocity of 100 m s^{-1} . The phase velocity of subbranch 5a increases for higher frequencies following the dashed arrow. Subbranch 5b cuts off at 4.6 kHz and approaches an infinite phase velocity. This behavior is explained by a standing vibration mode in the lateral directions at the cut-on frequency. Branch 6 cuts on at 6.4 kHz with an infinite phase velocity, which decreases with increasing frequencies. For frequencies below the cut-on frequencies, the corresponding wave modes are not able to propagate acoustically.⁷

For many engineering problems, one is particularly interested in the group velocity c_g . It describes the velocity at which wave packages propagate, and it is therefore identical to the energy velocity in the undamped case. Group velocities are derived from the dispersion curves²⁵ as

$$c_g = \frac{\partial \omega}{\partial \kappa}. \quad (24)$$

A numerical derivation, from discrete pairs (ω, κ) may result in relatively large errors. For this reason, it is more convenient to directly evaluate the energy velocity using energy densities and sound power properties.²⁶ The energy velocity is equal to the ratio of transmitted power P and energy density \bar{E} in the waveguide,

$$c_g = \frac{P}{\bar{E}} = \frac{P_f + P_s}{\bar{T}_s + \bar{T}_f + \bar{U}_s + \bar{U}_f}. \quad (25)$$

Equation (25) is decomposed into solid domain and fluid domain contributions. For the discrete acoustic-structure interaction problem, energy densities are computed by means of finite element system matrices and discrete wave modes, which are given by the eigenvectors in Eq. (17). The kinetic energy T_s and the potential energy U_s in a linear-elastic solid segment are given as

$$T_s = \int_{\Omega_s} \frac{1}{2} \rho_s \bar{v}^2 d\Omega, \quad U_s = \int_{\Omega_s} \frac{1}{2} \Re \left(\tilde{\sigma} : \tilde{\varepsilon}^H \right) d\Omega, \quad (26)$$

where v is the particle velocity, ε is the linear strain tensor, and σ is the Cauchy stress tensor. The subscript $(\cdot)^H$ denotes complex conjugate transposed values. The tilde operator $(\tilde{\cdot})$ denotes root mean square values, and it takes into account the effective time-averaged energy density in the acoustic field for the case of harmonic wave propagation. The discrete finite element representation of the energy densities in a pipe segment of length $\Delta \ell$ becomes

$$\bar{T}_s = \frac{1}{4\Delta \ell} \omega^2 \hat{\mathbf{u}}^H \mathbf{M}_s \hat{\mathbf{u}}, \quad \bar{U}_s = \frac{1}{4\Delta \ell} \hat{\mathbf{u}}^H \mathbf{K}_s \hat{\mathbf{u}}. \quad (27)$$

The time-averaging leads to a factor of 4 in the denominator, since $\tilde{\mathbf{u}} = \hat{\mathbf{u}}/\sqrt{2}$. In the acoustic domain, kinetic and potential energy are expressed as²⁷

$$T_f = \int_{\Omega_f} \frac{1}{2} \rho \bar{v}^2 d\Omega, \quad U_f = \int_{\Omega_f} \frac{1}{2\rho c^2} \bar{p}^2 d\Omega. \quad (28)$$

Substituting definitions for the acoustic system matrices from Eq. (6) into Eq. (28), and applying Euler's equation in the frequency domain for the dissipation-free case,

$$i\omega \rho \hat{\mathbf{v}} = -\nabla \hat{p}, \quad (29)$$

leads to expressions for energy densities in the acoustic fluid for discrete wave modes,

$$\bar{T}_f = \frac{1}{4\Delta \ell \rho \omega^2} \hat{\mathbf{p}}^H \mathbf{K}_f \hat{\mathbf{p}}, \quad \bar{U}_f = \frac{1}{4\Delta \ell \rho} \hat{\mathbf{p}}^H \mathbf{M}_f \hat{\mathbf{p}}. \quad (30)$$

The time-averaged sound power flows through the pipe cross sections Γ_s and Γ_f (Fig. 3, right) are

$$P_s = \int_{\Gamma_s} \tilde{\sigma} \tilde{\mathbf{v}} d\Gamma, \quad P_f = \int_{\Gamma_f} \tilde{p} \tilde{\mathbf{v}} d\Gamma. \quad (31)$$

For discrete wave modes in the frequency domain, the power flow through the solid cross section Γ_s is

$$P_s = \frac{1}{2} \omega \Im \left(\hat{\mathbf{f}}_s^H \hat{\mathbf{u}} \right), \quad (32)$$

and the power flow through the fluid cross section Γ_f is

$$P_f = \frac{1}{2\rho\omega} \Im \left(\hat{\mathbf{f}}_f^H \hat{\mathbf{p}} \right). \quad (33)$$

Although some algebraic effort is required to derive Eqs. (25)–(33), the energy velocities are computed accurately using Eq. (25) without large numerical effort.

Energy velocities for all propagating wave modes in the corrugated pipe are depicted in Fig. 9. The first cut-on mode (branch 5b) has a slightly negative group velocity. A negative group velocity indicates that the wave carries energy in the opposite direction to the direction of the phase propagation. This “backward” wave phenomenon²⁸ has also been reported for sandwich panels.¹¹

At low frequencies, the group velocities of the basic wave modes are similar to the previously described phase velocities. It is worth mentioning that the torsional wave

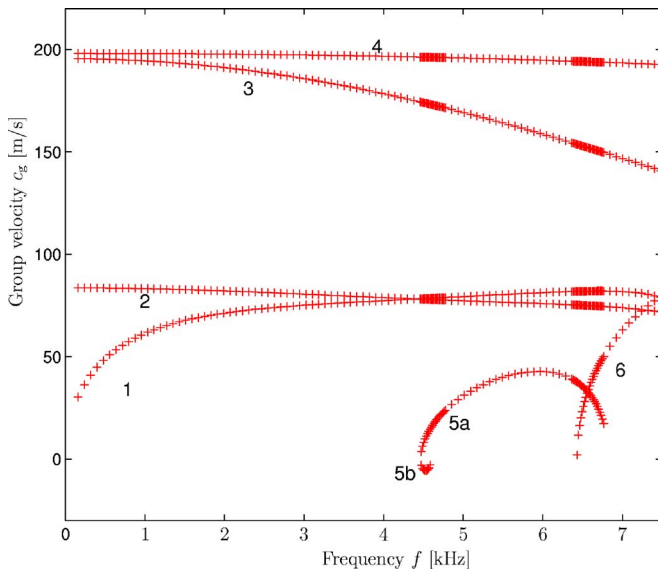


FIG. 9. (Color online) Group velocities.

mode has the highest wave speed at low frequencies. In fact, this unexpected behavior is due to the segmentation over the radii, which inserts additional flexibility for the longitudinal structural dynamics. As a result, the effective longitudinal stiffness is reduced. Furthermore, the fluid has added hydro-mass properties which increase the inertia of the structure. For these reasons, the effective wave speed of the extensional-compressional longitudinal wave decreases to as low as about 90 m s^{-1} . By comparison, the longitudinal structural wave in a straight uniform pipe of the same material would have a wave speed²⁹ of $c_1 = \sqrt{\frac{E}{\rho_s(1-\nu^2)}} = 598 \text{ m s}^{-1}$.

B. Energy ratio

The modal distribution of vibrational energy in the fluid and solid partitions for free wave propagation is determined using Eqs. (26) and (28) and the computed eigenvectors. This analysis is helpful to understand whether the vibrational energy is transmitted via the structural domain or the fluid domain over long distances. This knowledge leads to more effective design measures for noise reduction either on the pipe shell (for example, support elements or friction dampers) or in the fluid path (for example, pulsation dampers). The ratio of vibrational energy in the fluid and structure domains is given by $\bar{E}_f/\bar{E}_s = (\bar{T}_f + \bar{U}_f)/(\bar{T}_s + \bar{U}_s)$. Note that the ratio of energy densities is equal to the ratio of vibrational energies, since energy quantities are scaled by the constant segment length $\Delta\ell$. The modal energy ratios as a function of frequency are depicted in Fig. 10.

It is found that no wave mode contains more vibrational energy in the fluid than in the shell. Even in the so-called “plane-wave fluid” mode (branch 3), the fluid contains only up to about one-third of the overall energy. Thus, a considerable amount of the vibrational energy is located in the shell. On the other hand, the “structural longitudinal compressional-extensional” mode (branch 2) has a ratio 0.2–0.4 for all computed frequencies. A similar energy ratio is found for the bending wave mode (branch 2). However, un-

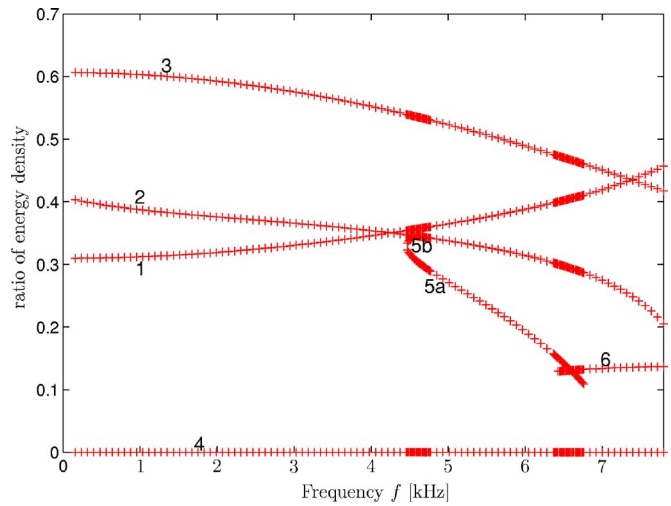


FIG. 10. (Color online) Distribution of vibrational energy \bar{E}_f/\bar{E}_s as a function of frequency.

like the longitudinal mode, the energy in the fluid increases with frequency. Considerable portions of the vibrational energy are located in the acoustic domain for both “structural” wave modes. This is not surprising since the comparable impedances of the shell material and of the engine fuel lead to strong acoustic-structure interaction.

The torsional mode (branch 4) is characterized by the fact that all the energy is located in the shell, and no vibrational energy is located in the fluid. This behavior is explained by the fact that the twisting of the segments generates negligible out-of-plane deformation on the acoustic-structure interface. As a result, the interior acoustic field is not influenced by the torsional wave motion of the shell. Both cut-on modes have energy ratios of 0.1–0.31, indicating that the larger part of the vibrational energy is distributed in the shell. However, the mode of branch 5a increases its energy portion in the shell for increasing frequencies.

Finally, it should be noted that this analysis of the energy ratio does not include information about the excitability of these wave modes, since free solutions of wave motion have an indeterminate amplitude. Mathematically, this physical behavior is reflected by the eigenvalue problems in Eq. (17), where eigenvectors can be scaled with any nonzero scalar number. Therefore, all properties related to absolute amplitudes of the wave vectors are scalable (e.g., vibrational energies, lateral wave shapes), whereas wave numbers, wave speeds, and energy ratios remain unaffected. Also, the ratio of sound powers depends on where the analysis is performed. For example, periodicity may be formulated between locations having minimum or maximum radii. The two formulations would lead to different sound power ratios. This is a major difference as compared to uniform waveguides, where the ratio of sound power remains constant and equals the energy ratio.⁷

C. The influence of damping on the dispersion curves

Damping effects in the structure and in the fluid alter the propagation characteristics of waves. In the presence of

damping, the amplitudes of wave modes decay as they propagate, and the wavelength may also be altered. In general, damping is present in both the structural and in the fluid domain.

The corrugated pipe consists of a flexible polymer with a considerable material damping. In the case of the structural pipe shell, frequency-dependent damping is experimentally determined using modal measurements. In general, a large number of damping models exist.³⁰ One of the simplest approaches uses a Rayleigh damping model. Specifically, a damping matrix \mathbf{D}_s having the form

$$\mathbf{D}_s = \alpha_s \mathbf{M}_s + \beta_s \mathbf{K}_s \quad (34)$$

is integrated into the FE equations (5). This damping matrix model with the fitted parameters $\alpha_s = 42 \text{ s}^{-1}$ and $\beta_s = 8.1 \times 10^{-6} \text{ s}$ approximates measured modal damping fairly accurately in the bandwidth of interest.

On the other hand, the mechanism of acoustic damping in the fluid is more complex. In general, fluids exhibit viscous damping. The wave equation from Eq. (1) is derived by considering Stoke's assumption¹⁸ as

$$\nabla^2 p + \frac{4\nu_k}{3c^2} \frac{\partial}{\partial t} \nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0. \quad (35)$$

Here, ν_k is the kinetic viscosity. From Eq. (35) and with the definitions of the finite element matrices, frequency-proportional damping of the form

$$\mathbf{D}_f = \beta_f \mathbf{K}_f \quad (36)$$

is derived, where $\beta_f = 4\nu_k/3c^2$. In general, the viscosity is not the same for different engine fuels, and it shows a strong temperature-dependency. Typical kinetic viscosities range from $\nu_k = 10$ to 100 cSt (10^{-4} – $10^{-5} \text{ m}^2/\text{s}$). Therefore, the damping parameter is on the order of $\beta_f = 10^{-10} \text{ s}$. However, this damping model only provides a very low bound for damping, since other damping mechanisms occur.³¹ For example, the existence of wall shear stress, particularly in small diameter pipes, leads to rapid decay of wave amplitudes. In the case of damped plane-wave modes in the fluid, damping properties depend on the particle velocity profile, which in turn depends on the frequency of the pressure pulsations. Models in one dimension have been described mathematically and verified experimentally by Theissen³² for uniform pipes. Furthermore, heat conduction between the fluid and the shell may also influence the attenuation of the waves. However, a general linear model including damping for all fluid modes in the corrugated pipe is not available at this point. For this reason, the viscous damping property in Eq. (36) is set to an artificially increased value of $\beta_f = 10^{-8} \text{ s}$ to study qualitatively the impact of damping on free wave propagation. It is worth mentioning that other linear damping models can be integrated into the finite element model using a similar approach.

Although the damping values for the fluid have been somewhat arbitrarily selected and do not account for influences from velocity gradients or heat conduction, the selected assumptions enable qualitative evaluation of the influence of damping on the dispersion curves. The dispersion

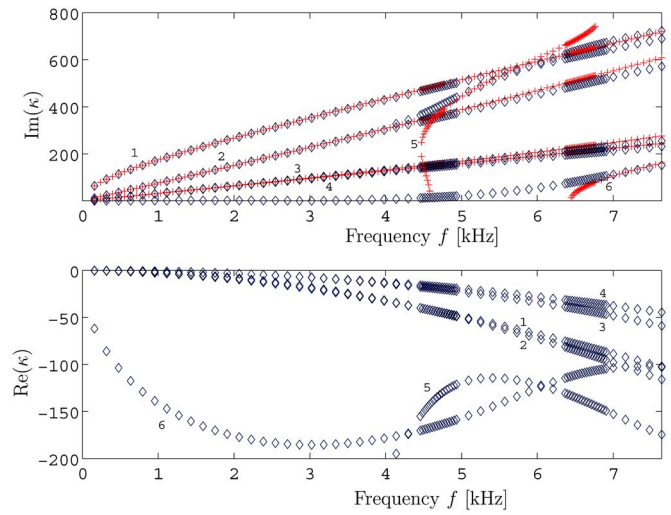


FIG. 11. (Color online) Influence of the material damping on the wave numbers κ . Without damping (+); including material damping (\diamond).

curves for the corrugated pipe model with damping are now shown in Fig. 11 (diamonds). The dispersion curves for the corrugated pipe model without damping are displayed in the same figure (plus signs) for comparison. Both the imaginary and real parts of the wave number κ are considered. It is obvious that damping leads to lower wave numbers for constant frequency for all low-frequency wave modes (branches 1–4). This indicates that the wave speeds increase with damping, which is a typical result of viscous damping due to the increased dynamic stiffness in the system. Moreover, the previous cut-on modes (branches 5 and 6) are largely modified. Both modes are now oscillatory modes for the complete observed frequency domain. Subbranch 5b disappears, and the corresponding wave mode becomes single-valued. Additionally, both modes 5 and 6 have large negative real parts of the wave number, thus indicating a strong attenuation in the propagation direction. Similar phenomena have been reported for cut-on modes in damped plate-like structures.⁸ The acoustic attenuation in alternative units [dB/m] as a function of frequency for the individual modes is depicted in Fig. 12. It becomes obvious that the previous cut-on modes (branches 5 and 6) are no longer far-propagating wave modes under practical considerations.

It should be stressed that no results have been shown for the energy densities or for the energy velocities in the damped case. However, the consistent momentum equation belonging to the damped wave equation (35) is given in terms of the fluid velocities \vec{v} by

$$\rho \dot{\vec{v}} = -\nabla p - \beta \nabla \frac{\partial p}{\partial t}. \quad (37)$$

In the frequency domain, one obtains

$$i\omega\rho\vec{v} = -(1 + i\omega\beta) \nabla p. \quad (38)$$

As a result, the expression for the kinetic energy density in the fluid as given by Eq. (30) has to be corrected by a factor of $(1 + \omega^2\beta^2)$. However, this factor tends very close to one in the given frequency range and for the applied value of β , where $\omega\beta < 5 \times 10^{-4}$.

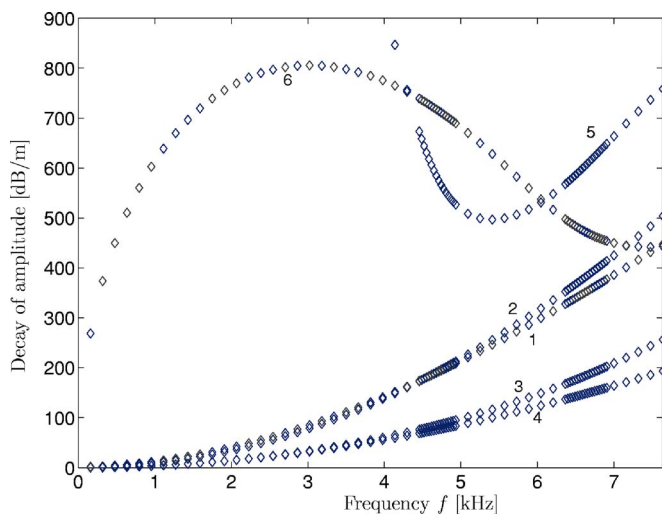


FIG. 12. (Color online) Attenuation of the wave modes due to the material damping as a function of frequency.

V. CONCLUSIONS

The characteristics of free wave motion in periodic corrugated fluid-filled pipes have been computed numerically. The combination of a finite element model for one periodic segment with the wave analysis approach offers the possibility of accurately predicting complex wave modes numerically. The method allows for simultaneous computation of all wave modes using a single model. Furthermore, the acoustic–structure interface is fully included in the discrete field equations and allows the treatment of the complex interaction between fluid- and structure-borne sound.

The results show that the wave motion in the corrugated pipe is considerably different as compared to a uniform pipe, since the segmentation of the pipe shell leads to complex mechanisms of compressional, bending, and acoustic fluid dynamics. The method renders accurate results for wave numbers less than 785.4 m^{-1} . Thus, all dispersion curves up to 6.7 kHz are completely evaluated. In postprocessing steps, phase and group velocities as well as the ratio between the fluid and structural vibrational energy for individual modes are computed. The influence of damping on the dispersion curves is studied qualitatively, thereby allowing prediction of the attenuation for the wave modes.

In general, any other fluid-filled waveguide having periodic properties in the direction of wave propagation can be investigated using this method. This study takes advantage of the discretization capabilities of a standard FE package for computing the system matrices of a segment model. No additional finite elements need to be implemented. The wave analysis method is implemented in a separate tool, which acts as a postprocessor to transform the segment models into dispersion curves.

ACKNOWLEDGMENTS

Funding of this project by the German Research Society DFG in the Transfer Unit TFB 51 “Simulation and Active Control of Hydroacoustics in Flexible Piping Systems” is

gratefully acknowledged. Furthermore, the authors wish to thank Dr.-Ing. Matthias Fischer at Robert Bosch GmbH for providing pipe data.

- ¹P. Kohmann, “Ein Beitrag zur Lärminderung bei flüssigkeitsbefüllten Rohrleitungen auf Schiffen,” Ph.D. thesis, University of Stuttgart, Stuttgart, 1995 (in German).
- ²C. A. F. de Jong, *Analysis of Pulsations and Vibrations in Fluid-Filled Pipe Systems* (TNO Institute of Applied Physics, Delft, Eindhoven, 1998).
- ³H. Lamb, “On the velocity of sound, as affected by the elasticity of the wall,” *Manchester Memoirs* **42**, 1–16 (1898).
- ⁴Y. Guo, “Approximate solutions of the dispersion equation for fluid-loaded cylindrical shells,” *J. Acoust. Soc. Am.* **95**, 1435–1445 (1994).
- ⁵W. Kuhl, “Die Eigenschaften wassergefüllter Rohre für Widerstands- und Schallgeschwindigkeitsmessungen,” (“Characteristics of water-filled pipes for measurements of acoustic flow and sound velocity”), *Acustica* **3**, 111–123 (1953).
- ⁶G. Pavic, “Acoustical analysis of pipes with flow using invariant field functions,” *J. Sound Vib.* **263**, 153–174 (2003).
- ⁷C. Fuller and F. Fahy, “Characteristics of wave propagation and energy distributions in cylindrical elastic shells filled with fluid,” *J. Sound Vib.* **81**, 501–518 (1982).
- ⁸I. Bartoli, A. Marzani, F. di Scalea, and E. Viola, “Modeling wave propagation in damped waveguides of arbitrary cross-section,” *J. Sound Vib.* **295**, 685–707 (2006).
- ⁹S. Finnveden and C.-M. Nilsson, “Waveguide finite elements for fluid-shell coupling,” in 11th international Congress on Sound and Vibrations, Stockholm, **2003**, pp. 501–508.
- ¹⁰S. Finnveden, “Evaluation of modal density and group velocity by finite element method,” *J. Sound Vib.* **273** (2004).
- ¹¹B. Mace, D. Duhamel, M. Brennan, and L. Hinke, “Wave number prediction using finite element analysis,” in 11th International Congress on Sound and Vibration, St. Petersburg, **2004**, pp. 3241–3248.
- ¹²B. Mace, D. Duhamel, M. Brennan, and L. Hinke, “Wave number prediction using finite element analysis,” *J. Acoust. Soc. Am.* **117**, 2835–2843 (2005).
- ¹³M. Maess and L. Gaul, “Dispersion in fluid-filled pipes by analyzing finite element models,” Proceedings of the 31st Annual Meeting of the German Acoustical Society DAGA, München, 2005.
- ¹⁴M. Maess, N. Wagner, and L. Gaul, “Dispersion curves of fluid-filled elastic pipes by standard FE models and eigenpath analysis,” *J. Sound Vib.* **296**, 264–276 (2006).
- ¹⁵M. Maess, “Methods for efficient acoustic–structure simulation of piping systems,” Ph.D. thesis, Bericht aus dem Institut für Angewandte und Experimentelle Mechanik, Stuttgart, 2006.
- ¹⁶M. Maess and L. Gaul, “Substructuring and model reduction of pipe components interacting with acoustic fluids,” *Mech. Syst. Signal Process.* **20**, 45–64 (2006).
- ¹⁷B. Brevart and C. Fuller, “Effect of an internal flow on the distribution of vibrational energy in an infinite fluid-filled thin cylindrical elastic shell,” *J. Sound Vib.* **167**, 149–163 (1993).
- ¹⁸L. Kinsler, A. Prey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
- ¹⁹P. Fahy, *Structural Sound and Vibration* (Academic, London, 1998).
- ²⁰O. Zienkiewicz and R. Taylor, *The Finite Element Method*, (Butterworth-Heinemann, Oxford, 2000), vol. I.
- ²¹D. Duhamel, B. Mace, and M. Brennan, “Finite element analysis of the vibrations of waveguides and periodic structures,” *J. Sound Vib.* **294**, 205–220 (2006).
- ²²ANSYS, Inc., “ANSYS Documentation 7.1” (2004).
- ²³A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics* (Springer, New York, 2000).
- ²⁴D. Korteweg, “Über die fortpflanzungsgeschwindigkeit des Schalles in elastischen Röhren,” *Ann. Phys.* **5**, 525–542 (1878).
- ²⁵J. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier North-Holland, Amsterdam, 1984).
- ²⁶S. Akrouf, M. Ichchou, J.-M. Mencik, and L. Jezequel, “Structural guided waves energy velocities by finite element methods,” in 12th International Conference on Sound and Vibration, Lisboa, 2005.
- ²⁷L. Cremer and M. Möser, *Technische Akustik* (Springer Berlin, 2003).
- ²⁸P. Marston, “Negative group velocity Lamb waves on plates and applications to the scattering of sound by shells,” *J. Acoust. Soc. Am.* **113**, 2659–2662 (2003).

- ²⁹L. Cremer, M. Heckl, and E. Ungar, *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies* (Springer, Berlin, 1990).
- ³⁰L. Gaul, "The influence of damping on waves and vibrations," *Mech. Syst. Signal Process.* **13**, 1–30 (2004).
- ³¹D. Ronneberger and C. Ahrens, "Wall shear stress caused by small amplitude perturbations of turbulent boundary layer flow—An experimental investigation," *J. Fluid Mech.* **83**, 433–464 (1977).
- ³²H. Theissen, "Die Berücksichtigung instationärer Rohrströmung bei der Simulation hydraulischer Anlagen," ("Simulation of hydraulic systems including instationary pipe flow"), Ph.D. thesis, RWTH Aachen, Aachen, 1983 (in German).

Radiation force and shear motions in inhomogeneous media

Lev Ostrovsky

*Zel Technologies/NOAA ESRL, 325 Broadway Boulder, Colorado 80305, and
University of Colorado at Boulder, Boulder, Colorado 80302*

Alexander Sutin

*Stevens Institute of Technology, Hoboken, New Jersey 07030 and
ARTANN Laboratories, Inc., 1459 Lower Ferry Road, Trenton, New Jersey 08618*

Yuri Il'inski

Applied Research Laboratories, University of Texas at Austin, Austin, Texas 78713-8029

Oleg Rudenko

Faculty of Physics, Lomonosov Moscow State University, Leninskie Gory, 119992, Moscow, Russia

Armen Sarvazyan

ARTANN Laboratories, Inc., 1459 Lower Ferry Road, Trenton, New Jersey 08618

(Received 24 May 2006; revised 13 November 2006; accepted 13 November 2006)

An action of radiation force induced by ultrasonic beam in waterlike media such as biological tissues (where the shear modulus is small as compared to the bulk compressibility) is considered. A new, nondissipative mechanism of generation of shear displacement due to a smooth (nonreflecting) medium inhomogeneity is suggested, and the corresponding medium displacement is evaluated. It is shown that a linear primary acoustic field in nondissipative, isotropic elastic medium cannot excite a nonpotential radiation force and, hence, a shear motion, whereas even smooth inhomogeneity makes this effect possible. An example is considered showing that the generated displacement pulse can be significantly longer than the primary ultrasound pulse. It is noted that, unlike the dissipative effect, the nondissipative action on a localized inhomogeneity (such as a lesion in a tissue) changes its sign along the beam axis, thus stretching or compressing the focus area. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2532113]

PACS number(s): 43.25.Qp [MFH]

Pages: 1324–1331

I. INTRODUCTION

The acoustic radiation force is, in general, defined as a period-average force exerted on the medium by a sound wave. Various aspects of this phenomenon in fluids have been discussed by numerous authors, beginning with Lord Rayleigh (e.g., Refs. 1–3). A classical example is sound pressure on interfaces or bodies in fluids; its application was proposed by Wood and Loomis as early as in the 1920s to measure total power in ultrasonic beams.⁴ As regards the distributed radiation force in homogeneous fluids, one of the most thorough studies has been carried out by Eckart,⁵ who consistently derived equations for the average force and the resulting motion in a viscous fluid. He showed that for a nonmodulated, harmonic sound in a homogeneous fluid, only a rotational (solenoidal) radiation force component can create fluid motion, and this is due to the presence of attenuation (viscosity). He also gave an example of acoustic streaming (“acoustic wind”) induced by such force. In the same work it was shown that the potential part of the averaged force (acoustic pressure) causes a static compression. Later it was shown that in a nonsteady (modulated) sound wave, the potential force can create low-frequency sound that was used in the well-known parametric arrays⁶ and, more recently, in nonlethal acoustic weapons.⁷

At present, there exists a large body of literature concerning “radiation force,” “radiation pressure,” and “radia-

tion stress” in fluids (e.g., Ref. 8) and in solids (e.g., Ref. 9). Note that the discussions are generally devoted to normal force (radiation pressure) acting at partially reflecting obstacles and/or scattering bodies (spheres, bubbles, etc.). In the cited paper by Eckart,⁵ the vortical (solenoidal) forces are considered that exist due to ultrasound attenuation in the medium and generate acoustic flows.

In waterlike elastic media such as biological tissues, the shear modulus is finite but small as compared to the bulk compressibility, and the action of radiation force can generate significant shear displacements. This effect is currently being extensively explored in medical applications related to remote assessment of viscoelastic properties of internal tissues and bodily fluids.⁹ Low-frequency displacements generated by focused ultrasound became the basis for numerous emerging diagnostic techniques such as the so-called shear wave elasticity imaging,^{10–12} acoustic radiation force impulse imaging,^{13–15} supersonic imaging,^{16,17} and vibroacoustography.^{18,19} Measurements of shear displacements produced by ultrasound beam can be conducted using magnetic resonance imaging^{11,20} and x-ray techniques.^{21,22} The expanding field of applications dictates more thorough analysis of physical mechanisms of shear radiation force generation in waterlike media.

As shown in this paper, there exist previously unexplored mechanisms of shear displacement generation by high

intensity ultrasonic beam other than the conventional mechanism related to attenuation of ultrasound wave. In particular, such nondissipative mechanism for generation of nonpotential radiation force can be related to medium inhomogeneity. Besides the conceptual significance of this problem, it would expand the opportunities for the medical diagnostic methods based on generation of shear stress in biological media by focused ultrasound.

This paper presents a theoretical model for shear stress and shear motions generation by a narrow-angle acoustic beam due to inhomogeneity of the acoustic parameters of a waterlike medium. We consider smooth inhomogeneities that do not lead to the reflection of acoustic waves. The forthcoming consideration is applicable to both a liquid and waterlike medium, although the amplitudes of the resulting motions can differ significantly due to the presence of small shear modulus in the latter. The paper is organized as follows. First, we obtain general expressions for average stress produced by a narrow-angle ultrasonic beam in an inhomogeneous elastic medium and then derive the expression for the corresponding nondissipative force, which is compared with that for the dissipative force. It is shown that the solenoidal part of the period-average force disappears in the absence of inhomogeneity and dissipation. Then the equation for the shear displacements in an inhomogeneous medium is obtained and solved for an example characteristic of a biological tissue. In the Appendix, the proof of the potentiality of the radiation force generated by a linear potential primary beam in a homogeneous medium is extended to a general isotropic elastic medium.

II. AVERAGE STRESS TENSOR IN AN ACOUSTIC BEAM

Radiation stress and radiation pressure in an elastic medium have been considered in numerous publications (e.g., Ref. 9). Here we give a particular expression for radiation stress components in the field of a low-angle acoustic beam. In the framework of the classical five-constant description, nonlinear stress in an isotropic solid is given by the known Piola-Kirchhoff pseudo tensor (e.g., Refs. 23 and 24):

$$\begin{aligned} \sigma_{ik} = & \mu \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) + \lambda \left(\frac{\partial u_l}{\partial x_l} \right) \delta_{ik} + \left(\mu + \frac{A}{4} \right) \left(\frac{\partial u_l}{\partial x_i} \frac{\partial u_l}{\partial x_k} \right. \\ & + \frac{\partial u_i}{\partial x_l} \frac{\partial u_k}{\partial x_l} + \frac{\partial u_i}{\partial x_l} \frac{\partial u_l}{\partial x_k} \left. \right) + \left(\frac{\lambda + B}{2} \right) \left[\left(\frac{\partial u_l}{\partial x_m} \right)^2 \delta_{ik} \right. \\ & + 2 \frac{\partial u_i}{\partial x_k} \frac{\partial u_l}{\partial x_l} \left. \right] + \frac{A}{4} \frac{\partial u_k}{\partial x_l} \frac{\partial u_l}{\partial x_i} + \frac{B}{2} \left(\frac{\partial u_l}{\partial x_m} \frac{\partial u_m}{\partial x_i} \delta_{ik} \right. \\ & \left. + 2 \frac{\partial u_k}{\partial x_i} \frac{\partial u_l}{\partial x_l} \right) + C \left(\frac{\partial u_l}{\partial x_l} \right)^2 \delta_{ik}. \end{aligned} \quad (1)$$

Here standard notations are used; namely, \mathbf{u} is the displacement vector, λ and μ are Lamé coefficients (linear moduli), and A, B, C are third-order Landau moduli. This tensor is, in general, not symmetric, but for a potential field, when $\partial u_i / \partial x_k = \partial u_k / \partial x_i$, it becomes symmetric. As usual, we represent the medium displacement in the standard form, $\mathbf{u} = \mathbf{v} + \mathbf{U}$, where \mathbf{v} is oscillating part, so that $\langle \mathbf{u} \rangle = \mathbf{U}$, where

angular brackets denote averaging over the period T of the high-frequency oscillations, i.e., $\langle \dots \rangle = T^{-1} \int_T \dots dt$.

In a general case, the radiation stress tensor, R_{ik} , can be defined as period average of Eq. (1):

$$R_{ik} = \langle \sigma_{ik}^N \rangle, \quad (2)$$

where σ_{ik}^N is nonlinear part of stress, i.e., the expression (1) less its first two, linear terms. As usual, nonlinear terms are supposed small as compared with the linear ones, which is practically always true. We limit our consideration by the case of a narrow-angle primary acoustic beam that is true for ultrasound focused system with small convergence angles. For a narrow angle beam the following inequalities are fulfilled:

$$\frac{\partial}{\partial x} \propto k \gg \frac{\partial}{\partial y}, \frac{\partial}{\partial z}; \quad u_x \gg u_y, u_z.$$

Under these conditions, from (1) it follows in the main-order approximation:

$$\begin{aligned} \langle \sigma_{xx} \rangle = G_1 \left\langle \left(\frac{\partial u_x}{\partial x} \right)^2 \right\rangle, \quad G_1 = \frac{3}{2} \lambda + 3\mu + A + 3B + C, \\ \langle \sigma_{yy} \rangle = \langle \sigma_{zz} \rangle = G_2 \left\langle \left(\frac{\partial u_x}{\partial x} \right)^2 \right\rangle, \quad G_2 = \frac{\lambda}{2} + B + C, \\ \langle \sigma_{xy} \rangle = \langle \sigma_{yx} \rangle = G_3 \left\langle \left(\frac{\partial u_x}{\partial x} \frac{\partial u_x}{\partial y} \right) \right\rangle, \\ G_3 = \lambda + 3\mu + A + 2B, \\ \langle \sigma_{xz} \rangle = \langle \sigma_{zx} \rangle = G_3 \left\langle \left(\frac{\partial u_x}{\partial x} \frac{\partial u_x}{\partial z} \right) \right\rangle. \end{aligned} \quad (3)$$

Note that nondiagonal components of this tensor are small as compared with the diagonal ones. However, as it will be seen below, they can give a comparable contribution in the shear displacements.

As shown below, the parameter

$$G_1 - G_2 = G_3 = \lambda + 3\mu + A + 2B = Q \quad (4)$$

determines the magnitude of shear radiation force.

III. EQUATIONS FOR DISPLACEMENTS INDUCED BY THE RADIATION FORCE

Now we obtain the equations for period-average motions in an inhomogeneous medium. The starting point is the standard equation for waves in an elastic medium (in Lagrangian coordinates):

$$\rho \ddot{u}_i = \frac{\partial \sigma_{ik}}{\partial x_k}. \quad (5)$$

The stress tensor is represented as a sum of linear and nonlinear parts, $\sigma_{ik} = \sigma_{ik}^L + \sigma_{ik}^N$, where

$$u_{ik}^L = 2\mu \left(u_{ik}^L + \frac{\sigma}{1-2\sigma} u_{ij}^L \delta_{ik} \right), \quad u_{ik}^L = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) \quad (6)$$

are the linear parts of stress and strain (u_{ik}) tensors, respectively, and σ_{ik}^N is a small nonlinear part; σ is Poisson ratio.

It is supposed that both second- and third-order moduli are, in general, dependent on coordinates and this dependence is slow as compared with the ultrasound wavelength. From (5) and (6), a modified nonlinear wave equation follows in the form

$$\rho[\ddot{\mathbf{u}} - c_l^2 \nabla^2 \mathbf{u} + (c_l^2 - c_t^2) \nabla(\nabla \cdot \mathbf{u})] - \mathbf{S} = \Phi, \quad (7)$$

where $c_l = \sqrt{(\lambda + 2\mu)/\rho}$ and $c_t = \sqrt{\mu/\rho}$ are the velocities of linear longitudinal and transverse waves, respectively, and the term

$$S_i = \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) \frac{\partial \mu}{\partial x_k} + \frac{\partial}{\partial x_i} \left(\frac{\sigma \mu}{1-2\sigma} \right) \nabla \cdot \mathbf{u} \quad (8)$$

reflects spatial variation of linear parameters.

The right-hand side of (7) is $\Phi = \partial \sigma_{ik}^N / \partial x_k$, and its period average defines the radiation force per unit volume acting on the medium:

$$\langle \Phi \rangle = \frac{\partial \langle \sigma_{ik}^N \rangle}{\partial x_k}. \quad (9)$$

In a linear, homogeneous material ($\mathbf{S}=0, \Phi=0$), Eq. (7) becomes the classical equation describing elastic wave propagation on solids:²³

$$\ddot{\mathbf{u}} - c_l^2 \nabla^2 \mathbf{u} + (c_l^2 - c_t^2) \nabla(\nabla \cdot \mathbf{u}) = 0. \quad (10)$$

In what follows we consider the case when the medium parameters depend on one coordinate x that is directed along the primary beam axis. In this case, the expression (8) reads

$$\frac{\mathbf{S}}{\rho} = \frac{\partial c_l^2}{\partial x} \left(\frac{\partial \mathbf{u}}{\partial x} + \nabla(u_x) \right) + \mathbf{x}_0 \frac{\partial}{\partial x} \left(\frac{\sigma \mu}{1-2\sigma} \right) \nabla \cdot \mathbf{u}. \quad (11)$$

Here \mathbf{x}_0 is unit vector in the x direction. Note that in the waterlike media, when $c_t^2 \ll c_l^2$, σ is close to $\frac{1}{2}$.

Now we average Eq. (7) over the period of high-frequency wave to obtain an equation for the average displacement, \mathbf{U} . In the corresponding equation,

$$\rho[\ddot{\mathbf{U}} - c_l^2 \nabla^2 \mathbf{U} + (c_l^2 - c_t^2) \nabla(\nabla \cdot \mathbf{U})] - \langle \mathbf{S} \rangle = \langle \Phi \rangle, \quad (12)$$

the right-hand side is determined by (9) where the tensor $\langle \sigma_{ik}^N \rangle$ is, in the particular case, given by (3).

For a homogeneous elastic medium the expression for $\langle \Phi \rangle$ was obtained before (e.g., Ref. 9).

Here, as usual for linear elastic waves (e.g., Ref. 23), one can represent \mathbf{U} as a sum of two vectors, a potential one, \mathbf{U}_1 so that $\nabla \times \mathbf{U}_1 = 0$, and a solenoidal, \mathbf{U}_2 , for which $(\nabla \cdot \mathbf{U}_2) = 0$. Similarly, we represent $\langle \mathbf{S} \rangle = \mathbf{S}_1 + \mathbf{S}_2$ and $\langle \Phi \rangle = \langle \Phi \rangle_1 + \langle \Phi \rangle_2$. Taking the divergence of Eq. (7) and keeping only the terms with no curl, we have

$$\frac{\partial^2 \mathbf{U}_1}{\partial t^2} - c_l^2 \nabla^2 \mathbf{U}_1 - \frac{\mathbf{S}_1}{\rho} = \frac{\langle \Phi \rangle_1}{\rho}. \quad (13)$$

Correspondingly, taking the curl of Eq. (7) yields

$$\frac{\partial^2 \mathbf{U}_2}{\partial t^2} - c_t^2 \nabla^2 \mathbf{U}_2 - \frac{\mathbf{S}_2}{\rho} = \frac{\langle \Phi \rangle_2}{\rho} = \mathbf{F}. \quad (14)$$

(Here \mathbf{F} can be interpreted as shear force per unit mass.)

Indeed, both $\nabla \times$ and $\nabla \cdot$ of Eqs. (13) and (14) give identities, which secures their validity (cf. Ref. 23). In general, separating the potential and solenoidal parts of the vector $\langle \Phi \rangle$ needs a solution of the equations for scalar and vector potentials. However, under the simple geometries considered below it can be done in a simplified way.

With an appropriate addition of linear dissipative terms in the left-hand sides, Eqs. (13) and (14) can be considered as general equations describing period-average motions produced by radiation force. In the limiting case of fluid when $\mu=0, A=0$, the corresponding expressions can be simplified by taking into account that, according to Ref. 25, in this case the elastic parameters become $\lambda = \rho c_l^2$, $B = -\lambda$, and $C = [\lambda - \rho^2 (dc_l^2/d\rho)]/2$, so that in Eq. (4), $Q = -\rho c_l^2$.

Various known specific cases follow from these equations. In particular, for a homogeneous fluid or a waterlike material such as a tissue, Eq. (13) for the potential part can be readily reduced to the well-known equation first obtained by Westervelt²⁶ for average (low-frequency) pressure, p , generated by a narrow acoustic beam in a parametric array:

$$\nabla^2 p - \frac{1}{c_l^2} \frac{\partial^2 p}{\partial t^2} = -\frac{\varepsilon}{\rho c_l^4} \frac{\partial^2 \langle p^2 \rangle}{\partial t^2}, \quad (15)$$

where $\varepsilon = (\gamma + 1)/2$ is the nonlinearity parameter (γ is adiabatic ratio). A similar equation for an elastic medium was obtained by Cantrell.⁹ However, as shown below, in all these cases the radiation force is potential so that the right-hand side of Eq. (14) vanishes.

Note that in recent papers (e.g., Ref. 27), a notion of “dynamic” radiation force is introduced as opposed to the “static” force created by continuous sound. The force in Eq. (15) that is classical in the literature on parametric arrays and self-demodulation is evidently “dynamic.”

Here we concentrate on the generation of shear displacements [Eq. (14)] that can highly exceed the potential displacements due to the fact that in biological tissues $c_t^2 \ll c_l^2$, i.e., the “shear resistance” is small [see the comment after Eq. (32) below].

IV. EQUATIONS FOR SHEAR DISPLACEMENT

A. Radiation force created by a narrow-angle beam

Let us return to Eq. (14) for shear displacement, which is of main interest here. We consider a quasi-longitudinal, quasi-harmonic wave. Then in the primary ultrasonic beam, the displacement can be presented in the form

$$\mathbf{u}_x = \frac{1}{2} V(x, z, t) \exp [i(kx - \omega t) + i\theta(x, z, t)] + \text{c. c.}, \quad (16)$$

where V and θ are slowly (at the scale of ultrasound period and wavelength) varying amplitude and phase; without loss of generality they can be considered real functions.

First we consider a simple 2-D case with plane geometry when a primary (collimated) acoustic beam propagates along

the x axis in the (x, z) plane. To find the shear motions, we determine the solenoidal part of the force in (12) by taking its curl:

$$\Psi = (\nabla \times \langle \Phi \rangle)_y = \left(\frac{\partial \langle \Phi_x \rangle}{\partial z} - \frac{\partial \langle \Phi_z \rangle}{\partial x} \right). \quad (17)$$

Evidently, in the case considered, the rotation is directed along the third axis, y , so that the resulting expression is essentially scalar.

To evaluate the shear forcing, Ψ , we use the above expressions (3) for the average nonlinear part of stress tensor. According to (2) and (3), components of the nonlinear force Φ are

$$\begin{aligned} \langle \Phi_x \rangle &= \frac{\partial R_{xx}}{\partial x} + \frac{\partial R_{xz}}{\partial z} = G_1(x) \frac{\partial}{\partial x} \left(\frac{\partial \bar{u}_x}{\partial x} \right)^2 \\ &+ G_3(x) \frac{\partial}{\partial z} \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_x}{\partial z} \right) + \frac{\partial G_1}{\partial x} \left(\frac{\partial \bar{u}_x}{\partial x} \right)^2, \end{aligned} \quad (18)$$

$$\begin{aligned} \langle \Phi_z \rangle &= \frac{\partial R_{zx}}{\partial x} + \frac{\partial R_{zz}}{\partial z} = G_3(x) \frac{\partial}{\partial x} \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_x}{\partial z} \right) \\ &+ G_2(x) \frac{\partial}{\partial z} \left(\frac{\partial \bar{u}_x}{\partial x} \right)^2 + \frac{\partial G_3}{\partial x} \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_{x1}}{\partial z} \right). \end{aligned} \quad (19)$$

Hence, in (17),

$$\begin{aligned} \Psi &= (G_1 - G_2) \frac{\partial^2}{\partial x \partial z} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \right)^2 \right\rangle + G_3 \left(\frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial x^2} \right) \\ &\times \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_x}{\partial z} \right) \right\rangle + \frac{\partial(G_1 - G_2)}{\partial x} \frac{\partial^2}{\partial z^2} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \right)^2 \right\rangle \\ &- 2 \frac{\partial G_3}{\partial x} \frac{\partial}{\partial x} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_x}{\partial z} \right) \right\rangle - \frac{\partial G_3}{\partial x^2} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_{x1}}{\partial z} \right) \right\rangle. \end{aligned} \quad (20)$$

As mentioned, we suppose that the primary field is a narrow-angle beam, so that in the average field $\partial/\partial z \gg \partial/\partial x$, hence (from $\nabla \cdot u = 0$), $u_x \gg u_z$. Thus, one can neglect the last two terms in (20) that are of second order in the x derivatives of slowly varying functions. Then, taking (4) into account, we have a simplified expression:

$$\begin{aligned} \Psi &\approx Q \frac{\partial}{\partial z} \left[\frac{\partial}{\partial x} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \right) \right\rangle^2 + \frac{\partial}{\partial z} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \frac{\partial \bar{u}_x}{\partial z} \right) \right\rangle \right] \\ &+ \frac{\partial Q}{\partial x} \frac{\partial}{\partial z} \left\langle \left(\frac{\partial \bar{u}_x}{\partial x} \right)^2 \right\rangle = \frac{\partial F_x}{\partial z}. \end{aligned} \quad (21)$$

Evidently, in this approximation the expression denoted here as F_x is the x component of the solenoidal force (provided $F_x = 0$ at $|z| \rightarrow \infty$).

In a similar way we evaluate the term S_2/ρ in Eq. (14). Taking the curl of the vector (11), we have

$$\begin{aligned} |\nabla \times \mathbf{S}/\rho| &= \frac{\partial c_l^2}{\partial x} \nabla^2 W - \frac{\partial^2 c_l^2}{\partial x^2} \left(\frac{\partial U}{\partial z} + \frac{\partial W}{\partial x} \right) \\ &\approx - \frac{\partial^2}{\partial z \partial x} \left(\frac{\partial c_l^2}{\partial x} U \right), \end{aligned} \quad (22)$$

and thus

$$\frac{S_2}{\rho} \approx - \frac{\partial}{\partial x} \left(\frac{\partial c_l^2}{\partial x} U \right). \quad (23)$$

In Eq. (22), U and W are the x and z components of the vector \mathbf{U}_2 , respectively. In transition to Eq. (23), the above equality $\nabla \cdot \mathbf{U}_2 = 0$ is used and the higher-order terms are neglected. Evidently, for a narrow-angle beam, S_2 is a second-order value, and in what follows it will be neglected.

Finally, we neglect the term $c_l^2 U_{xx}$ in the left-hand part of Eq. (14) as it is evidently small as compared with the term $U_{tt} \approx c_l^2 U_{xx}$. In a general formulation (when shear wave is not sufficiently localized in transverse direction) this term, as well as the ‘‘geometrical’’ term, Eq. (23), should be taken into account. As a result for the wave (16), Eq. (14) becomes

$$U_{tt} - c_l^2 U_{zz} = \frac{Q}{\rho} \left[\frac{\partial M_a^2}{\partial x} + \frac{1}{k} \frac{\partial}{\partial z} \left(M_a^2 \frac{\partial \theta}{\partial z} \right) \right] + M_a^2 \frac{1}{\rho} \frac{\partial Q}{\partial x}. \quad (24)$$

Here

$$M_a = \sqrt{\frac{\langle (\partial u_x / \partial t)^2 \rangle}{c_l^2}} = \frac{\omega |V|}{c_l} \quad (25)$$

is dimensionless acoustic velocity (Mach number) in the ultrasound.

Then in the first approximation

$$\left(\frac{\partial \bar{u}_1}{\partial x} \right)^2 = \frac{k^2 |V|^2}{2} \approx M_a^2, \quad \frac{\partial \bar{u}_1}{\partial x} \frac{\partial \bar{u}_1}{\partial z} \approx \frac{k |V|^2}{2} \frac{\partial \theta}{\partial z} = \frac{M_a^2}{k} \frac{\partial \theta}{\partial z}. \quad (26)$$

To evaluate the first term on the r.h.s. of Eq. (24), we use the fact that a linear narrow beam satisfies the classical parabolic equation for complex amplitude, $V = a \exp(i\theta)$ for a low-angle, 2-D wave beam in a linear medium^{28,29} and we consider media with small inhomogeneities that do not affect significantly the parameters of the beam:

$$i a_x + \frac{a_{zz}}{2k} = 0. \quad (27)$$

The imaginary part of this equation is

$$a_x + \frac{1}{2k} (a \theta_{zz} + 2a_z \theta_z) = 0. \quad (28)$$

Multiplying this by $k^2 a / 2$, we see that the left-hand part of this equation coincides with the term in the square brackets of Eq. (24). Hence, a linear, nondissipative beam in a homogeneous medium does not produce shear force. As shown in the Appendix, this statement is true in a general three-dimensional case, for a general case of a homogeneous elastic medium.

B. Dissipative and nondissipative forces

Dissipative effects can be readily incorporated into Eq. (24). It is important that, in general, there exist two dissipative factors. One is the attenuation of the primary beam, and another is the attenuation of the generated shear wave. The ultrasound wave attenuation leads to the replacement of M_a^2 by $M_{nd}^2 \exp(-2\alpha x)$, where α is the wave attenuation coefficient, and M_{nd} corresponds to the absence of dissipation. In this case, Eq. (24) reads

$$\frac{U_{tt}}{c_t^2} - U_{zz} = \frac{Q}{\rho c_t^2} M_{nd}^2 \exp(-2\alpha x) \left(\frac{1}{Q} \frac{\partial Q}{\partial x} - 2\alpha \right). \quad (29)$$

The expression for the dissipative part of radiation force was semi-empirically introduced in Ref. 11 from the evaluation of momentum loss. As mentioned, a general derivation of dissipative shear force in fluids was made by Eckart.⁵

In homogeneous media, Eq. (29) has the form

$$\frac{U_{tt}}{c_t^2} - U_{zz} = -\frac{2\alpha Q}{\rho c_t^2} M_{nd}^2 \exp(-2\alpha x). \quad (30)$$

Taking into account that, in waterlike media, $\mu \ll \lambda$ and $A \ll B$, we have from (4)

$$\frac{Q}{\rho c_t^2} = \frac{\lambda + 3\mu + A + 2B}{\mu} \approx -\frac{c_l^2}{c_t^2}. \quad (31)$$

Indeed, as mentioned, in the limit of a fluid and, approximately, in waterlike media such as biological tissues, $B \approx -\lambda = -\rho c_l^2$, so that

$$Q = -\rho c_l^2 + O(\mu, A), \quad (32)$$

and the absolute value of the ratio (31) is large. Due to this, the shear radiation force is typically much stronger than the potential one [besides, according to Eq. (15), the latter exists only in modulated beams]. Indeed, writing Eq. (15) in terms of average displacement, $U_l \approx pT/\rho c_l$, where T is a characteristic modulation time (pulse duration), it is easy to see that the ratio of the right-hand sides of the corresponding equation for U_l and of Eq. (30) is of order $(c_l/c_t)^2/(\varepsilon \alpha c_l T)$. For the waterlike media considered here, this ratio is small in all realistic cases. The same is true for the nondissipative force (33) considered below. Another estimate follows from the known expression for the cumulative growth of average displacement in a plane nonlinear (simple) wave: $U_l \approx (\varepsilon/4)M_a^2 x$. Letting $x=5$ cm be the length of the focal area and the value of $M_a \approx 10^{-3}$ in this area (as it is taken for estimates below), we obtain displacement of order 50 nm (which is typically too small from the viewpoint of, e.g., medical applications), whereas the shear displacements can reach at least 10 μm (see below).

Now we specifically consider the nondissipative shear force in inhomogeneous media defined by the expression

$$F = \frac{1}{\rho c_t^2} \frac{\partial Q}{\partial x} M_a^2 \approx -M_a^2 \left(\frac{c_l^2}{c_t^2} \right) \left(2 \frac{\partial c_l / \partial x}{c_l} + \frac{\partial \rho / \partial x}{\rho} \right). \quad (33)$$

This force is due to spatial variation of the parameter Q , which is a combination of linear and nonlinear moduli.

Among the features that distinguish the nondissipative radiation force from the dissipative one are that it is practically independent of ultrasound frequency and that its direction depends on the sign of the gradient of the parameter Q . Thus, the displacement produced by nondissipative force can be directed both from and to the transducer, depending on the direction of media parameters variations. For a localized inhomogeneity such as a lesion, the nondissipative force changes its direction along the acoustic beam and thus it tends to either compress or stretch the lesion.

As follows from Eq. (32), for water $Q \approx 2.25 \times 10^9$ Pa. For biological tissues, the value of the sound speed c_l differs in different organs typically by several dozens of m/s.²⁹ Hence, it is reasonable to suppose that in areas between organs or between a healthy tissue and a lesion, the value of Q can vary by 10–20%, which we shall take here for estimates. Note that, although the shear modulus μ is small in the media considered here, it reveals the strongest spatial variation,^{29,30} in such cases it can noticeably contribute to the effect. The spatial scale of variations for, e.g., a lesion, can vary from few millimeters to a few centimeters. For the applicability of the simplified Eq. (30) it is, strictly speaking, necessary that the variation is smooth so that its scale, L , exceeds the characteristic beam width, D .

Let us compare the nondissipative force proportional to Q_x/Q , with the viscous one, proportional to α . Their ratio is

$$N = \frac{F}{F_d} = \left| \frac{Q_x}{2\alpha Q} \right|. \quad (34)$$

In real tissues the attenuation coefficient α is typically proportional to the ultrasound frequency f . For example, typical ultrasound attenuation for liver²⁹ is about 0.5 dB/cm MHz so that the attenuation distance in cm is

$$\alpha^{-1} = 17.3/f, \quad (35)$$

where f is taken in MHz. Taking as an example the parameter variation scale of $L=0.5$ cm, we have $N \approx 3.46/f$. Thus, starting from at least 1 MHz down, the nondissipative effect prevails for a variation of Q at a scale of 1 cm and less.

Note that because, as mentioned, the nondissipative force can have any sign whereas the dissipative force is always directed to the ultrasound propagation direction, these two forces can either add up or compensate each other. Measurements of shear displacements produced by radiation force can be used for reconstruction of spatial distribution of density and sound speed in tissues.

V. CYLINDRICAL BEAM: SOME SOLUTIONS

Consider now a more practical case of a cylindrically symmetric beam, where all variables depend on x and the cylindrical radius r . In the absence of angular dependence, the shear force that contains only x derivative has the same form as in Eq. (29). Besides, to obtain realistic results, one has to take into account attenuation of the shear wave, which is known to be quite strong. Here we describe it by a constant shear viscosity, ν , which agrees with the Voigt model (e.g., Refs. 31 and 32), to obtain

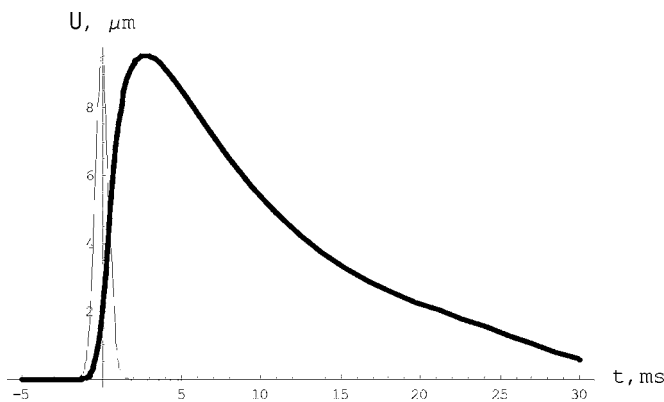


FIG. 1. Time dependence of shear displacement at the beam axis, $r=0$ (thick line). Thin line—primary ultrasonic impulse (arbitrary units).

$$U_{tt} - \left(c_t^2 + \nu \frac{\partial}{\partial t} \right) \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial U}{\partial r} \right) = \frac{\partial Q}{\partial x} \frac{1}{\rho} M_a^2 = F. \quad (36)$$

Consider a Gaussian primary beam. As long as the left-hand side of (36) is local with respect to the x coordinate, the problem is reduced to a local action of radiation force in any given cross-section of the beam. Hence, we represent the force in the form

$$F = \frac{\partial Q}{\partial x} \frac{1}{\rho} M_0^2 \exp \left(-\frac{r^2}{D^2} - \frac{t^2}{T^2} \right). \quad (37)$$

Here D is characteristic beam radius, T is characteristic pulse duration, and M_0 corresponds to the beam axis.

Let us estimate shear displacement variations in the beam focus in the case when the nondissipative force prevails. For estimates we consider a focused beam creating a pressure of 2 MPa in the focus, which corresponds to the Mach number, $M_0=10^{-3}$. As shown above, in tissues $Q \approx -\rho c_t^2$. In our example, $c_t=1500$ m/s, $c_l=2$ m/s, and the beam diameter at a half-intensity level is of about 0.4 cm, so that in (37), $D=0.17$ cm. The primary pulse duration at the half-intensity level is taken as 1 ms, so that $T=0.6$ ms. As to the factor dQ/dx , as mentioned, we suppose that $|Q|$ varies by a factor of 0.2 at a scale of 0.5 cm.

A least known factor is shear viscosity that is typically large so that the shear wave attenuates at distances comparable with its wavelength. The measured values of this parameter can differ drastically depending on frequency and

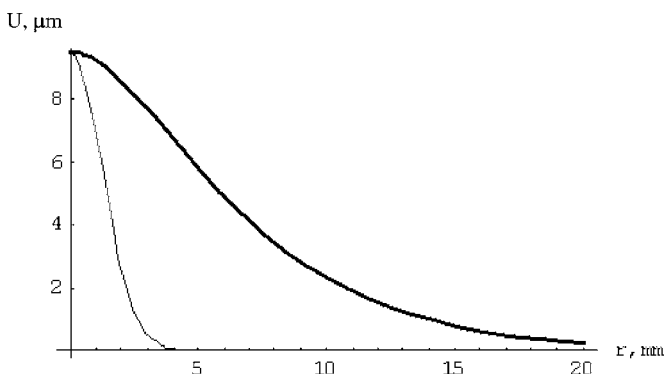


FIG. 2. Radial profile of displacement at the moment $t=2.7$ ms (thick line). Thin line corresponds to the primary ultrasonic beam (in arbitrary units).

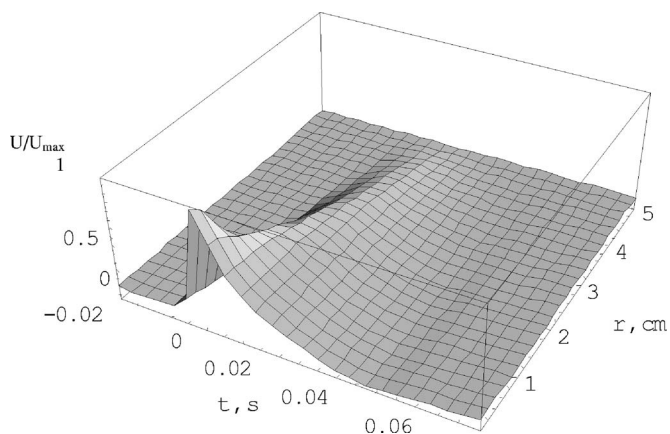


FIG. 3. 3-D plot of normalized displacement on the (t,r) plane (in this figure, the time is given in seconds).

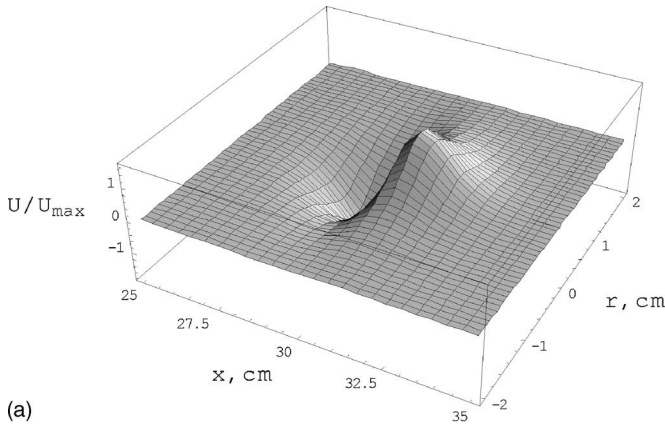
type of biological material. Most of the measurements refer to high ultrasonic frequencies (e.g., Ref. 31). In Ref. 32 the low-frequency measurements (100 Hz and lower) were performed for an arterial wall, and for the dynamic shear viscosity the value of $\rho\nu=15$ Pa·s (so that $\nu=0.015$ m²/s) was given that tends to increase with frequency in that range. In our estimates, for a millisecond-range pulse, we take the same value, $\nu=0.015$ m²/s (that is four orders larger than the molecular viscosity in “pure” water).

With these parameters, Eq. (36) was solved by using the Mathematica program with the initial conditions $U(t=0)=0$ and $U_t(t=0)=0$, and boundary conditions $U_r(r=0)=0$ and $U(r=5 \text{ cm})=0$. The results for shear displacement are shown in Figs. 1 and 2. It is seen that shear displacements can reach the values of order 10 μm for moderate ultrasound intensities. The scales of the perturbation in time and space are significantly larger than those in the primary beam. This is due to both elastic inertia and viscous relaxation: the damping scale of the generated shear wave (propagating almost radially from the primary beam) is typically comparable with the characteristic wave length.

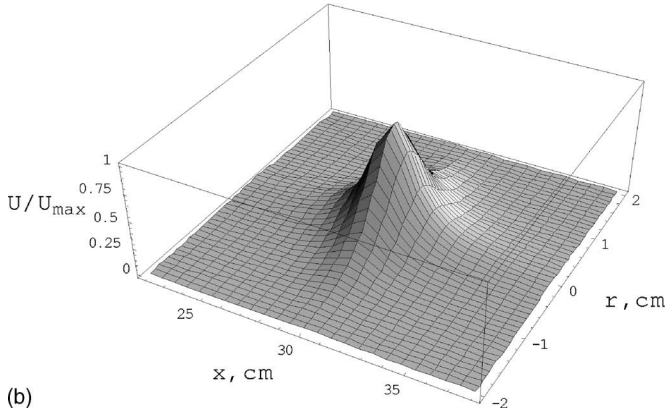
Three-dimensional Fig. 3 shows transverse propagation of the shear wave in the (t,r) plane. It is seen that the crest of the plot is sloped, which corresponds to the propagation at a velocity of about 2 m/s, in accordance with the shear wave velocity given above. After the main maximum, there are some oscillations, but in this case they are small due to large shear viscosity.

The distribution of shear displacement in a given cross section of the beam (Fig. 2) and its time behavior are in agreement with calculations (made for a homogeneous medium with losses) and experimental data available in the literature (e.g., Refs. 11 and 33). However, the distribution along the beam axis can differ drastically for a dissipative force in a homogeneous medium (where the x dependence is only due to the beam intensity variation) and a nondissipative force existing due to inhomogeneity. This is illustrated by Figs. 4(a) and 4(b) showing each of these effects.

As more detailed calculations show, the result strongly depends not only on the ultrasound intensity and viscosity



(a)



(b)

FIG. 4. (a) 3-D plot of normalized displacement distribution around the focus of a beam formed by an aperture of 4 cm, with a focal length of 30 cm, focused at the middle of a lesion about 10 cm long with hardness parameter $|Q|$ increased by 20% along the longitudinal coordinate x , in the horizontal (x, r) plane (positive and negative r denote the different sides from the axis in this plane). (b) The same for the pure dissipative effect ($Q=\text{const}$), with the same beam parameters.

but it also increases with the beam focal width and with pulse duration; the latter is in agreement with the experimental data given in Ref. 33.

The quantitative experimental data on shear wave generation in focused ultrasound seem too scarce yet to reliably separate dissipative and nondissipative effects. Still, the inhomogeneity may be responsible for a sharp enhancement of shear displacement in a lesion area that was reported in Ref. 34. It was expected the “classical” radiation force should lead to a decrease of shear displacement because of decreased ultrasonic attenuation and increased elasticity modulus of the lesion, whereas the experiment showed the opposite. The observed increase of shear displacement can be explained by gradients of elasticity in the lesion area leading to the nondissipative force considered here.

VI. CONCLUSIONS

Theoretical consideration presented above has demonstrated that radiation force generated by a linear potential primary wave in a homogeneous, nondissipative elastic medium is always potential and cannot induce shear displacement. It is shown, however, that such force can produce shear displacements in *inhomogeneous* fluids and waterlike media even without attenuation or reflection of primary ul-

trasound. Such a nondissipative force can be comparable with and even exceed the classical dissipative force, especially for a relative low-frequency (sub-MHz) ultrasound that can be of practical interest. The nondissipative radiation force can be directed both outward and toward the transducer, depending on the direction of a medium parameter gradient. It is recommended to perform the experiments aimed at the determining the frequency dependence of the effect and spatial distribution of displacements along the beam, which would allow us to evaluate the relative role of the effects of dissipation and inhomogeneity.

APPENDIX: THE IMPOSSIBILITY OF SHEAR DISPLACEMENT GENERATION BY A POTENTIAL ULTRASOUND BEAM

It was shown above for a low-angle ultrasonic beam that a linear, potential, nondissipative acoustic wave in a homogeneous medium can only create a potential radiation force that does not generate shear displacements. Here we show that this statement is true in a general case. Consider the radiation stress produced in an arbitrary isotropic elastic medium by a potential acoustic wave field so that the displacement $u_i = \partial\varphi/\partial x_i$, where potential φ oscillates with frequency ω and obeys the Helmholtz equation:

$$\nabla^2\varphi + \frac{\omega^2}{c_l^2}\varphi = 0. \quad (\text{A1})$$

Consider the radiation force vector (9):

$$\langle\Phi\rangle = \left\langle \frac{\partial\sigma_{ik}^N}{\partial x_k} \right\rangle. \quad (\text{A2})$$

As shown below, a potential acoustic field cannot generate shear waves in homogenous media, so that $\nabla_x\langle\Phi\rangle=0$, i.e.,

$$\frac{\partial\langle\Phi_i\rangle}{\partial x_k} = \frac{\partial\langle\Phi_k\rangle}{\partial x_i}. \quad (\text{A3})$$

Vector $\langle\Phi\rangle$ is contributed by the following terms in Eq. (1) for stress tensor:

- (1) Diagonal nonlinear terms of type $(\partial u_l/\partial x_l)^2\delta_{ik}$ give a contribution in Eq. (A2) in the form of $\langle\Phi^{(1)}\rangle = -\nabla f$, where f is a scalar. For these terms, Eq. (A3) is evidently true.
- (2) Terms of type $(\partial u_k/\partial x_k)(\partial u_l/\partial x_l)$. From (A1) it follows that

$$\frac{\partial u_l}{\partial x_l} = \frac{\partial^2\varphi}{\partial x_l^2} = -\frac{\omega^2}{c_l^2}\varphi. \quad (\text{A4})$$

Thus, for the corresponding component of $\langle\Phi\rangle$ we have

$$\begin{aligned} -\frac{\omega^2}{c_l^2}\langle\Phi_i^{(2)}\rangle &= \frac{\partial}{\partial x_k} \left(\varphi \frac{\partial^2\varphi}{\partial x_i \partial x_k} \right) = \left(\frac{\partial\varphi}{\partial x_k} \frac{\partial^2\varphi}{\partial x_i \partial x_k} \right. \\ &\quad \left. + \varphi \frac{\partial^3\varphi}{\partial x_i \partial x_k^2} \right) = \frac{1}{2} \left(\frac{\partial\varphi}{\partial x_k} \frac{\partial^2\varphi}{\partial x_i \partial x_k} - \frac{\omega^2\partial^2\varphi}{c_l^2\partial x_k} \right) \\ &= \frac{1}{2} \frac{\partial}{\partial x_i} \left[\left(\frac{\partial\varphi}{\partial x_k} \right)^2 - \frac{\omega^2}{c_l^2}\varphi^2 \right]. \end{aligned} \quad (\text{A5})$$

As the last expression is an x_i derivative of a scalar, and hence $\Phi_k^{(2)}$ is the x_k derivative of the same expression, Eq. (A3) is fulfilled for this part of the force.

(3) Finally, terms of type $(\partial u_i / \partial x_l)(\partial u_l / \partial x_i)$. The corresponding force component is

$$\begin{aligned} \langle \Phi_i^{(3)} \rangle &= \frac{\partial}{\partial x_k} \left(\frac{\partial^2 \varphi}{\partial x_l \partial x_i} \frac{\partial^2 \varphi}{\partial x_l \partial x_k} \right) = \frac{\partial^3 \varphi}{\partial x_l \partial x_i \partial x_k} \frac{\partial^2 \varphi}{\partial x_l \partial x_k} \\ &+ \frac{\partial^3 \varphi}{\partial x_l \partial x_k^2} \frac{\partial^2 \varphi}{\partial x_l \partial x_i} = \frac{1}{2} \frac{\partial}{\partial x_i} \left[\left(\frac{\partial^2 \varphi}{\partial x_l \partial x_k} \right)^2 \right. \\ &\left. - \frac{\omega^2}{c_l^2} \left(\frac{\partial \varphi}{\partial x_l} \right)^2 \right]. \end{aligned} \quad (\text{A6})$$

For the same reason as above, $\langle \Phi_i^{(3)} \rangle$ satisfies Eq. (A3). Two other terms in brackets in the term with a factor of $(\mu + A/4)$ in Eq. (1) are transformed similarly.

Hence, Eq. (A3) is true for the entire radiation force $\langle \Phi \rangle$. As a result, the period-average force created by a non-dissipative, linear primary field of an arbitrary form in a homogeneous elastic medium is always potential and does not produce shear displacements. On the contrary, as shown in this paper, in an inhomogeneous medium a nondissipative radiation force is possible and can be significant.

¹R. T. Beyer, "Lord Rayleigh and nonlinear acoustics," *J. Acoust. Soc. Am.* **98**, 3032–3034 (1995).

²R. T. Beyer, "Radiation pressure—the history of a mislabeled tensor," *J. Acoust. Soc. Am.* **63**, 1025–1030 (1978).

³W. L. Nyborg and J. A. Rooney, "Acoustic radiation pressure produced by a beam of sound," *J. Acoust. Soc. Am.* **72**, 1673–1687 (1982).

⁴R. W. Wood and A. L. Loomis, "The physical and biological effects of high-frequency sound-wave of great intensity," *Philos. Mag.* **4**, 417–436 (1927).

⁵C. Eckart, "Vortices and streams caused by sound waves," *Phys. Rev.* **73**, 68–76 (1948).

⁶P. J. Westervelt, "Parametric acoustic array," *J. Acoust. Soc. Am.* **35**, 535–537 (1963).

⁷T. Irvine, "Long Range Acoustic Device (LRAD)," *Vibration Newsletter*, 2–4, December 2005, http://www.vibrationdata.com/Newsletters/December2005_NL.pdf (Last time viewed 13 November 2006).

⁸G. P. Lee and T. G. Wang, "Acoustic radiation pressure," *J. Acoust. Soc. Am.* **94**, 1099–1109 (1993).

⁹J. Cantrell, "Acoustic-radiation stress in solids. I. Theory," *Phys. Rev. B* **30**, 3214–3220 (1984).

¹⁰O. V. Rudenko, A. P. Sarvazyan, and S. Y. Emelianov, "Acoustic radiation force and streaming induced by focused nonlinear ultrasound in a dissipative medium," *J. Acoust. Soc. Am.* **99**, 2791–2798 (1996).

¹¹A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear Wave Elasticity Imaging—a new ultrasonic technology of medical diagnostics," *Ultrasound Med. Biol.* **24**, 1419–1435 (1998).

¹²E. A. Barannik, S. A. Girnyk, V. V. Tovstiyak, A. I. Marusenko, V. A. Volokhov, A. P. Sarvazyan, and S. Y. Emelianov, "The influence of viscosity on the shear strain remotely induced by focused ultrasound in viscoelastic media," *J. Acoust. Soc. Am.* **115**, 2358–2364 (2004).

¹³K. Nightingale, M. Palmeri, and G. Trahey, "Analysis of contrast in images generated with transient acoustic radiation force," *Ultrasound Med.*

Biol. **32**, 61–72 (2006).

¹⁴G. F. Pinton, S. A. McAleavey, J. J. Dahl, K. R. Nightingale, and G. E. Trahey, "Real-time acoustic radiation force impulse imaging," *Proc. SPIE* **5750**, 226–235 (2005).

¹⁵G. E. Trahey, M. L. Palmeri, R. C. Bentley, and K. R. Nightingale, "Acoustic radiation force impulse imaging of the mechanical properties of arteries: in vivo and ex vivo results," *Ultrasound Med. Biol.* **30**, 1163–1171 (2004).

¹⁶J. 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).

¹⁷J. Bercoff, M. Pernot, M. Tanter, and M. Fink, "Monitoring thermally-induced lesions with supersonic shear imaging," *Ultrason. Imaging* **26**, 71–84 (2004).

¹⁸A. Alizad, L. E. Wold, J. F. Greenleaf, and M. Fatemi, "Imaging mass lesions by vibro-acoustography: modeling and experiments," *IEEE Trans. Med. Imaging* **23**, 1087–1093 (2004).

¹⁹A. Alizad, D. H. Whaley, J. F. Greenleaf, and M. Fatemi, "Potential applications of vibro-acoustography in breast imaging," *Technol. Cancer Res. Treat.* **4**, 151–158 (2005).

²⁰O. D. Kripfgans, K. J. Haworth, D. D. Steele, and S. D. Swanson, "Magnetic resonance elastography using time reversed acoustics," *Proc. SPIE* **5746**, 323–332 (2005).

²¹T. J. Hamilton, C. J. Bailat, C. Rose-Petruck, and G. J. Diebold, "Acoustically modulated x-ray phase contrast imaging," *Phys. Med. Biol.* **49**, 4985–4996 (2004).

²²C. J. Bailat, T. J. Hamilton, C. Rose-Petruck, and G. J. Diebold, "Acoustic radiation pressure: A "phase contrast" agent for x-ray phase contrast imaging," *Appl. Phys. Lett.* **85**, 4517–4519 (2004).

²³L. D. Landau and E. M. Lifschitz, *Theory of Elasticity* (Pergamon, New York, 1986).

²⁴R. N. Thurston, "Wave propagation in fluids and normal solids," in *Physical Acoustics*, edited by W. P. Mason (Academic Press, New York, 1964), Vol. 1, Part A, pp. 1–110.

²⁵M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "Separation of compressibility and shear deformation in the elastic energy density," *J. Acoust. Soc. Am.* **116**, 41–44 (2004).

²⁶H. O. Berkta, "Possible exploitation of non-linear acoustics in underwater transmitting applications," *J. Sound Vib.* **2**, 435–461 (1965).

²⁷G. Silva, "Dynamic ultrasound radiation force of acoustic waves," *Phys. Rev. E* **71**, 056617 (2005).

²⁸L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).

²⁹A. Sarvazyan, "Elastic properties of soft tissue," in *Handbook of Elastic Properties of Solids, Liquids and Gases. Volume III: Elastic Properties of Solids: Biological and Organic Materials* (Academic, New York, 2001), pp. 107–127.

³⁰E. L. Madsen, H. J. Sathoff, and J. A. Zagrebski, "Ultrasonic shear wave properties of soft tissues and tissue-like materials," *J. Acoust. Soc. Am.* **74**, 1346–1355 (1983).

³¹K. Sunagawa and H. Kanai, "Measurement of shear wave propagation and investigation of estimation of shear viscoelasticity for tissue characterization of the arterial wall," *J. Med. Ultrason.* **32**, 39–47 (2005).

³²S. Calle, J.-P. Remenieras, O. B. Matar, M. E. Hachemi, and F. Patat, "Temporal analysis of tissue displacement induced by a transient ultrasound radiation force," *J. Acoust. Soc. Am.* **118**, 2829–2840 (2005).

³³V. G. Andreev, V. N. Dmitriev, Yu. A. Pishchal'nikov, O. V. Rudenko, O. A. Sapozhnikov, and A. P. Sarvazyan, "Observation of shear waves excited by focused ultrasound in a rubber-like medium," *Acoust. Phys.* **43**, 123–128 (1997).

³⁴A. Shrestha, B. Otteson, H. Yao, and E. Ebbini, "Active ultrasonic imaging method for lesion visualization," *J. Acoust. Soc. Am.* **118**(3), 1183 (2005).

Noncontact transportation in water using ultrasonic traveling waves

Shinfuku Nomura^{a)}

Department of Mechanical Engineering, Ehime University, 3 Bunkyo-cho Matsuyama Ehime 790-8577, Japan

Thomas J. Matula

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

Jun Satonobu

Department of Intelligent Machine Engineering, Hiroshima Institute of Technology, 2-1-1 Miyake, Saeki-ku, Hiroshima 731-5193, Japan

Lawrence A. Crum

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

(Received 19 March 2006; revised 7 September 2006; accepted 3 January 2007)

A noncontact transport experiment in water using ultrasonic traveling waves was investigated. Acrylic, aluminum, and brass discs were used as test objects. Traveling waves were generated using two ultrasonic transducers attached at the ends of a vibrating plate. One side was used as the wave-source side and the other side was used as the wave-receiving side. Acrylic plates cemented to the sides of the vibrating plate formed a tank to hold water. Object transportation was accomplished by adding a small amount of water to the vibrating structure. The transport velocity of floating objects in water is faster than for floating transport in air because of buoyancy. The transport velocity of an object depends on water height. The minimum value of the velocity occurs when the disc thickness is equal to the water height. The transport velocity increases as the height of water increases. For very shallow depths, the largest velocity is obtained when cavitation-induced streaming occurs. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436708]

PACS number(s): 43.25.Uv, 43.25.Yw, 43.25.Ts, 43.25.Nm, 43.25.Qp [MFH] Pages: 1332–1336

I. INTRODUCTION

Noncontact manipulation of particles,^{1,2} droplets,^{3,4} and bubbles^{5,6} using ultrasonic energy as a levitation force is an active area of investigation. It is possible to use near-field acoustic waves (from several 10 to several 100 μm) to levitate a planar object.^{7–9} Furthermore, by exciting flexural vibration mode traveling waves, it is possible to provide noncontact transport of a planar object.^{10–12} If multiple transducers are used, it should be possible to transport large items long distances.¹³

We are currently investigating this technology for the postwashing processes used in the manufacture of such components as wafers and glass substrates for liquid crystal displays, thereby preventing mechanical contact or recontamination by airborne impurities. Recently, an in-liquid plasma chemical vapor deposition process has been proposed for the high-speed synthesis of material substrates.^{14,15} Such developments enhance the desire for realizing a precise, noncontact transport technology for use in liquids. In addition to differences in the density and the sound velocity of liquids compared to air, buoyancy effects suggest that liquids can be used for speedy transport of heavier objects. However,

acoustic standing waves must be minimized so that transport momentum is maintained. Furthermore, cavitation must be considered, since it may induce drag, or cause damage to the object.

The results of our studies presented in this paper show how traveling waves can be maintained. We also report initial studies related to transport velocity as a function of water height.

II. EXPERIMENTAL APPARATUS AND METHOD

Figure 1 shows a schematic diagram of the experimental apparatus. There are two transducers affixed to either end of the vibrating plate; one is used as a source (driving side), the other is an absorber (receiving side). These Langevin transducers having a nominal frequency of 27 kHz were used along with a stepped horn to amplify the vibration. As traveling waves are generated from the driving transducer, a disc propagates without contact in the direction of the traveling waves.

The vibrating plate is made from stainless steel, measures 50 × 250 mm and is 3 mm thick. The receiving transducer is connected to a parallel resonant circuit comprised of coils and resistors to absorb vibrational energy and convert it into electrical energy. Except for the size of the vibrating plate, the basic design is nearly the same as our device used

^{a)}Electronic mail: shin_nomu@eng.ehime-u.ac.jp

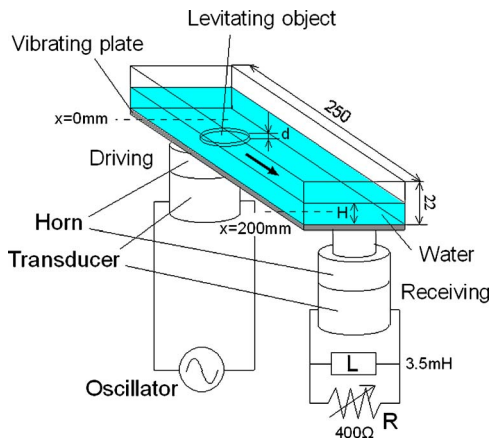


FIG. 1. (Color online) Experimental design. H is the water height and d is the disc thickness. The transducers were affixed to the vibrating plate by flat head bolts. The receiving transducer is connected to a matching load consisting of a resistor and inductor.

for transporting objects in air.^{10,13} The distance between the driving transducer and the receiving transducer is 200 mm. The vibrating plate has acrylic plate walls 22 mm high and 1 mm thick to hold the water. The input power of the amplifier was fixed while the frequency of the oscillator was adjusted so that maximum power was input to the transducer. For air-based comparison studies, the resonant frequency was 27.2 kHz. This was adjusted to 27.3 kHz for experiments in water where the height of the water was in a range $H=0-20$ mm. The velocity of transported objects was calculated using still images from a high-speed video camera (200 fps). Acrylic discs with 30 mm diameters, and thicknesses ranging from 2 to 10 mm, an aluminum disc, and a brass disc were used as floating planar objects (disc parameters are shown in Table I). Note that the diameters of the discs are greater than the half-wavelength of the ultrasound.

The input electrical power to the transducers varied from 10 to 30 W. A laser Doppler vibroscope (LDV) mounted above the vibrating plate was used to measure the particle velocity along the center axis of the vibrating plate.

III. RESULTS OF THE EXPERIMENT AND OBSERVATIONS

A. Oscillating velocity and phase of vibrating plate

Figure 2 shows the measured particle velocity of the vibrating plate when the water height is 16 mm. In the LDV measurement, the lock-in amplifier detects only the driving

TABLE I. Physical parameters for levitation objects.

		Diam (mm)	Thick (mm)	Mass (g)	Unit mass (kg/m ²)
Acrylic	A 1	30	2	1.6	2.2
	A 2	30	4	3.3	4.6
	A 3	30	8	6.6	9.3
	A 4	30	10	8.3	11.7
Aluminum	M 1	30	4	7.6	10.8
Brass	M 2	43.5	1.5	19.1	12.9

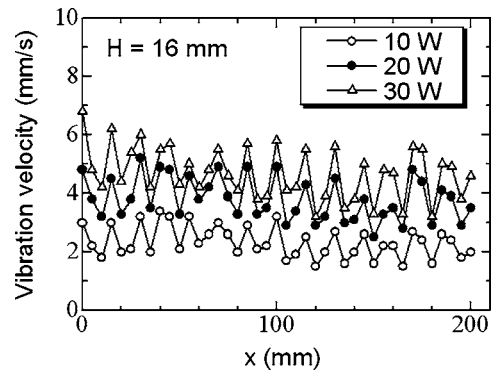


FIG. 2. Measured vibration distribution of traveling wave on the vibrating plate for $H=16$ mm.

frequency components. The position of the LDV changed little even if the height of water was changed; thus, the particle velocity on vibration plate was measured, not the velocity on the surface of water. There is little experimental error by water motion.

The particle velocity increased in proportion to an increase in output power. There is a slight reduction in particle velocity from the driving side to the receiving side. In addition, there is a periodicity of approximately 18–20 mm. Figure 3 shows the measured particle velocity when a water height is changed by 2 mm increments. Once water is introduced into the apparatus, the vibration velocity is reduced in comparison to vibration in air ($H=0$ mm). The particle velocity gradually changed with the height of the water. The difference of particle velocity with water height was within 20%.

Direct evidence of the traveling wave characteristics of our transporter is shown in Fig. 4. The changes in phase of the traveling wave for air and for water are plotted in order to determine the extent of traveling wave excitation on the vibrating plate. The traveling waves pass from left to right in Fig. 4. Since the traveling waves on the vibrating plate advance to the right and can be expressed as $y=e^{-(kx-\omega t)}$, the phase of the traveling waves can be plotted as the negative inclination. The phase changes from 180° to -180° linearly,

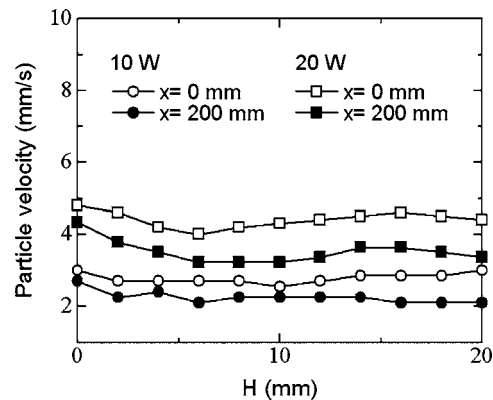


FIG. 3. Particle velocity as a function of water height. The series of open symbols are the measurement of driving point ($x=0$ mm) and the series of closed symbols are that of receiving point ($x=200$ mm). The circles are for 10 W electrical power; the squares, 20 W electrical power.

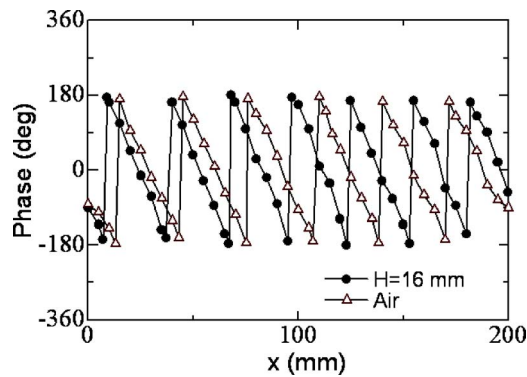


FIG. 4. (Color online) Phase distribution of traveling waves generated on the vibrating plate. The triangle is the phase in air, and the circle is the phase in water when $H=16$ mm.

which exhibits the characteristic of traveling waves, so we can conclude that excitation of traveling waves takes place both in water and air.

B. Measurement of transport velocity

The center axis for the section from $x=40$ mm to $x=160$ mm on the vibrating plate was divided into 9 increments of 15 mm each. Figure 5 shows the average velocity of the transported disc as calculated in each of these increments. The transport velocity is obtained by taking the average value of approximately 10 measurements. Data corresponding to when the disc came in contact with the sidewalls were not included in the calculation. In air, a steady-state transport velocity, or terminal velocity, was reached after approximately 90 mm. For liquid transport, steady-state velocity was reached earlier. In addition, a reduction in velocity was observed as the disc approached the receiving side. The maximum disc velocities were calculated using the average velocity as the disc passed through the 70–100 mm sections for water transport and the 100–130 mm sections for air transport.

The transport velocity of the disc in the present experiment is decreased in comparison with that in experiments with a large space between the acrylic wall and disc. It is possible that disc transport is reduced because of the interaction of ultrasound with the walls. However, the diameter of the discs in the experiment is greater than the half-

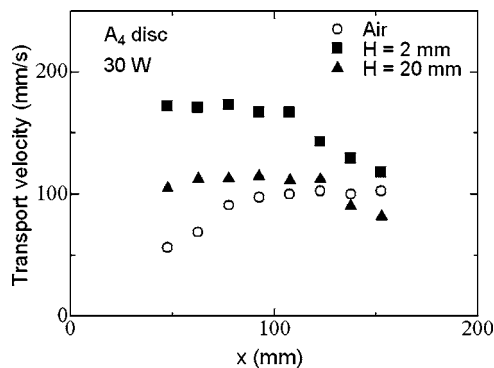


FIG. 5. Measurement of the transport velocity. The center axis of the vibrating plate is divided into 15 mm segments from 40 to 160 mm and the average velocity of the disc at each of these sections is calculated.

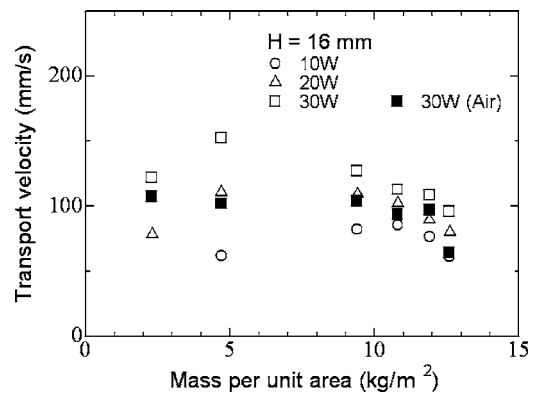


FIG. 6. The results of a transport experiment in water. The series of open symbols are measurements in water when the height of the water (H) is 16 mm, and the closed symbol is that in air at 30 W.

wavelength; the influence of the walls should be relatively equal. There are no large qualitative differences in the experimental results because the disc is transported by the component of ultrasonic traveling wave along the plate.

Figure 6 shows the maximum transport velocity as a function of mass per unit area of the levitated object when the water height $H=16$ mm. The results qualitatively agree with previous results in which the transport velocity is proportional to the increase of vibrating amplitude.¹⁰ This is because an increase in ultrasound output implies an increase in the amplitude of the vibrating plate. The results are given in terms of the maximum velocity, and there is a region in which the velocity of a heavy-mass disc is faster than that of a light disc. The reason may be due to the disc thickness. The reflected pressure amplitude from a thin disc may be significantly lower than for the thick disc. For the thicker disc, the sound pressure amplitude may be enough to generate cavitation. The occurrence of cavitation bubbles can give rise to streaming, which may be the same direction of traveling wave. The streaming induced by cavitation bubble influences the transport object, and the amplitude of sound pressure is thus indirectly a concern on the transport object.

For comparison purposes, we also plot the transport velocity of discs in air at $P=30$ W. The transport velocity in water is faster than the transport velocity in air at the same power, in some cases, 50%. However, the buoyancy of discs is not considered in Fig. 6. In Fig. 7, we replot the data by subtracting the buoyancy mass from that of the disc. The transport velocity has the maximum value around 1–1.5 kg/m^2 ; this value has a tendency to decrease as the mass per unit area increases.

C. Effect of water height

The water height can dramatically change the transport velocity in liquid; this is shown in Fig. 8. Note that when $H/d > 1$, the disc is completely submerged in water, and when $H/d < 1$, part of the disc is not submerged. The maximum velocity is obtained at the smallest H/d value. The transport velocity decreases with increasing H/d , reaching a minimum when the disc thickness is the same as the water height. This is because the fluid resistance increases with surface area, thus, the transport velocity near $H/d=1$ results

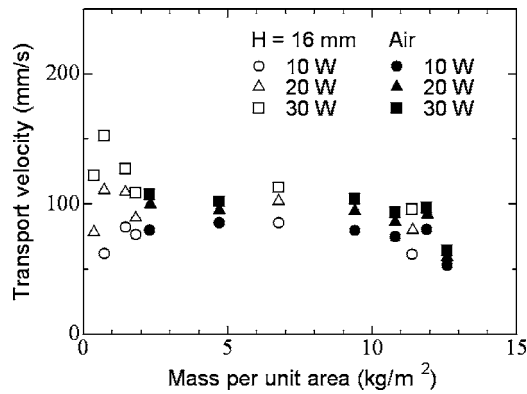


FIG. 7. The relationship between transport velocity and mass per unit area taking into account buoyancy. Open symbols correspond to measurements in water when the water height (H) is 16 mm, and the series of closed symbols is that in air.

in the minimum velocity. At larger H/d values, the transport velocity increases until saturation is reached at the transport velocity in water as shown in Fig. 6. It means that the effect of drag on the discs is negligible under water. The maximum transport velocity of 170 mm/s with 10 mm thickness acrylic disc (A_4) was attained when $H/d=0.2$. The smallest water depth in the present experiment was 2.0 mm. As the mass per unit area of the levitating object becomes heavier, the levitation gap becomes proportionally smaller as a function of weight per unit area.^{9,12} When the space between the disc and the vibrating plate is sufficiently narrow, that gap changes sinusoidally because of ultrasonic vibration, and implies that the squeeze effect creates a lifting force that levitates the object.^{11,16} Since the viscosity coefficient of water is larger than that of air, it is possible that the squeeze effect that originates from the nonlinearity of viscous forces associated with the liquid also takes part in levitation. The acoustic radiation force and squeeze effect may be closely related to the noncontact transportation of objects, but further work is needed to understand this better.

In particular, in liquid film transportation where the liquid height is much less than the thickness of the disc, a large amplitude sound pressure is generated and cavitation bubbles are created between the disc and the vibration plate. Therefore, in the cases of both A_2 and A_4 , if it is possible to make

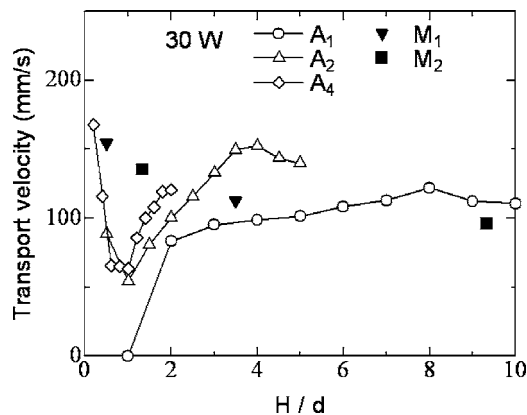


FIG. 8. The effect of water height on transport velocity. Here H is the height of the water and d is the thickness of the acrylic disc.

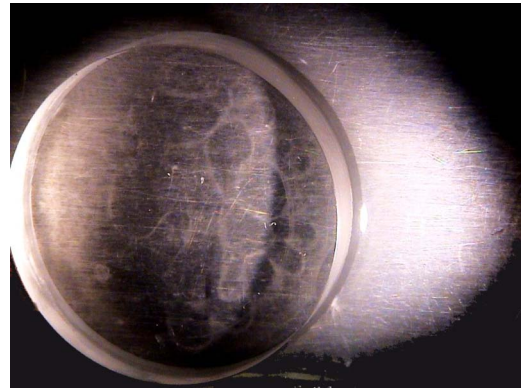


FIG. 9. (Color online) Cavitation bubbles created in the gap between the disc and the vibrating plate. There are several gaps, associated with movement of cavitation bubbles. The cavitation bubbles appear to move from the right side (driving side) to the left side (receiving side) of the photograph. The bright light is from the source set for photographing cavitation bubbles.

the liquid film thinner, the maximum transport velocity will be obtained at $H/D < 1$. Although the effect of buoyancy is smaller when generating cavitation bubbles, the transport velocity increases as H/D becomes smaller than unity. The difference in transport velocity due to buoyancy cannot be confirmed when cavitation appears.

Figure 9 is a top-view photograph of the acrylic disc during transit. This allows us to confirm the cluster streaming of acoustic cavitation that extends in the travel direction of the disc (from the driving side to the receiving side). The creation of acoustic cavitation accelerates the fluid flow, creating a large velocity gradient in the levitating gap between the vibrating plate and the object. Until now, cavitation streaming has been studied in standing wave fields.^{17,18} It was reported that the cluster streaming velocity of acoustic cavitation ranges from several cm/s to a few m/s in a standing wave field.^{18,19}

In the future, it will be necessary to better clarify streaming in such a narrow gap. It is also possible to consider that the viscous force in the horizontal direction restrains the object being transported.²⁰ In addition, it is speculated that during liquid film transport, the creation of bubbles can lead to greater velocity, because the disc experiences lower fluid resistance. It is suggested that cavitation bubbles are generated between the vibration plate and the disc though no levitation distances from vibrating plate are being measured in the actual experiment.

Figure 10 shows a comparison of velocities at different water heights ($H=2$ mm, $H=16$ mm, and in air) for $P=30$ W. The highest velocity was obtained at $H=2$ mm. The transport velocities of $H=2$ mm are about 1.5 times greater than those of air when mass per unit area is near 10 kg/m².

Figure 11 is a photograph of the condition of the water on the vibrating plate during the output of 30 W in a water depth of 2 mm. The surface of the water is pushed up by acoustic radiation pressure when the water is this shallow. As a result, a bumpy striped pattern appears and the disc becomes trapped, making it difficult for a thin disc to be transported smoothly. For example, when using the A_1 disc, the standing waves create an undulating pattern that cause the disc to stop midway during transport. With the thin disc (see

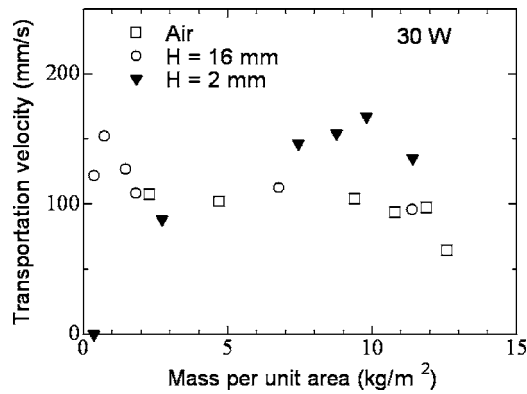


FIG. 10. Comparison of transport velocity.

A_1 in Fig. 8), the velocity becomes zero when near $H/d=1$. As was presented in the previous section, the mass of the surface has to be sufficiently large in order to stably transport it in liquids. In summary, various factors influence object transportation in water, including water depth, streaming induced by cavitation bubbles, the amplitude of the sound pressure between the disc and vibration plate, and the shape of the water surface caused by radiation pressure of ultrasound.

We also confirmed that during liquid film transfer, the thin disc attached itself to the vibrating plate and stopped. This suction force is reported on in Ref. 21. There is a need to carefully examine the relationship between the occurrence of acoustic cavitation and the suction force.

IV. CONCLUSIONS

We performed initial basic research into the application of noncontact transport technology using traveling-wave vi-



FIG. 11. (Color online) The undulating pattern formed by the vibrating plate.

bration in water. We confirmed the levitated transport of a disc in a liquid. The transport velocity in water is faster than that of air. The transport velocity changed as the ratio of the height of water (H) to the disc thickness (D) increases. We showed that the transport velocity at $H/d=0.2$, which is the smallest value used in the present experiment, is the maximum value. The minimum value is near $H/d=1$. After that, the transport velocity increases as H/d becomes larger, and finally, when the disc is completely submerged, it approaches a steady-state velocity.

- ¹T. Kozuka, T. Tuziuti, H. Mitome, and T. Fukuda, "Control of a standing wave field using a line-focused transducer for two-dimensional manipulation of particles," *Jpn. J. Appl. Phys., Part 1* **37**, 2974–2978 (1998).
- ²M. Takeuchi and K. Yamanouchi, "Ultrasonic micromanipulation of a small particle in Liquid," *Jpn. J. Appl. Phys., Part 1* **33**, 3045–3047 (1994).
- ³T. Otsuka and T. Nakane, "Ultrasonic levitation for liquid droplet," *Jpn. J. Appl. Phys., Part 1* **41**, 3259–3260 (2002).
- ⁴Z. C. Feng and Y. H. Su, "Numerical simulations of the translational and shape oscillations of a liquid drop in an acoustic field," *Phys. Fluids* **9**, 519–529 (1997).
- ⁵S. Nomura and K. Nishida, "Numerical simulation of a single bubble rising in an ultrasonic standing wave field," *Jpn. J. Appl. Phys., Part 1* **42**, 2975–2980 (2003).
- ⁶J. Halzfuss and M. Ruggeberg, "Micromanipulation of sono luminescing bubbles," *Phys. Rev. E* **69**, 056304-1–8 (2004).
- ⁷Y. Hashimoto, Y. Koike, and S. Ueha, "Acoustic levitation of planar objects using a longitudinal vibration mode," *J. Acoust. Soc. Jpn. (E)* **16**, 189–191 (1995).
- ⁸H. Nomura and T. Kamakura, "Theoretical and experimental examination of near-field acoustic levitation," *J. Acoust. Soc. Am.* **111**, 1578–1583 (2002).
- ⁹Y. Hashimoto, Y. Koike, and S. Ueha, "Near-field acoustic levitation of planar specimens using flexural vibration," *J. Acoust. Soc. Am.* **100**, 2057–2061 (1996).
- ¹⁰Y. Hashimoto, Y. Koike, and S. Ueha, "Transporting objects without contact using flexural traveling waves," *J. Acoust. Soc. Am.* **103**, 3230–3233 (1998).
- ¹¹A. Minikes and I. Bucher, "Noncontacting lateral transportation using gas squeeze film generated by flexural traveling waves-Numerical analysis," *J. Acoust. Soc. Am.* **113**, 2464–2473 (2003).
- ¹²S. Ueha, Y. Hashimoto, and Y. Koike, "Non-contact transportation using near-field acoustic levitation," *Ultrasonics* **38**, 26–32 (2000).
- ¹³Y. Hashimoto, Y. Koike, and S. Ueha, "Magnification of transportation range using non-contact acoustic levitation by connecting vibration plates," *Jpn. J. Appl. Phys., Part 1* **36**, 3140–3145 (1997).
- ¹⁴S. Nomura and H. Toyota, "Sonoplasma generated by a combination of ultrasonic waves and microwave irradiation," *Appl. Phys. Lett.* **83**, 4503–4505 (2003).
- ¹⁵S. Mukasa, S. Nomura, and H. Toyota, "Measurement of temperature in sonoplasma," *Jpn. J. Appl. Phys., Part 1* **43**, 2833–2837 (2004).
- ¹⁶K. Hahiba, K. Terao, and T. Kunoh, "A generalization of squeeze film theory and acoustic radiation theory (second report, inertia effects and relationship between squeeze film pressure and Langevin radiation pressure)," *Trans. JSME* **64**, 2455–2461 (1998) (in Japanese).
- ¹⁷S. Nomura, S. Mukasa, M. Kuroiwa, Y. Okada, and K. Murakami, "Cavitation bubble streaming in ultrasonic-standing-wave field," *Jpn. J. Appl. Phys., Part 1* **44**, 3161–3164 (2005).
- ¹⁸S. Nomura, K. Murakami, and Y. Sasaki, "Streaming induced by ultrasonic vibration in a water vessel," *Jpn. J. Appl. Phys., Part 1* **39**, 3636–3640 (2000).
- ¹⁹T. Kubo, M. Kuwabara, and Jian Yang, "Visualization of acoustically induced cavitation bubbles and microjets with the aid of a high-speed camera," *Jpn. J. Appl. Phys., Part 1* **44**, 4647–4652 (2005).
- ²⁰J. Hu, K. Nakamura, and S. Ueha, "Stability analysis of an acoustically levitated disc," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50**, 117–127 (2003).
- ²¹T. Hatanaka, Y. Koike, K. Nakayama, S. Ueha, and Y. Hashimoto, "Characteristics underwater near-field acoustic radiation force acting on a planar object," *Jpn. J. Appl. Phys., Part 2* **38**, L1284–L285 (1999).

Experiments on sound generation in corrugated pipes with flow

Ulf R. Kristiansen^{a)} and Geir A. Wiik^{b)}

Acoustic Research Center, Norwegian University of Science and Technology, O.S. Bragstads plass 2b, 7491 Trondheim, Norway

(Received 19 June 2006; revised 15 December 2006; accepted 15 December 2006)

The article reports acoustic measurements on short corrugated pipes with flow. Such pipes might generate high sound levels associated with length resonances. One of the main objectives of the study was to estimate the location of the effective sources by studying the energy flow through the pipes. It was found that a short section of corrugations will only produce sound effectively when placed at the inflow end, while for fully corrugated pipes, the sound-producing regions are located around the pressure maxima of the observed standing waves. It was further found that the net energy flow is in the upstream direction for nearly the complete length of pipe. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434241]

PACS number(s): 43.28.Ra, 43.20.Mv [GCL]

Pages: 1337–1344

I. INTRODUCTION

Air flow through corrugated pipes might produce high sound levels at frequencies associated with the length resonances of the pipes. A well-known example is the musical toy, consisting of an approximately 1 m long and a few centimeter diameter corrugated pipe, which when whirled in the air generates a humming sound. Other examples are found in industrial applications, where for instance “singing risers,” or noise problems connected to the corrugated pipes used in the natural gas industry is an important one.

The flow acoustic interaction represents an intriguing source of sound, which to some extent has been discussed in the literature. Probably the first reference to the subject is a short note by Burstyn, who discussed sound production in corrugated pipes at the meeting of the German Society for Technical Physics in June 1921.¹ The following year, Cermak presented an experimental study of the phenomenon.² Cermak experimented with 0.5-m-long tubes of narrow (10 mm outer diameter) bore, and reported that the fundamental tone was difficult to excite, but when extracted as the difference between consecutive higher partials, it was found to be lower than the length resonance frequencies expected from smooth pipe geometrical considerations. From his investigations, he ruled out that end conditions could be the reason for this, thereby implying a reduction in the effective wave speed. The jump in tones, that is that a tone will remain constant over a flow velocity interval, and then jump to the next harmonic once a certain velocity is exceeded, was also remarked upon. As to the generating mechanism, Burstyn suggested that the corrugations act as “numerous lips,” while Cermak calculates a “bump frequency” based on flow velocity and corrugation pitch and finds this to be close to the generated tones. The sound generation as basically a flow acoustic interaction was hence assumed already in the early studies.

The papers by Silverman and Cushman, Crawford, Cadwell, Serafin and Kojs, and Nakamura and Fukamachi were inspired by the musical toy and report further information on the frequencies heard and associated flow velocities as well as required turbulence levels and the characterizing Strouhal numbers.^{3–7} They do not seem to have been aware of the earlier German papers. As Cermak, Crawford³ states that the pipe will sing when the “bump frequency,” or the frequency at which air bumps into the corrugations, equals a natural harmonic of the pipe. Testing pipes with more marked differences between cavity length and pitch indicate that the cavity length is a better parameter to use in the description of the flow acoustic phenomenon.⁸ Common to all the studies however is that when combining frequency, flow velocity, and cavity length/pitch to a Strouhal number, the value of this number will remain fairly constant for a given pipe when the velocity changes. This implies that a small-scale flow phenomenon is present and, at least partly, responsible for the sound generation. By testing tubes of different designs, Cadwell⁵ also finds that a Reynolds number based on the corrugation pitch must exceed 500 for narrow bore pipes to sing. He interprets this as a minimum turbulence level of this scale is required.

Binnie reported in Ref. 9 results from experiments on similar tubes as well as a theoretical analysis of wave propagation in periodically modulated waveguides. Binnie also studied water waves in a shallow canal with streaming water and corrugated banks.¹⁰ Although the ratio of wave velocity to stream velocity in the experiments by Binnie are much higher than the Mach numbers of the present study, it is of interest that the water waves were observed to only propagate in the upstream direction.

Petrie and Huntley¹¹ are especially concerned with reducing the acoustic output of domestic appliances incorporating corrugated pipes. They demonstrated that roughening the cavities, or placing an obstruction in the pipe inlet, will to a large extent achieve this.

A recent contribution to the acoustics of corrugated pipes is presented by Elliot in Ref. 12. An analysis of the shear layer movement and the recognition that the real part

^{a)}Electronic mail: ulf.kristiansen@iet.ntnu.no

^{b)}Presently with Multiconsult AS, Bergen, Norway.

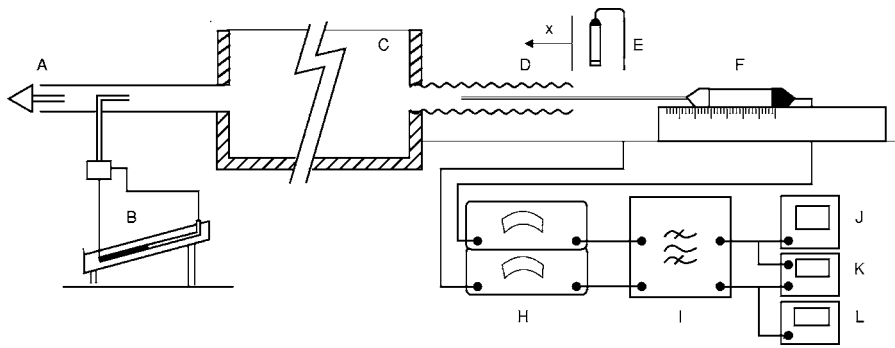


FIG. 1. Schematic of the experimental setup. The different elements are not drawn to scale. A: Pipe connecting box to vacuum cleaner; B: Pitot static tube and manometer; C: box; D: corrugated pipe; E: stationary microphone; F: microphone supplied with probe tube; H: measuring amplifiers; I: bandpass filters; J: voltage meter; K: phase meter; L: frequency counter. Note direction and origin of the distance x .

of the admittance presented by the cavities to the pipe must be negative for sound generation enabled the author to calculate a Strouhal number range for the acoustic excitation that corresponds well with measured values. A theoretical model for the sound propagation in corrugated pipes by Cummings is also detailed in the same reference. Cummings assumes the effect of the cavities to be similar to compressible springs presenting a reactance to the acoustic field along the cylindrical boundary. By assuming an idealized cavity shape, a formula predicting the pipe's acoustic resonances was developed by Cummings. The effect of the corrugations was seen to slow down the acoustic waves and thereby lower the resonant frequencies relative to the ones of a smooth pipe. The formula is interesting in the sense that for low (negligible compared to unity) Mach numbers, it will lower the resonant frequencies by a constant ratio, a feature which is apparent in the experiments reported in the present paper.

Kopiev, Mironov, and Solntseva presented measurements in Ref. 13, showing how an acoustic signal from a loudspeaker in a plenum chamber next to a corrugated pipe would decrease or increase when propagating through the pipe depending on frequency and flow velocity.

Several publications address sound generation by flow over a single cavity, and the sound production caused by the interaction of the unsteady vortical flow and the acoustic domain. General references to the subject are found in Howe.¹⁴ Our experiments involve relatively low velocities, where relevant single cavity studies are the papers by Bruggeman *et al.*,¹⁵ Dequand *et al.*,¹⁶ and Kook and Mongeau.¹⁷

The primary interest of the present work was to investigate if we could find out more about the flow acoustic phenomenon by doing measurements inside the pipe while it sounds at one of its resonant frequencies.

II. EXPERIMENT

The experimental setup is illustrated in Fig. 1. The corrugated pipe has one end (inflow) open to the laboratory, and the other attached to a box supplied with a vacuum cleaner in order to draw air through the system. The box has the dimensions $0.57(\text{length}) \times 0.29 \times 0.325 \text{ m}^3$ and is lined on the inside with a layer of mineral wool, 50 mm thick with a density of 18 kg/m^3 . One side ($0.29 \times 0.57 \text{ m}$) is not lined, but supplied with a glass window to allow visual inspection. The acoustic properties of the box were evaluated by measuring the pressure reflection factor experienced by waves in a

smooth pipe of inner diameter 28 mm terminated by the box at “no flow” conditions for the frequency range of interest. The pipe has a loudspeaker as sound source at one end, and is supplied with two Brüel & Kjær $\frac{1}{4}$ in. microphones situated 25 mm apart, allowing for reflection measurements to be performed using the method described by Chung and Blazer,¹⁸ and implemented in the WinLMS¹⁹ measuring system. The result is shown in Fig. 2 where also the reflection factor of the end open to the laboratory is included for comparison. Figure 2 shows that there are no large differences in the reflection factor curves in this frequency range. Flow has however been found to increase the reflection factor magnitude at the discharge end of smooth surfaced flow pipes open to free space.^{20,21}

The experimental laboratory is not anechoic. The experiment is placed on a laboratory table, and the guide rail is placed about 10 cm upstream of the pipe opening, see Fig. 1. However, an upstream end reflection factor measurement for a configuration where the table and guide rail were removed showed that these structures only have a minimal influence on the reflection factors in the frequency range of interest.

The vacuum cleaner is attached via a smooth pipe of length 665 mm and inner diameter 34 mm in which the centerline velocity is measured using a Pitot static tube attached to a manometer. The tip of the Pitot tube was positioned 465 mm downstream of the box. The flow velocity was calculated by the formula $U = \sqrt{2\Delta P/\rho}$, where ρ is the air den-

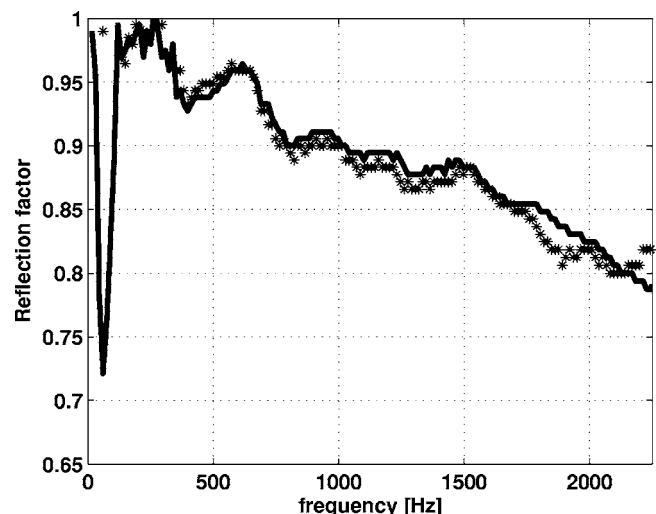


FIG. 2. Modulus of the pressure reflection factor of 12 mm radius smooth pipe openings. Complete line: box end; stars: end open to laboratory.

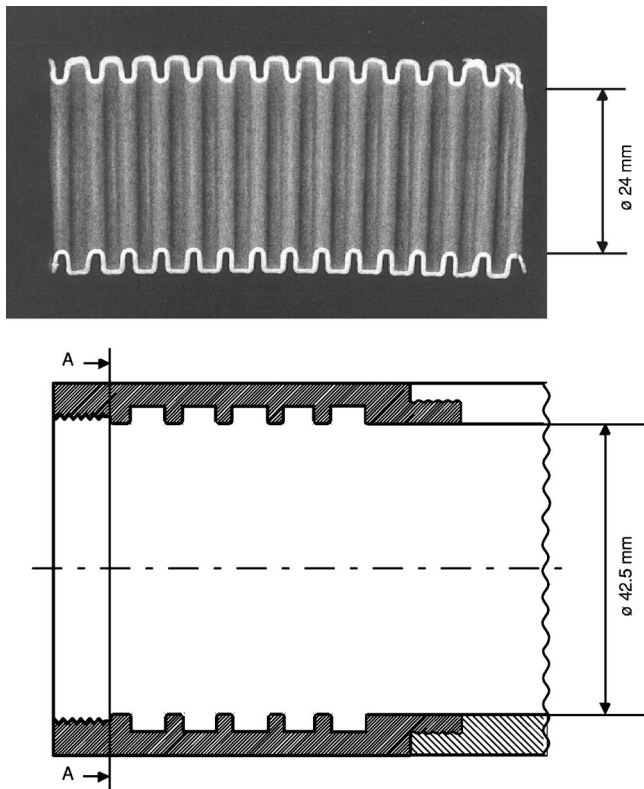


FIG. 3. Sections of the two pipe types used. Upper panel: A short section of the corrugated pipe; lower panel: Five cavity section at inflow end of longer smooth pipe.

sity and ΔP the pressure difference measured by the Pitot tube. The acoustic pressures were measured using $\frac{1}{2}$ in. Brüel & Kjær microphones (type 4166). One of the microphones was supplied with a 230-mm-long probe tube of 3.2 mm inner diameter (Brüel & Kjær probe kit UA0040), extended to a length of 700 mm using a 3.2-mm-o.d., 2.6-mm-i.d. tube. This allowed pressure measurements to be made inside the corrugated pipe. The probe tube microphone preamplifier was supported by a guiding rail having a millimetric scale. The stationary microphone was normally positioned a few centimeters off axis and upstream of the pipe entry (4–5 cm). The microphone signals were fed measuring amplifiers (Brüel & Kjær type 2606) and filtered by 1/3 octave filters (Norsonic type 719). The sound level amplitudes were measured by a Brüel & Kjær rms voltmeter (type 2417), normally using a 3 s averaging time, and the phase difference between the microphone signals was measured by a Hewlett-Packard 3575A gain phase meter displaying a value every 25 ms. The reference microphone was calibrated using a piston phone, and the probe microphone was calibrated by placing the probe tip and reference microphone close together in a free field generated by a loudspeaker at the frequency in question.

The pipes tested are illustrated in Fig. 3. The fully corrugated pipe has a length of 615 mm, with an inner (open area) pipe diameter of 24 mm. The cavity depth is 3.3 mm, and the cavity pitch 5.4 mm. The cavity edges are rounded. The pipe was chosen as it sings readily, and has a diameter large enough so that the microphone probe tube traversing

the pipe would not appreciably alter the sound. In one test, a 410-mm-long pipe of the same geometry was also tested.

Figure 3 also shows one of two Five-cavity sections, positioned at a smooth pipe inlet in the picture. These short cavity sections could be placed at different positions in a pipe of total length 800 mm made up by short (50 and 100 mm) smooth sections. The smooth pipe has an inner diameter of 42.5 mm, and in the two sections with cavities, the cavities were machined to a depth of 2.75 mm, length 5 mm, and pitch 7.5 mm. The leading and trailing cavity edges are slightly rounded.

It was possible to regulate the velocity of air through the system by adjusting the vacuum cleaner setting. For the measurements, the velocity was adjusted so that a clear tone was heard. The microphone probe tip was then moved along the pipe to measure the sound pressure level at regular interval positions, Δx , apart. This was achieved by manually moving the microphone preamplifier along the guide rail. The sound pressure level and relative microphone phase were read visually on the instrument displays and entered a measurement log. After the completion of a measurement series, the data were analyzed on a PC. We extracted the axial (x direction) intensities, or energy fluxes from the magnitudes and relative phases of the pressure values. The idea was that regions with a strong increase in axial intensity would indicate regions of sound energy generation. Regions of constant or decreasing intensity values along the axis would indicate regions of no effective sound sources or energy loss.

For the low Mach numbers involved, the intensity was calculated using no flow expressions. For a single frequency sound field, the time averaged active intensity is given as the expression

$$I_x = \frac{1}{2} \text{Re}\{p(\omega)u_x^*(\omega)\}, \quad (1)$$

where $p(\omega)$ and $u(\omega)$ are complex quantities representing the acoustic pressure and particle velocity. The star represents complex conjugation, and the subscript x designates the axial direction. The particle velocity is related to the gradient of the acoustic pressure by the Euler equation

$$u_x = \frac{1}{i\omega\rho} \frac{\partial p}{\partial x}, \quad (2)$$

where ρ and ω represent the fluid density and circular frequency, while “ i ” is the imaginary unit.

A simple approximation to the pressure gradient is a second-order finite difference expression, yielding the particle velocity as

$$u_x(x) \approx \frac{1}{i\omega\rho} \frac{p(x-H) - p(x+H)}{2H}, \quad (3)$$

where H is a small distance compared to the wavelength of sound. The pressure value at the point x is approximated by the arithmetic mean of the point pressure values

$$p(x) \approx \frac{p(x-H) + p(x+H)}{2}. \quad (4)$$

If the complex pressures further are written in amplitude and phase as

$$p(x-H) = |p(x-H)|e^{i\phi(x-H)},$$

$$p(x+H) = |p(x+H)|e^{i\phi(x+H)}, \quad (5)$$

with ϕ being the phase angle between the signals at the measurement position and a reference microphone position, the substitution of Eq. (5) into Eqs. (4), (3), and (1) gives

$$I(x) \approx \frac{1}{2\rho\omega(2H)} |p(x-H)||p(x+H)| \sin(\phi(x-H) - \phi(x+H)), \quad (6)$$

for the axial intensity at point x at the circular frequency ω . Note that in this expression, the direction of the intensity vector is given by the sign of the difference between the two phase angles and this is a point of interest for the interpretation of the experimental results.

The approximations given as Eqs. (3) and (4) form the basis for the so-called “two microphone technique” used in acoustic intensity measurements,²² and expression (6) could be regarded as a single frequency version of the more general expression:

$$I_x \approx \frac{1}{2\rho\omega(2H)} \int_{-\infty}^{\infty} \frac{\text{Im}\{C_{21}(\omega)\}}{\omega} d\omega, \quad (7)$$

where C_{21} denotes the cross-spectral density between the pressures p_2 and p_1 .²³ Exploring pure tone sound fields by measuring complex pressure values in a network of positions by the use of a single microphone has also been reported in Refs. 23 and 24. An advantage of this method is that it eliminates the possible instrument phase mismatches associated with the use of two separate microphone chains. For stable sound fields in still air, the measurement errors involved in the estimation of intensity by this technique are linked to the approximations given as Eqs. (3) and (4) and hence directly related to the separation distance between the two microphone points.^{22,23} For measurements in unsteady flows, further errors are introduced. In particular it should be noted that Eq. (2) is strictly only valid for one-dimensional flow. The unstable flow in the pipe, to a large extent caused by the corrugated geometry, must therefore be expected to cause fluctuations in the data.

III. RESULTS

At the onset of each experiment, we adjusted the velocity until a clear tone was heard. It should be mentioned that repeating experiments on a given mode would often yield slightly different frequencies and different excitation levels. Once a resonance was established, the probe was moved manually to the different measurement positions and the sound pressure levels and phase values were registered from the instruments. The sound pressure levels could vary by typically ± 0.2 dB between repeated measurements at a given position. Generally there was a monotonic increase in the phase meter values toward the downstream end of the fully corrugated pipe. For measurements useful for intensity calculations, repeated measurements at a given position could vary approximately $\pm 0.25^\circ$. Other experimental results had

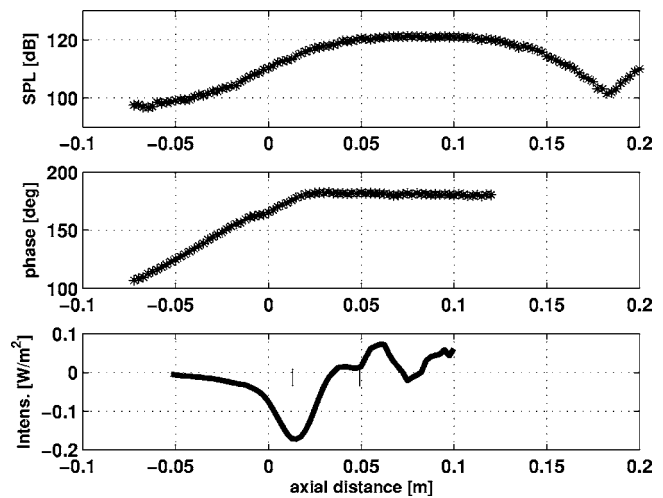


FIG. 4. Sound pressure level, phase, and axial intensity along part of the pipe (0.76 m long) at 860 Hz. The distances are measured in the direction of the flow. The tube opening has the axial distance “0.” $\Delta x=2.5$ mm, $H=4\Delta x$, and $F=5$.

to be discarded for such calculations due to too much variation in phase values at a given position. The data were spatially smoothed in the subsequent data analysis. Two methods were used to achieve this, the distance between the points used in the estimation of the intensities were increased, that is letting H correspond to multiples of Δx , and also the obtained intensity values were filtered by a moving average filter of frame size F . Since the frequencies in question involved acoustic wavelengths much longer than the pipe diameter, we expected plane acoustic waves in the pipe. This was confirmed by letting the probe tip traverse the pipe cross section at fixed axial positions. We did not measure any radial pressure variation in the flow pipe beyond normal pressure fluctuations.

A. Short corrugated section

As an initial investigation, a steel tube made up by sections 50 and 100 mm long and 42.5 mm diameter sections was tested. In two of the sections, five cavities were machined along the inside surface. The interest of this test was twofold. We wanted to find out whether such a limited number of cavities would generate the high sound levels observed for fully corrugated pipes, and also whether the intensity measurements would ascertain that the acoustic energy is generated in the cavity region and thereby give credibility to the experimental design.

Pipes of different lengths made up by the smooth sections (maximum total length 700 mm) were tested with cavity sections at different positions in the pipes. For the velocity range tested, up to 12 m/s, the pipe would only sound strongly with a cavity section at the inflow end (see Fig. 3). It should also be mentioned that in a test where the small upstream protruding lip was machined away from the cavity section (the fine thread region to the left of section A-A in Fig. 3), the strong sound disappeared.

Figure 4 shows measurements in the cavity region of the pipe. The frequency corresponds to the fourth mode of an open-open pipe. The short vertical lines in the lower figure

TABLE I. Columns 1–4: Mode number, theoretical and measured frequencies, and their difference expressed in percent. Column 5: Typical midpipe flow velocity for excitation of that particular mode. Columns 6 and 7: Reynolds and Strouhal numbers based on cavity pitch. The Reynolds number involves the kinematic viscosity ν , whose value was taken as $1.51 \times 10^{-5} \text{ m}^2/\text{s}$. The cavity pitch l_p is 5.4 mm.

Mode No.	f_t (theoretical) (Hz)	f_m (measured) (Hz)	$\frac{f_t - f_m}{f_t} \times 100$ (%)	\bar{U}_0 (m/s)	Re $\frac{\bar{U}_0 l_p}{\nu}$	St $\frac{f_m l_p}{\bar{U}_0}$
1	273	—	—	—	—	—
2	543	496	9.1	7.6	2718	0.353
3	815	740	9.6	12.0	4290	0.333
4	1087	990	9.3	16.8	6007	0.318
5	1358	1230	9.8	21	7508	0.316
6	1630	1497	8.5	25.8	9225	0.314
7	1902	1731	9.4	31.2	11,155	0.300
8	2174	1968	9.8	36.0	12,871	0.296
9	2445	2216	9.7	39.2	14,076	0.306

mark the leading edge of the upstream, and the trailing edge of the downstream cavities. The data were measured at intervals $\Delta x = 2.5 \text{ mm}$. The intensity at each measuring point was then obtained using the formulas given in the experimental section, with $H = 4\Delta x$, and further filtered using a frame size $F = 5$.

rms sound pressure level and phase relative to a stationary microphone are shown in the two top figures. The pressure plot is extended to the first minimum, showing the standing wave ratio to be close to 20 dB. From the phase curve, it is seen that the slope changes its sign in the cavity region, implying a change in the energy flow direction.

The intensity is shown in the lower figure, where the intensity in the upstream direction has negative values due to the chosen sign convention. The intensity plot shows that the sound energy is generated in the cavity region. The highest intensity values correspond to upstream propagating energy. The intensity in this direction decays rapidly toward and into the outside field. Note that the energy decrease toward the outside field manifests itself from a distance inside of the physical tube opening. The downstream propagating part displays a much more rugged curve, indicating the difficulty of

obtaining exact intensity values when the phase slope is small and easily influenced by fluctuations in the phase values.

B. Fully corrugated pipes

Most of the measurements were performed on a 615-mm-long pipe, and a number of resonances were found for the available velocities. A summary of the experimental values are given in Table I. The velocities were measured in the smooth pipe connecting the box to the vacuum cleaner. The velocities given in Table I are the measured values multiplied by the cross-sectional area fraction of the two pipes, where the inner open area section was used for the corrugated pipe. The accuracy of the velocity measurements is estimated to be $\pm 1 \text{ m/s}$. From Table I, we see that the lowest mode was not excited for this pipe length (see the last paragraph of Sec. III B). The resonant frequencies were found to be lower than the ones calculated theoretically for a smooth pipe by about 9.5%. The theoretical values were calculated from the formula $f = nc/2L'$, where L' is the effective pipe length. Plane waves are assumed for the pipe's acoustic field and end corrections of 60% of the pipe radius, corresponding to those for

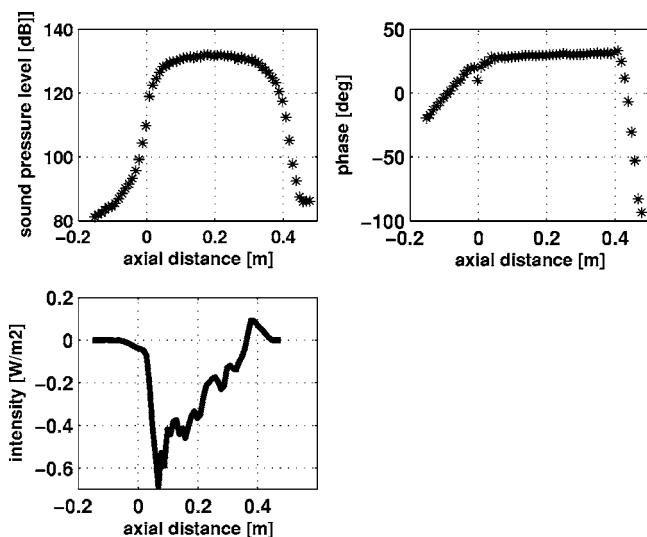


FIG. 5. Sound pressure level, phase, and axial intensity for the fundamental mode of 410 mm pipe. $f = 370 \text{ Hz}$, $\Delta x = 10 \text{ mm}$, $H = 5\Delta x$, and $F = 3$.

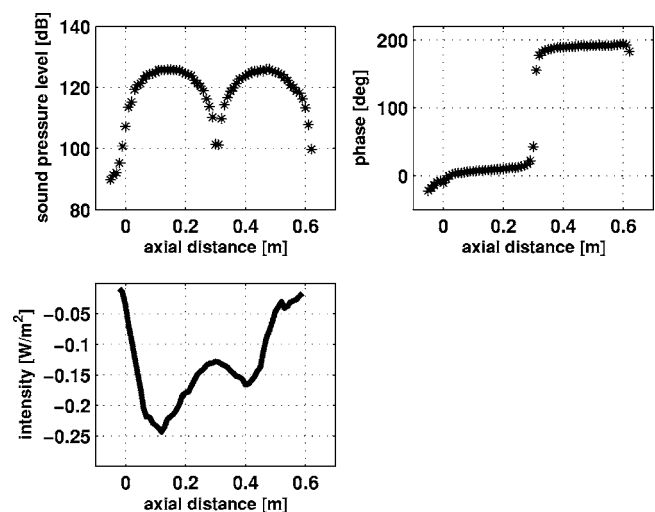


FIG. 6. Sound pressure level, phase, and axial intensity for the second mode of 615 mm pipe. $f = 497 \text{ Hz}$, $\Delta x = 10 \text{ mm}$, $H = 3\Delta x$, and $F = 7$.

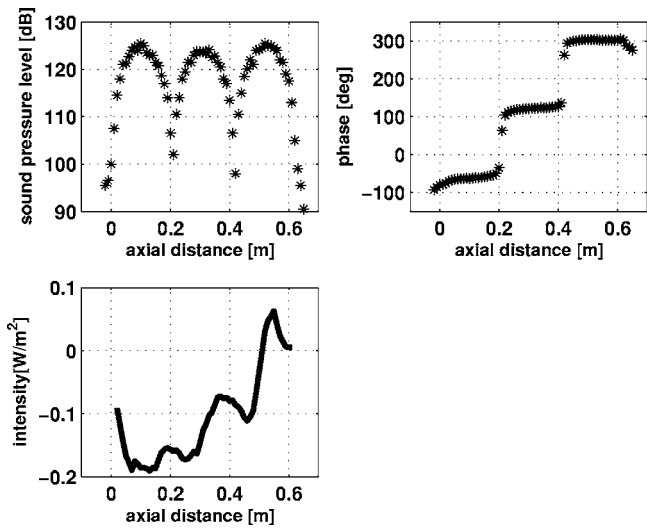


FIG. 7. Sound pressure level, phase, and axial intensity for the third mode of 615 mm pipe. $f=749$ Hz, $\Delta x=10$ mm, $H=2\Delta x$, and $F=7$.

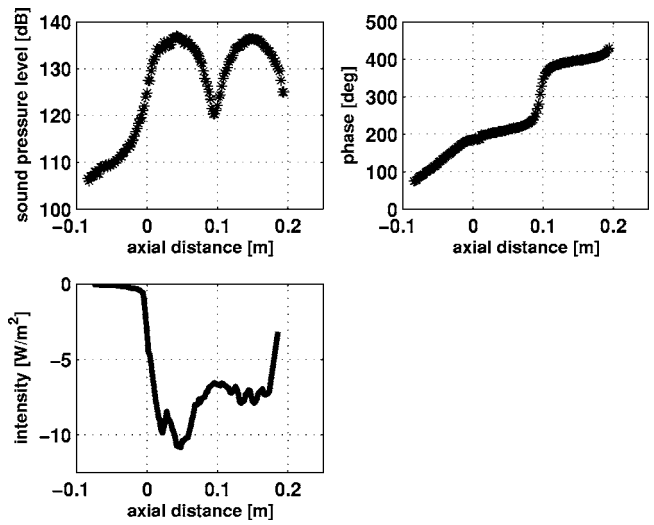


FIG. 9. Sound pressure level, phase, and axial intensity measured for the two first peaks of the sixth mode of 615 mm pipe. $f=1470$ Hz, $\Delta x=2$ mm, $H=2\Delta x$, and $F=13$.

unbaffled pipes, are added to the geometrical pipe length at both ends.²⁵ The speed of sound was taken as 343.4 m/s and the pipe radius as the free cross section radius of 12 mm. Using the maximum inner radius, 15.3 mm, would reduce the values of column 4 by 0.5%–0.6%. We believe this lowering in frequency to be a geometry effect caused by the cavities rather than a flow effect and will come back to this in the discussion. Table I also shows that the Strouhal number is reasonably constant.

It was not possible to excite the lowest mode strongly for the 615-mm-long pipe. For a shorter pipe, however, a possible excitation of this mode would be expected at a higher velocity. Figure 5 shows measurements on a 410-mm-long pipe where the fundamental mode was excited for a flow velocity of 6 m/s. The data were measured at intervals $\Delta x=10$ mm. The intensity values were derived using $H=5\Delta x$, and further filtered using a frame size $F=3$.

Similar plots are shown for the 615 mm pipe’s modes 2 and 3 in Figs. 6 and 7. We observe that the energy generation occurs in the region of pressure peaks, and that the lowering of intensity values around the standing wave nodes (where

the axial particle velocities are the highest) indicates some sound energy absorption in the pressure nodal regions. The phase plots show a monotonic increase in the phase values in the flow direction up until a short distance inside of the downstream opening. Here the slope changes direction, indicating energy flowing into the box. Due to relatively few data points in this region, and because fluctuations in the data are especially detrimental in this region where part of the phase curve is nearly flat, it was difficult to get good estimates of the positive direction intensity. For mode 3, the energy seems predominantly to be generated around the second and third pressure peaks counted from the inflow end.

For the sixth mode shown in Figs. 8 and 9 only the intensity value in the region of the first two maxima is plotted. The monotonic increase in phase observed up to near the end of the pipe (Fig. 8) does however clearly indicate the direction of the energy flow for the major part of the tube. Note also the general lowering in sound pressure level at the minima of the standing wave pattern with distance from the inflow end.

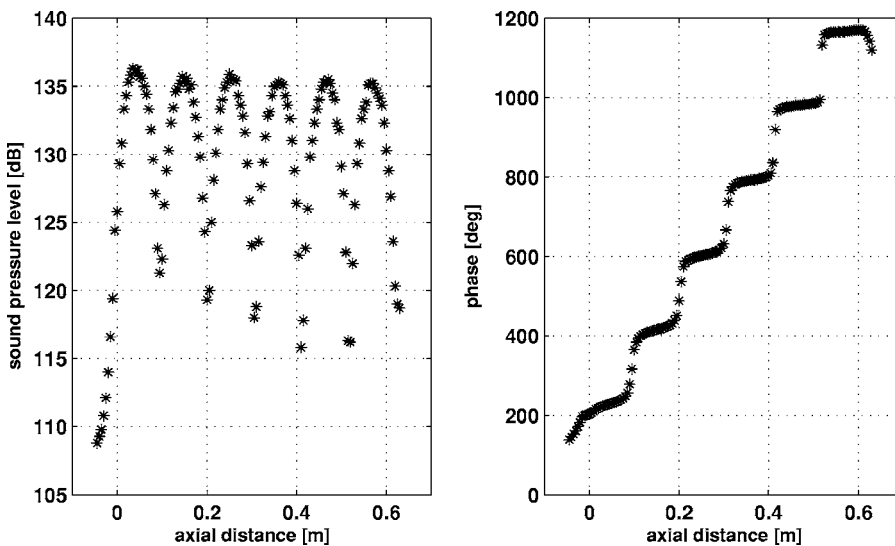


FIG. 8. Sound pressure level and phase for the sixth mode of 615 mm pipe. $f=1483$ Hz and $\Delta x=5$ mm.

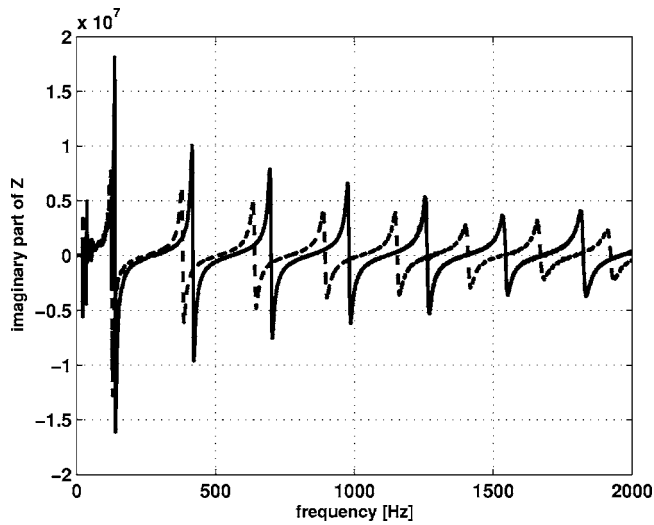


FIG. 10. Input impedances of smooth (complete line), and corrugated (dashed line) pipes of equal length in the no flow case.

The decrease in intensity toward the outside field starts also in these pipes a certain distance into the pipes' upstream openings. There is a possibility for energy dissipation in this region, but the present experimental setup did not allow an estimation of such.

Note that the frequencies of maximum excitation would vary slightly from test to test (compare Table I and the figure legends).

Low velocity excitation. When starting at low velocities, we found that the first resonances could be excited at relatively weak amplitudes in the velocity range 1–4 m/s for the 615-mm-long pipe. Increasing the velocity further, the pipe would restart the excitation beginning at mode 2, but now with much higher amplitudes. This indicates two different modes of excitation. Preliminary intensity measurements (not presented in the present paper) made at the low velocities indicate that the excitation takes place close to the inlet section of the pipe.

IV. DISCUSSION

A. The reduced frequencies

The lowering of the acoustic resonance frequencies by about 9% from what was calculated for a smooth pipe was evident in all our measurements involving fully corrugated pipes. We continued this investigation by measuring the input impedance of the corrugated pipe and compared it to the input impedance of an equally long, smooth pipe with an internal diameter of 25 mm. The measurements were done without an internal flow. A specially designed sound source and receiver unit²⁶ was fixed to one end while the other end was open to the laboratory, see also Ref. 27 for details on impedance measurements of this kind. Figure 10 shows the imaginary parts of the input impedances of the two pipes. The curves are similar in shape, and characterized by sharp positive and negative peaks. The curve for the corrugated pipe is seen to be shifted to lower frequency values by about 8.5%, which is close to what was found during the flow tests. As the lengths of the pipes are equal, and the difference in

diameter minimal, this is interpreted as a reduction in the effective wave speed for the corrugated pipe. Our conclusion is therefore that this reduction in wave speed and consequently also in resonant frequency is predominantly a geometry effect rather than a flow effect for the reported tests. The resonant frequencies are defined as the ones giving zero input reactance. For a pipe driven at one end and open at the other, the formula $f=nc/2L'$ is given in Ref. 25 for these frequencies where L' is an effective pipe length. Using 343.4 m/s as the speed of sound and an open unflanged end correction of 60% of the pipe radius, the above-presented expression gives the measured values to within 2% for the smooth pipe.

It is known that geometry variations along a pipe will cause changes in a pipe's resonant frequencies. Forbes and Pike²⁸ presented a method based on perturbation analysis of a duct to study the effect a simple localized constriction has on the eigenvalues. The same technique could probably be used to study a multicavity geometry, but no attempts to do so have been found in the literature.

The work by Cummings cited in Sec. I¹² relates this lowering in wave speed to the spring effect of the cavity volumes along the pipe surface. By filling a corrugated pipe with water and weighing the water volume, a single circumferential cavity was estimated to have the volume $V=632.3\text{ mm}^3$. The cavity depth d is 3.3 mm and the inner open cross section radius $R=12\text{ mm}$; consequently an equivalent rectangular cross-section cavity has an effective length $l_c=V/(\pi(R+d)^2-\pi R^2)=2.2\text{ mm}$. Cummings' formula for resonant frequencies is written

$$f_n = \frac{nc}{2L'} \frac{\left[1 - M^2 \left(1 - i \frac{P \langle \beta_\omega \rangle}{A k} \right)^{1/2} \right]}{\left(1 - i \frac{P \langle \beta_\omega \rangle}{A k} \right)^{1/2}}, \quad (8)$$

where $P=2\pi R$ and $A=\pi R^2$ designate the perimeter and cross-sectional area of the open section; $k=\omega/c$ is the wave number; and $\langle \beta_\omega \rangle$ is the spatially averaged wall admittance. L' is the effective pipe length.

Cummings further assumes that the cavities behave like spring-type impedances in the frequency region where the acoustic wavelengths are very much longer than the cavity dimensions:

$$\beta_\omega = ikV/S \text{ over the cavities,} \quad (9)$$

$$\beta_\omega = 0 \text{ elsewhere, and } \langle \beta_\omega \rangle = l_c/l_p ikV/S. \quad (10)$$

l_c and l_p designate the cavity length and pitch, respectively (2.2 and 5.4 mm). S is the surface presented by the cavity to the pipe, $S=2\pi R l_c(165.9\text{ mm}^2)$. In situations where M^2 is negligible compared to unity, the formula gives

$$f_n = 0.89 \frac{nc}{2L'} \quad (11)$$

for the dimensions measured for the present corrugated pipe. This reduction of 11% is close to the observed value (see Table I) and therefore suggests further that the spring effect

by the cavities might be the reason for the reduction in frequencies from the smooth pipe values.

B. The acoustic sources

Results from measurements on the pipe with a short corrugated section show that the increase in energy flow is a valid technique to pinpoint effective energy producing acoustic sources. For the fully corrugated pipes, sources were found in the regions of the pressure maxima. An interesting observation from the measurements on these pipes is that the net acoustic energy stream is always in the same direction for nearly the total length of the pipe, that is against the direction of the flow. Only close to the downstream end of the pipe did we measure intensity vectors in the air flow direction. This can be seen from the general inclination of the phase curves, indicating the energy flow direction [Eq. (6)].

The reflection factor measurements shown as Fig. 2 show that the pipe end open to the downstream box has close to the same reflection factors as the end open to the laboratory, for the frequency range of our experiments (370–2216 Hz) for “no flow” conditions. Flow effects at the box end might however increase the reflection factor,^{20,21} and thereby have an influence on the observed energy flow direction.

Studies on the sound production by flow over single cavities have demonstrated that the resulting acoustic fields can have strong upstream directivities. Two numerical studies showing this effect are the papers by Colonius *et al.*²⁹ and Gloerfelt *et al.*³⁰ The Mach numbers of these studies are however much higher than in the present experiments ($M=0.6$ and $M=0.7$), and the resulting sound fields have wavelengths comparable to cavity dimensions. Preferential radiation in the upstream direction was also evident in a numerical study by Hardin and Pope³¹ for a Mach number $M=0.1$. The flow field in their investigation was found to oscillate with a Strouhal number of $St=fl_c/U=0.58$, and the acoustic wavelength was approximately 17 cavity lengths. Analytical results by Howe¹⁴ at $M=0.1$ indicate the same acoustic behavior.

We therefore believe that also source directivity is a possible explanation for the observations made in the present study. Further studies on flow source directivity at low Mach numbers for multicavity systems such as those found in corrugated pipes, as well as the interaction of the flow and global acoustic field, should therefore be undertaken to more fully understand the phenomena.

ACKNOWLEDGMENT

The authors would like to thank Øyvind Lervik for his help with the experiments.

¹W. Burstyn, “Eine neue pfeife (a new pipe.),” *Z. Tech. Phys.* (Leipzig) **3**, 179–180 (1922).

²P. Cermak, “Über die tonbildung bei metallschläuchen mit eingedrücktem spiralgang (on the sound generation in flexible metal hoses with spiralling grooves),” *Phys. Z.* **23**, 394–397 (1922).

³F. S. Crawford, “Singing corrugated pipes,” *Am. J. Phys.* **42**, 278–288

(1974).

⁴M. P. Silverman and G. M. Cushman, “Voice of the dragon: The rotating corrugated resonator,” *Eur. J. Phys.* **10**, 298–304 (1989).

⁵L. H. Cadwell, “Singing corrugated pipes revisited,” *Am. J. Phys.* **62**, 224–227 (1994).

⁶S. Serafin and J. Kojs, “The voice of the dragon: A physical model of a rotating corrugated tube,” in *Proceedings of the Sixth International Conference on Digital Audio Effects (DAFex-03)*, London, 8–11 September 2003.

⁷Y. Nakamura and N. Fukamachi, “Sound generation in corrugated tubes,” *Fluid Dyn. Res.* **7**, 251–261 (1991).

⁸U. Kristiansen, T. A. Reinen, and G. A. Wiik, “Sound generation in corrugated pipes,” in *Proceedings of the Internoise 2005*, Rio de Janeiro, Brazil, 7–10 August 2005.

⁹A. M. Binnie, “Self induced waves in a conduit with corrugated walls. ii. Experiments with air in corrugated and finned tubes,” *Proc. R. Soc. London, Ser. A* **262**, 179–197 (1961).

¹⁰A. M. Binnie, “Self induced waves in a conduit with corrugated walls. i. Experiments with water in an open horizontal channel with vertically corrugated sides,” *Proc. R. Soc. London, Ser. A* **259**, 18–27 (1960).

¹¹A. M. Petrie and I. D. Huntley, “The acoustic output produced by a steady airflow through a corrugated duct,” *J. Sound Vib.* **70**, 1–9 (1980).

¹²*Lecture Notes on the Mathematics of Acoustics*, edited by M. C. M. Wright (Imperial College Press, London, 2005), Chap. 11.

¹³V. F. Kopiev, M. A. Mironov, and V. S. Solntseva, “Sound generation, amplification and absorption by air flow through waveguide with periodically corrugated boundary,” in *Proceedings of Forum Acusticum*, Budapest, Hungary, 29 August - 2 September 2005.

¹⁴M. S. Howe, “Mechanism of sound generation by low Mach number flow over a wall cavity,” *J. Sound Vib.* **273**, 103–123 (2004).

¹⁵J. C. Bruggeman, A. Hirschberg, M. E. H. van Dongen, A. P. J. Wijnands, and J. Gorter, “Self-sustained aero-acoustic pulsations in gas transport systems: Experimental study of the influence of closed side branches,” *J. Sound Vib.* **150**, 371–393 (1991).

¹⁶S. Dequand, S. J. Hulshoff, and A. Hirschberg, “Self-sustained oscillations in a closed side branch system,” *J. Sound Vib.* **265**, 359–386 (2003).

¹⁷Kook and L. Mongeau, “Analysis of the periodic pressure fluctuations induced by flow over a cavity,” *J. Sound Vib.* **251**, 823–846 (2002).

¹⁸J. 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).

¹⁹WinMLS Morset Sound Development, Trondheim, Norway.

²⁰R. M. Munt, “Acoustic transmission properties of a jet pipe with subsonic jet flow. I. The cold jet reflection coefficient,” *J. Sound Vib.* **142**, 413–436 (1990).

²¹S. Allam and M. Åbom, “Investigation of damping and radiation using full plane wave decomposition in ducts,” *J. Sound Vib.* **292**, 519–534 (2006).

²²F. J. Fahy, “Measurement of acoustic intensity using the cross-spectral density of two microphone signals,” *J. Acoust. Soc. Am.* **62**, 1057–1059 (1977).

²³O. K. ØPettersen, “Sound intensity measurements for describing acoustic power flow,” *Appl. Acoust.* **14**, 387–397 (1981).

²⁴F. J. Fahy, *Sound Intensity* (Spon, London, 1995).

²⁵L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 2000).

²⁶Brass Instruments Analysis System (BIAS), Institute for Musical Acoustics, Vienna.

²⁷J. P. Dalmont, “Acoustic impedance measurement. i. A review,” *J. Sound Vib.* **243**(3), 427–439 (2001).

²⁸B. J. Forbes and E. R. Pike, “Acoustical Klein-Gordon equation: A time-independent perturbation analysis,” *Phys. Rev. Lett.* **93**(5), 054301 (2004).

²⁹T. Colonius, A. J. Basu, and C. W. Rowley, “Computation of sound generation and flow/acoustic instabilities in the flow past an open cavity,” in *Proceedings of the FEDSM99 3rd ASME/JSME Joint Fluids Engineering Conference*, San Francisco, 18–23 July 1999.

³⁰X. Gloerfelt, C. Bailly, and D. Juvé, “Direct computation of the noise radiated by a subsonic cavity flow and application of integral methods,” *J. Sound Vib.* **266**, 119–146 (2002).

³¹J. C. Hardin and D. S. Pope, “Sound generation by flow over a two-dimensional cavity,” *AIAA J.* **33**, 407–412 (1995).

Vortex sound due to a flexible boundary backed by a cavity in a low Mach number mean flow

S. K. Tang^{a)}

Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

R. C. K. Leung and R. M. C. So

Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong, China

(Received 26 April 2006; revised 5 November 2006; accepted 15 December 2006)

Low frequency sound radiated due to the unsteady motion of an inviscid vortex in the proximity of a flexible membrane backed by an airtight cavity on an otherwise rigid plane is investigated theoretically. Results show that both monopole and dipole are created but the latter is important only when the vortex is traversing over the membrane. The monopole results from the membrane vibration and the dipole from the transverse motion of the vortex. It is also found that these sound fields tend to counteract each other. The increase in the mean flow speed in general results in a stronger acoustic power radiation, but sound attenuation may be possible if the membrane-cavity system is weak compared with the mean flow momentum. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434240]

PACS number(s): 43.28.Ra, 43.40.Rj, 43.50.Nm [GCL]

Pages: 1345–1352

I. INTRODUCTION

It is well known that the unsteady motion of turbulence is an important source of aerodynamic sound, especially in the presence of nonrigid solid surfaces.^{1,2} The regeneration noise inside and the breakout noise from an air duct in a building conveying low Mach number turbulent flow are typical examples of such sound.³ Silencing devices, such as the dissipative silencer and expansion chamber,⁴ which are adopted to attenuate fan noise in the ventilation system will also interact with the turbulent flow to produce noise. This low frequency aerodynamic noise is relatively difficult to attenuate in a cost-effective manner.

It has been confirmed that the low Mach number flow tends to amplify the low frequency aerodynamic sound power radiation (for instance, Ffowcs Williams and Lovely,⁵ Dowling,⁶ and Howe⁷). It also produces a Doppler effect which alters the directivity of the sound radiation. Recently, Huang⁸ has shown that a flexible membrane backed by a cavity inside a rigid walled duct can result in significant sound attenuation in the absence of a mean flow. In the presence of a turbulent flow, the unsteady motions of the turbulence and the flow-induced membrane vibration are expected to be sound producing and thus will deteriorate the performance of the membrane setup of Huang.⁸ In fact, the self-noise generated by various duct silencing devices in the presence of turbulent flow is one of the very hot research topics.

Direct computation of sound generated by a low Mach number flow is difficult enough because of the great disparity in scales between the acoustic and aerodynamic field.^{9,10} Sound generated by a turbulent flow in the presence of a flexible boundary further complicates the direct computation of sound because of the very stringent requirement in the

computation of the acoustic field, the flow field, and the flow-induced vibration of the flexible boundary. There have been studies concerning the noise radiated from plate vibration driven by a turbulent flow by assuming a certain turbulence spectrum (for instance, Davies,¹¹ Wu and Maestrello,¹² and Dowell¹³). However, such assumption tends to ignore the important coupling among the flow, the acoustic field, and the plate. Also, the underlying sound generation mechanisms have not been discussed.

Real turbulence is also difficult to handle theoretically, making the study of aerodynamic sound generation mechanisms difficult. Therefore vortices, though they are drastic simplifications of the turbulent flow, are often adopted to get insights into the sound generation process. Crighton¹⁴ used this vortex analogy to get insight into the radiation of the edge noise. More examples can be found in Howe,¹⁵ Leung and So,¹⁶ and Tang and Lau.¹⁷ The recent work of Tang *et al.*¹⁸ shows the presence of significant dipole and monopole radiation when a vortex moves over a flexible membrane in the absence of a mean flow. The former is due to the transverse motion of the vortex, implying that similar dipole radiation will be produced by the turbulent eddies inside a turbulent flow when they interact with a flexible boundary.

The present investigation is focused on the aeroacoustics when a vortex interacts with a finite length flexible boundary backed by an airtight cavity in the presence of a low Mach number mean flow. The setup is analogous to that of the “drumlike” silencing device of Huang.⁸ It is hoped that the results of the present study can provide information on the self-noise of such device in a practical flow duct and to enhance fundamental understanding of aeroacoustics.

II. THEORETICAL MODEL

The turbulent flow inside a practical duct is of high Reynolds number but low Mach number. The effect of viscosity is therefore of second importance in the present study.¹⁹ Fig-

^{a)}Author to whom correspondence should be addressed; electronic mail: besktang@polyu.edu.hk

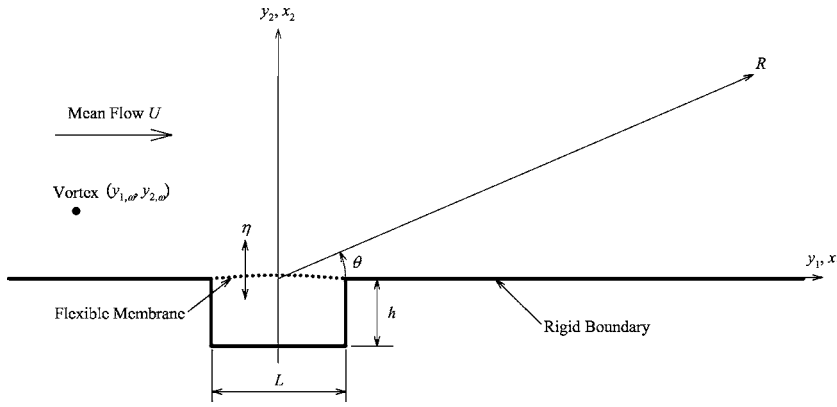


FIG. 1. Schematic of the present theoretical model.

ure 1 illustrates the schematic of the present study. An inviscid vortex with circulation Γ and initial height d is chosen. It moves toward the finite length flexible membrane backed by a cavity of depth h in the presence of a mean parallel flow of velocity U . This flow extends to the far field. The mechanical properties of the membrane are characterized by the tension per unit spanwise length T , the mass density M_m , and the damping factor D as in Huang⁸ and Tang *et al.*¹⁸ In the foregoing discussion, \mathbf{y} and \mathbf{x} denote positions in the near and far fields, respectively, with origin at the middle of the membrane.

A. Vortex dynamics and membrane vibration

Since the vibration displacement magnitude of the membrane is very small when compared to d ($< 1\%$ in the present study), the membrane is modeled as a rigid surface with a time varying velocity perturbation v in the present study as in Tang *et al.*¹⁶ The velocity of the vortex v_ω at the position $(y_{1,\omega}, y_{2,\omega})$ is

$$v_\omega = \left(U + \frac{\Gamma}{4\pi y_{2,\omega}} \right) \hat{y}_1 + \frac{\hat{y}_1}{\pi} \int_{-L/2}^{L/2} \frac{v(y_{1,\omega} - y_1)}{(y_{1,\omega} - y_1)^2 + y_{2,\omega}^2} dy_1 + \frac{\hat{y}_2}{\pi} \int_{-L/2}^{L/2} \frac{v y_{2,\omega}}{(y_{1,\omega} - y_1)^2 + y_{2,\omega}^2} dy_2, \quad (1)$$

where the last two terms on the right-hand side of Eq. (1) represent the fluid velocity induced by the vibrating flexible boundary at the position of the vortex, which are obtained by potential theory.²⁰

The unsteady vortex motion induces fluid pressure fluctuation over the flexible membrane and thus modulates $v(=\partial\eta/\partial\tau)$. The governing equation for the membrane vibration is

$$M_m \frac{\partial^2 \eta}{\partial \tau^2} = T \frac{\partial^2 \eta}{\partial y_1^2} - D \frac{\partial \eta}{\partial \tau} - \Delta p, \quad (2)$$

where Δp is the fluid pressure differential across the membrane. The fluid pressure above the membrane, p^+ , can be obtained from the linearized Bernoulli relationship²¹

$$p^+ = -\rho_0 \left(\frac{\partial \phi}{\partial \tau} + U \frac{\partial \phi}{\partial y_1} \right), \quad (3)$$

where ρ_0 is the fluid density and ϕ the incompressible velocity potential on the upper surface of the membrane:

$$\phi = -\frac{\Gamma}{2\pi} \tan^{-1} \left[\frac{2y_{2,\omega}(y_1 - y_{1,\omega})}{(y_1 - y_{1,\omega})^2 + (\eta^2 - y_{2,\omega}^2)} \right] + \frac{1}{\pi} \int_{-L/2}^{L/2} \frac{\partial \eta'}{\partial \tau} \log \sqrt{(y_1 - y_1')^2 + (\eta - \eta')^2} dy_1' + U y_1, \quad (4)$$

where the prime denotes a quantity along the membrane. The second term on the right-hand side of Eq. (4) can be regarded as a kind of fluid loading,²² but can be neglected in the present study as long as η is kept small and the vibration frequency is low. While fluid compressibility is not important above the membrane because of the much stronger hydrodynamic pressure resulting from the unsteady vortex motions and membrane vibrations, it is the only mean which creates pressure fluctuations inside the cavity. It will be shown later that the frequency of the membrane vibration is much lower than the first eigenmode frequency of the cavity, so that the fluid pressure inside the cavity can be assumed to be uniform. One can then write

$$p^- = c^2 \Delta \rho = -\frac{c^2 \rho_0}{Lh} \int_{-L/2}^{L/2} \eta dy_1 \quad (5)$$

and $\Delta p = p^+ - p^-$. The motions of the vortex and the membrane can be obtained by time integration of Eqs. (1) and (2) using a fourth-order Runge-Kutta procedure.

B. Acoustic far field

At a point \mathbf{X} fixed in the fluid and thus moving with the local fluid velocity in the far field, the quadrupole in the solutions of acoustic analogy^{1,2} at low Mach number M ($M = U/c$, where c is the ambient speed of sound) can be ignored so that the acoustic pressure at \mathbf{X} is

$$p(\mathbf{X}, t) = \frac{\rho}{2\pi} \left[\frac{\partial}{\partial t} \int_S \frac{\mathbf{u} \cdot \mathbf{n}}{r} dS(\mathbf{Y}) - \nabla_{\mathbf{x}} \cdot \int_S (\rho \mathbf{u}(\mathbf{u} \cdot \mathbf{n}) + p \mathbf{n}) \frac{dS(\mathbf{Y})}{r} \right], \quad (6)$$

where \mathbf{Y} is a point in the near flow field, $r = |\mathbf{X} - \mathbf{Y}|$, \mathbf{n} is the outward normal of the surface S , and the integrands are evaluated at the retarded time $t - r/c$. The first term on the right-hand side of Eq. (6) is due to the membrane vibration. Following the result of Ffowcs Williams and Lovely,⁶ which

is for a compact point source, this term can be evaluated by an integration along the spanwise direction of the system:

$$p_{\text{vib}}(\mathbf{X}, t) = \frac{\rho}{2\pi} \int_{-\infty}^{\infty} \frac{1}{|\mathbf{X} - \mathbf{Y}|(1 + M \cos \theta')^3} \frac{\partial^2}{\partial \tau^2} \times \int_{-L/2}^{L/2} \eta(\mathbf{y}, t - |\mathbf{X} - \mathbf{Y}|/c) dy_1 dy_3 + O(M^2), \quad (7)$$

and

$$\cos \theta' = \hat{\mathbf{Y}}_1 \cdot \frac{\mathbf{X} - \mathbf{Y}}{|\mathbf{X} - \mathbf{Y}|} = \frac{X_1 - Y_1}{r}. \quad (8)$$

The second term on the right-hand side of Eq. (6) represents the sound radiation due to the fluctuating forces and rate of momentum transfer at the vibrating membrane boundary, which must equal the force that gives rise to the unsteady motions of the vortex.²¹ The force per unit spanwise length, \mathbf{F} , acting on the fluid in the present two-dimensional system is thus

$$\mathbf{F} = \int_{-L/2}^{L/2} [\rho \mathbf{u}(\mathbf{u} \cdot \mathbf{n}) + p \mathbf{n}] dy_1 = \rho \Gamma \hat{\mathbf{y}}_3 \times (v_\omega - v'_\omega), \quad (9)$$

where v'_ω is the velocity of the image vortex.²³ The far field acoustic contribution from this force is

$$p_{\text{force}}(\mathbf{X}, t) = \frac{\rho}{2\pi c} \int_{-\infty}^{\infty} \frac{(\mathbf{X} - \mathbf{Y})}{|\mathbf{X} - \mathbf{Y}|^2 (1 + M \cos \theta')^2} \cdot \frac{\partial}{\partial \tau} \mathbf{F}(\mathbf{y}, t - |\mathbf{X} - \mathbf{Y}|/c) dy_3 + O(M^2). \quad (10)$$

As shown by Ffowcs Williams and Hawkins,²⁴ the integration along the direction y_3 in Eqs. (8) and (10) can be transformed into a time integration which is dominated by the contributions at $y_3 \rightarrow 0$. Therefore,

$$p(\mathbf{X}, t) = \frac{\rho}{\pi \sqrt{2R} (1 + M \cos \theta)^2} \left\{ \int_{-\infty}^{t-|\mathbf{X}|/c} \frac{1}{\sqrt{c(t-\tau) - R}} \times \left[\frac{1}{(1 + M \cos \theta)} \int_{-L/2}^{L/2} \frac{\partial^2 \eta}{\partial \tau^2} dy_1 + \frac{1}{c} \frac{\partial \mathbf{F}}{\partial \tau} \cdot \frac{\mathbf{X}}{|\mathbf{X}|} \right] d\tau \right\}, \quad (11)$$

where $R = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ and $\cos \theta = (x_1 - y_1)/R$. The line integral in Eq. (11) can be performed by a substitution depicted in Tang and Ko.²⁵

III. RESULTS AND DISCUSSIONS

In the present study, the mean flow is introduced abruptly into the system which is initially at rest. This results in an initial swelling of the membrane because of the mean fluid pressure difference between the near flow field and the cavity. The vortex path deflects away from the rigid plane at the same time. The air pressure on both sides of the membrane is not equalized initially in the present study because the flow speed inside a practical ventilation duct can vary according to the ventilation requirement and/or heating/

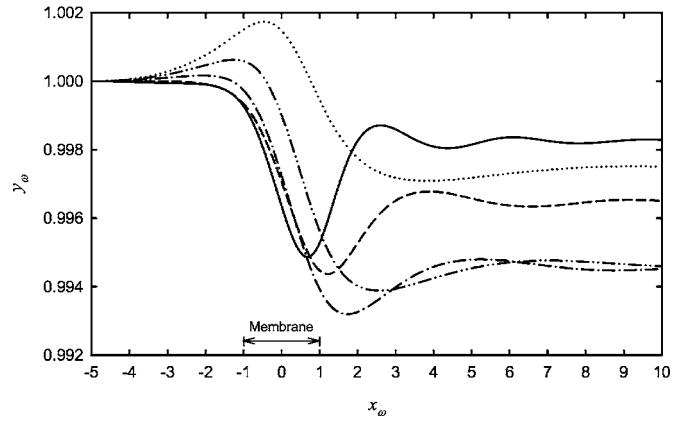


FIG. 2. Effects of mean flow velocity on vortex paths at $c_m=1$. (—) $U=0$; (---) $U=0.5$; (-·-) $U=1$; (---) $U=1.5$; and (···) $U=2$.

cooling load inside a centrally ventilated building automatically within operation hours. Therefore, the associated equalization of air pressure will require some kind of control system to be in place. Also, the flow is driven by a fan which should be started up from rest.

All length scales in the present study are normalized by d and all velocities by the initial vortex speed due to self-induction, which is $U_i = \Gamma/4\pi d$. The tension per unit length, T , the damping coefficient D , and the mass density of the membrane M_m are normalized by $\rho U_i^2 d$, ρU_i , and ρd , respectively. Since the structural damping within the membrane is in practice very weak, its effects are not discussed here. In the literature, for instance Frendi *et al.*,²⁶ D varies between 0.5 and 1 which refers to very weak damping condition.¹⁸ D is fixed at unity throughout the present investigation. The *in vacuo* wave speed along the membrane is $c_m = \sqrt{T/M_m}$, which is again normalized by U_i . To ensure a low Mach number condition, U_i is fixed at $0.1c$. Huang⁸ shows using a real material example that c_m can be set to $0.1c$ and thus the present range for c_m is chosen to be around 1 (maximum 2).

In reality, the vortex circulation and the mean flow speed are two related quantities as discussed by Lau and Tang²⁷ with a formula given in Saffman.²⁸ The general conclusion is that $U \sim 1$. However, since the formula in Saffman²⁸ is a general approximation, a range of U should be allowed in the present study. More detailed discussions on the relationship between the mean flow and the vortex circulation can be found in Lau and Tang.²⁷

The magnitude of the membrane deflection is kept below 1% of d (and $h=1$) throughout the present study. Given the very small degree of membrane stretching (maximum 0.5%) and that the membrane should be stretched to a great extent to maintain the tension, the effect of the tension variation should be insignificant. Such assumption was basically justified by the experiments of Choy and Huang²⁹ within engineering tolerance.

A. Vortex paths and membrane vibration

The flight paths of the vortex under different mean flow U with $c_m=1$ are shown in Fig. 2. The vortex moves toward the membrane due to the finite impedance of the latter.³⁰ The maximum transverse deviation from the original vortex po-

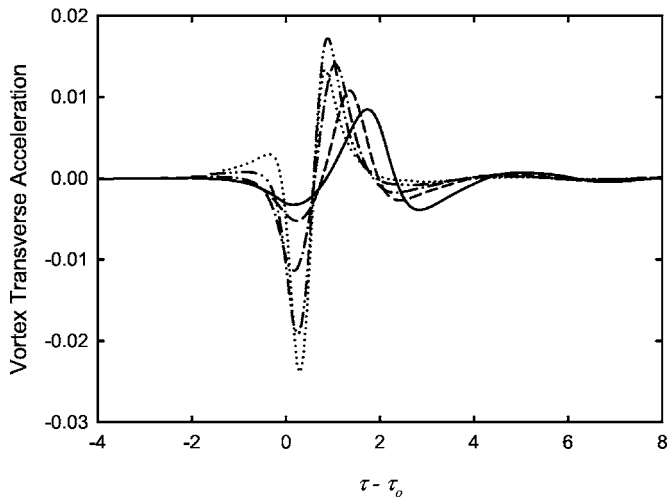


FIG. 3. Time variations of vortex transverse accelerations for $c_m=1$. Legends: same as those for Fig. 2.

sition increases with U for $U \leq 1$ and decreases upon further increase of U . The mean flow introduces a mean pressure difference on the two sides of the membrane. This tends to create an initial swelling of the membrane and thus the vortex paths, especially for strong U . The maximum $y_{2,\omega}$ increases monotonically with U because of the mean pressure difference induced by U mentioned before. This maximum is attained before the vortex traverses over the centerline of the membrane ($y_1=0$).

The relatively strong vortex accelerations at the instants the vortex reaches the leading and trailing edges of the membrane shown in Fig. 3 suggest significant vortex dipole sound is radiated only when the vortex is moving over the membrane. τ_0 denotes the instant the vortex is directly above the leading edge of the membrane. The instant the vortex reaches the trailing edge of the membrane is roughly $\tau_0 + 2/(U + U_i)$. The magnitude of the earlier acceleration peak increases with U , but the later one increases with U only up to $U \sim 1.5$

where the upward movement of the vortex is less affected by the restoring force of the membrane and the rigid edge at $y_1=-1$ as such high U tends to wash the vortex faster downstream, reducing the interaction time between the vortex and the membrane system. One can notice this also from Fig. 2, which illustrates that the vortex attains its minimum transverse position at increased y_1 as U increases. At $U=2$, this position is reached even when the vortex is downstream of the membrane.

The time variations of the membrane vibration velocity ($\partial\eta/\partial\tau$) at $U=1.25$ and 2 with $c_m=1$ are shown in Figs. 4(a) and 4(b), respectively. Maximum membrane/vortex path deflection occurs at $U \sim 1$ to 1.25 . The corresponding patterns for $U < 1.25$ resemble that of Fig. 4(a) with smaller magnitudes and thus are not presented. At $U=2$, the macroscopic vibration velocity pattern still looks similar to that at $U=1.25$, but it has become irregular with steeper variations. One can notice from Fig. 4 that membrane vibration frequency is very low. The highest vibration frequency in the present study occurs at $U=3$, $c_m=2$ which equals $0.518d/U_i$ and is roughly one-fifth of the lowest cavity eigenmode frequency ($5U_i/L$). The recent results of Huang³¹ confirm the validity of the present assumption of uniform air pressure inside the cavity at such low frequency ratio [Eq. (5)].

An increase in c_m does not bring about very significant change in the vortex paths as shown in Fig. 5(a). However, at a high $c_m=2$, one can observe a minor dip in the minimum y_ω at $U \sim 0.5$ but the lowest y_ω appears at $2.5 < U < 3$. The relatively strong membrane tension in this case limits the upward motion of the membrane due to the mean fluid pressure difference under the stronger mean flow. Computation is stopped at $U=3$, which corresponds to a Mach number U/c of 0.3 ; a value close to the limit of the acoustic analogy. Figure 5(b) illustrates the vortex path at $c_m=1/\sqrt{2}$. The weaker tension allows a larger upward stretch of the membrane. The lowest y_ω in this case appears at a U less than c_m .

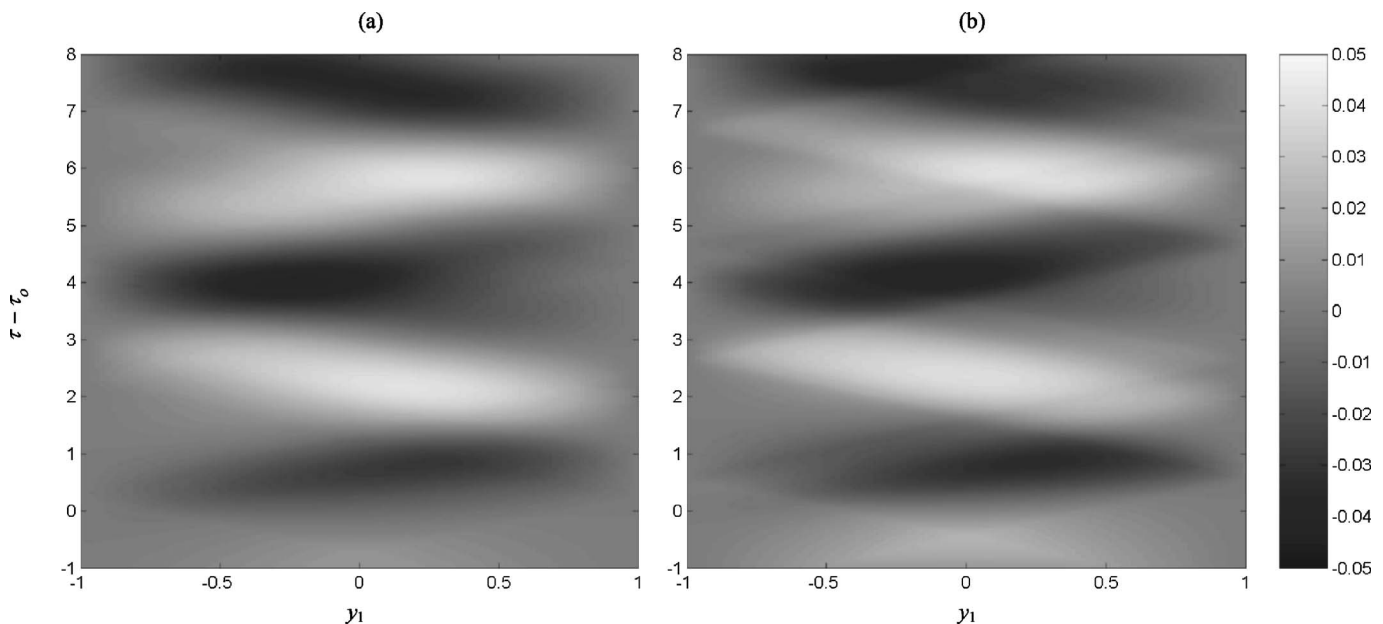


FIG. 4. Membrane vibration velocity patterns for $c_m=1$. (a) $U=1.25$ and (b) $U=2$.

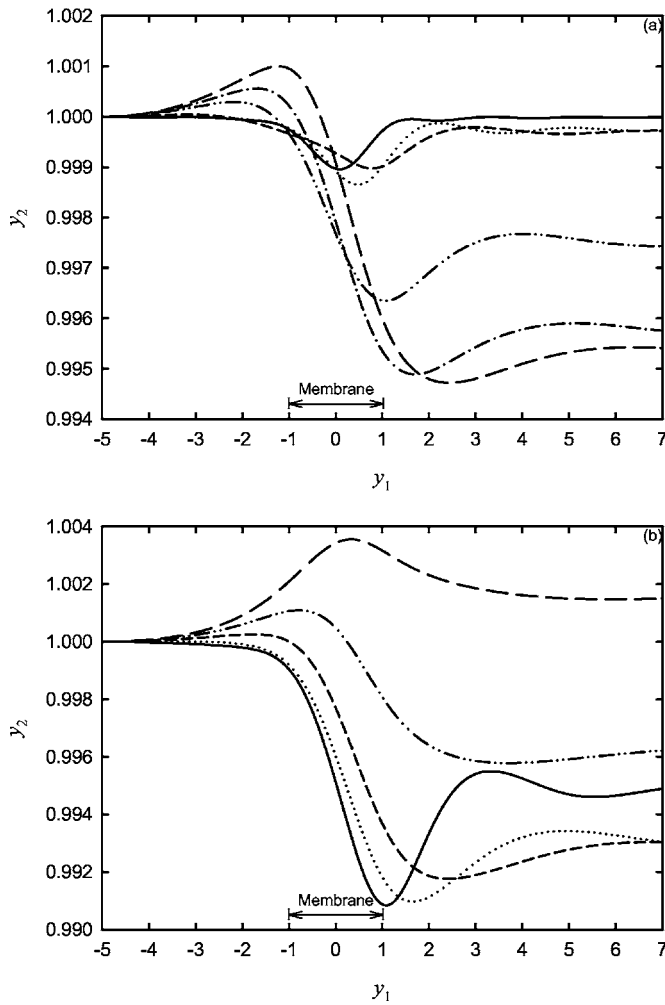


FIG. 5. Vortex paths at different c_m , (a) $c_m=2$ and (b) $c_m=1/\sqrt{2}$.

The pattern of the vortex path variation with U shown in Fig. 5(b) is similar to those at $U > c_m$ presented in Figs. 2 and 5(a).

The results shown in Fig. 6 suggest that the lowest $y_{2,\omega}$ is found at $U+0.5U_i \approx 1.63c_m$ for $c_m \geq 1$. There is also a

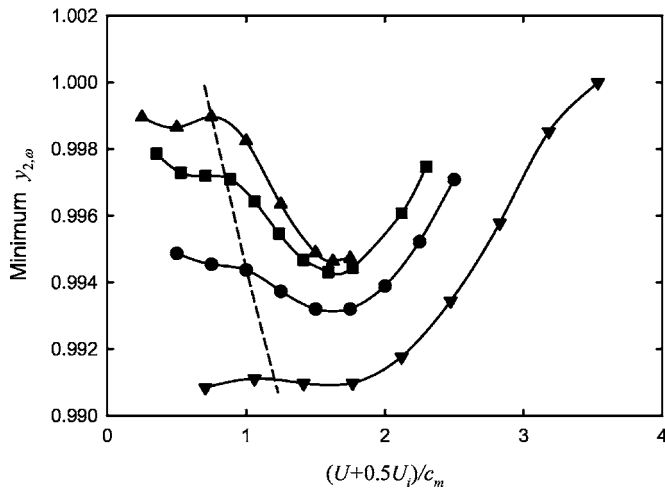


FIG. 6. Relationship between minimum vortex height and flow speed. (---) Approximate location of $U/c_m=0.5$; (∇) $c_m=1/\sqrt{2}$; (\bullet) $c_m=1$; (\blacksquare) $c_m=\sqrt{2}$; and (\blacktriangle) $c_m=2$.

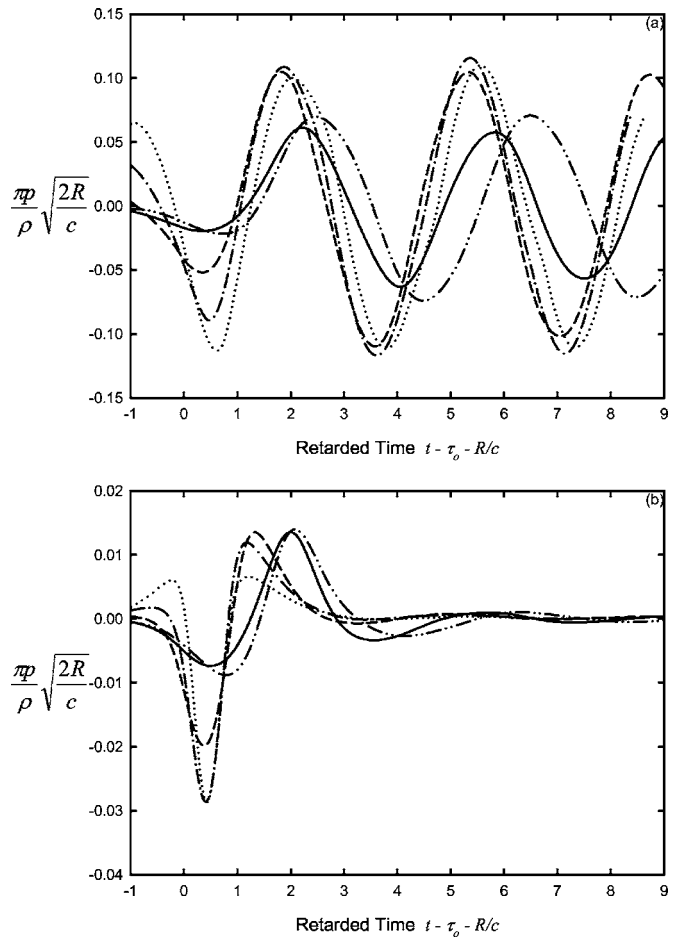


FIG. 7. Time variations of sound field strengths for $c_m=1$. (a) Monopole and (b) dipole. (—) $U=0$; (---) $U=1$; (-·-) $U=1.5$; (···) $U=2$; (---) $U=0$ (without cavity)— Ref. 16.

local minimum at a similar location for $c_m=1/\sqrt{2}$. Without the mean flow and the cavity, the lowest $y_{2,\omega}$ decreases monotonically with increasing c_m .¹⁶ For small U , the variation of the lowest $y_{2,\omega}$ is not significant, but it drops for $U/c_m > 0.5$ before a minimum is reached with a rate which increases with increasing c_m . The increase in the fluid pressure due to the vortex motion depends on the vortex speed and thus is partially related to U , but the latter reduces the mean fluid pressure above the membrane relative to that within the cavity counteracting the former effect. The effect of the mean pressure difference predominates at $U+0.5U_i \approx 1.63c_m$. The value 1.63 is believed to be associated with the volume of the cavity. It is left to further investigation.

B. The acoustic far field

The far field acoustic pressure fluctuations for $c_m=1$ are presented in Fig. 7 with the Doppler factor $(1+M \cos \theta)$ excluded. These pressure fluctuations are normalized in the same way as in Tang *et al.*¹⁸ The corresponding results of Tang *et al.*¹⁸ are included for the sake of easy comparison. The monopole radiations are shown in Fig. 7(a). The presence of the cavity below the membrane weakens the monopole radiation but increases the frequency of the radiation. The latter is expected as the cavity adds to the stiffness of the membrane. The increase in U produces a louder initial sound

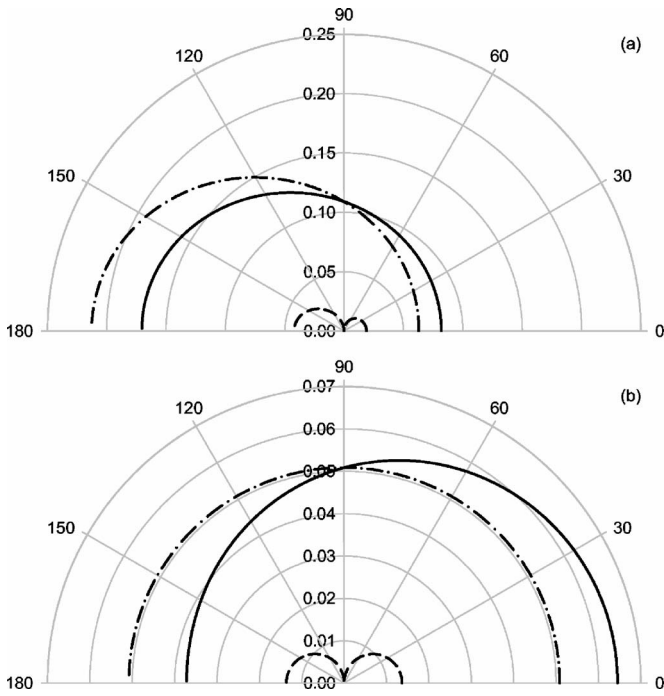


FIG. 8. Directivity of overall sound radiation at instants of strongest dipole radiation. (a) $U=2$ and (b) $U=0$. $c_m=1$. (—) Overall radiation; (---) monopole; and (- - -) dipole.

radiation. However, the corresponding sound amplitude peaks at around $U \sim 1.5$, probably due to the more dominant effect of the mean pressure difference between the flow and the cavity.

The corresponding dipole strengths are shown in Fig. 7(b). One can observe the eventual gain in dominance of the negative trough as U increases. The magnitude of the positive peak of this dipole time variation does not depend much on U for $U < 1$. Though the Doppler factor tends to attenuate more for the monopole radiation for $\theta < \pi/2$ as indicated in Eq. (11), the magnitude of the monopole remains more than three times that of the dipole at $\theta=0$ for $U=2$ ($M=0.2$). The corresponding ratio is ~ 3.7 in the absence of the mean flow. The increase in U shortens the period of active dipole radiation as expected.

Figure 7 indicates that the monopole and the dipole are nearly in phase (of the same sign) when the vortex is traversing in the proximity of the membrane for $0 \leq t - \tau_0 - R/c \leq 1$. Therefore, these two sound fields will be counteracting with each other for $\theta > \pi/2$. Figure 8(a) illustrates the directivity of sound radiation at $t - \tau_0 - R/c = 0.5$ for $U=2$, $c_m=1$, which corresponds to the instant of strongest dipole radiation [Fig. 7(b)]. The combined action of the two sound fields tends to reduce the effect of the Doppler factor. The corresponding sound fields in the absence of the mean flow at the instant of strongest dipole radiation are presented in Fig. 8(b). Without the mean flow, a downstream biased radiation as in Tang *et al.*¹⁸ is observed. The results with $U=1$ ($M=0.1$) still show a slightly downstream biased overall sound radiation (not presented here). Outside the active dipole radiation period, the monopole results in an upstream biased radiation. One can also notice the very strong convective amplification of the monopole in the presence of the mean

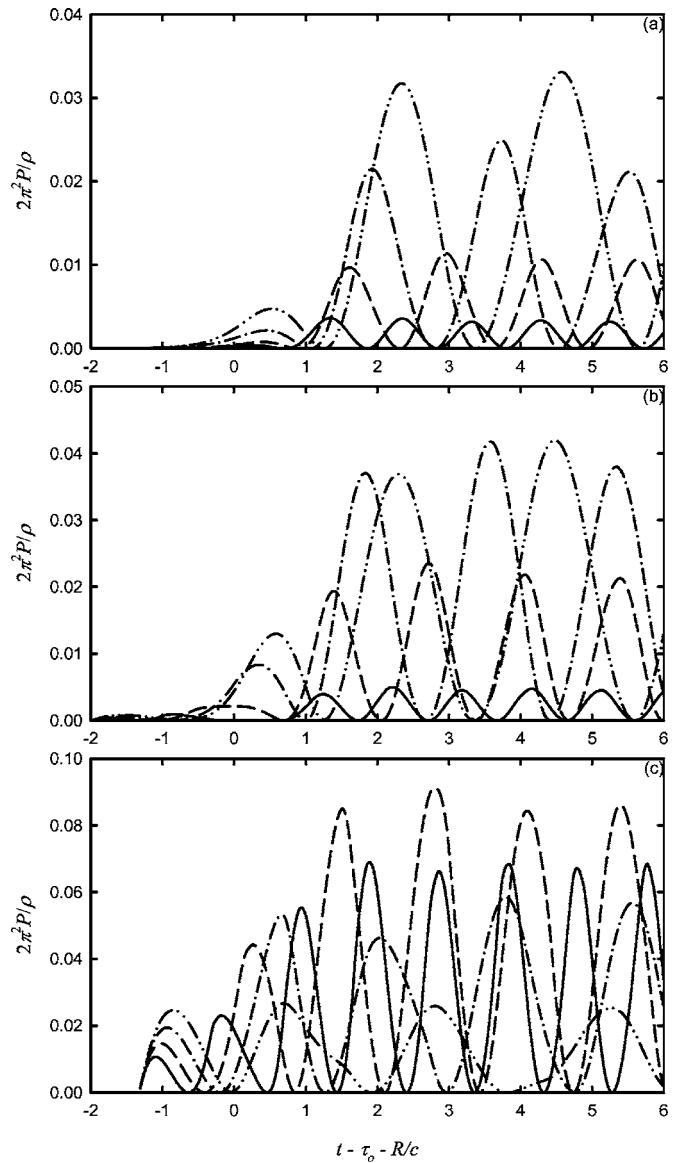


FIG. 9. Time variation of acoustic power radiation. (a) $U=0.5$; (b) $U=1$; and (c) $U=2$. (—) $c_m=2$; (---) $c_m=\sqrt{2}$; (- - -) $c_m=1$; and (- - -) $c_m=1/\sqrt{2}$.

flow, such that the overall sound power radiation is dominated by this monopole, which results from the vortex induced vibration of the membrane. Similar data variations as in Figs. 7 and 8 are observed at other values of c_m and thus the corresponding results are not presented.

The acoustic power radiated per unit spanwise length of the membrane, P , is summarized in Fig. 9. P is obtained by integrating $p^2(\mathbf{X}, t)/\rho c$ [Eq. (11)] over a distant cylindrical surface fixed in space with radius $|\mathbf{X}|$. For $U=0.5$, the acoustic power increases with decreasing c_m [Fig. 9(a)], but the trend is reversing gradually as U increases as illustrated in Figs. 9(b) and 9(c). For $c_m \geq 1$, the acoustic power increases with increasing U but the rate of increase depends on c_m . For $c_m=1$, the increase in the acoustic power radiated due to the increase of U toward 2 is limited. The rate of acoustic power increase is impressive for $c_m=2$ when U increases beyond 1.

The variation of the radiated acoustic power magnitude with U is nearly negligible for $c_m=1/\sqrt{2}$ and a reduction is

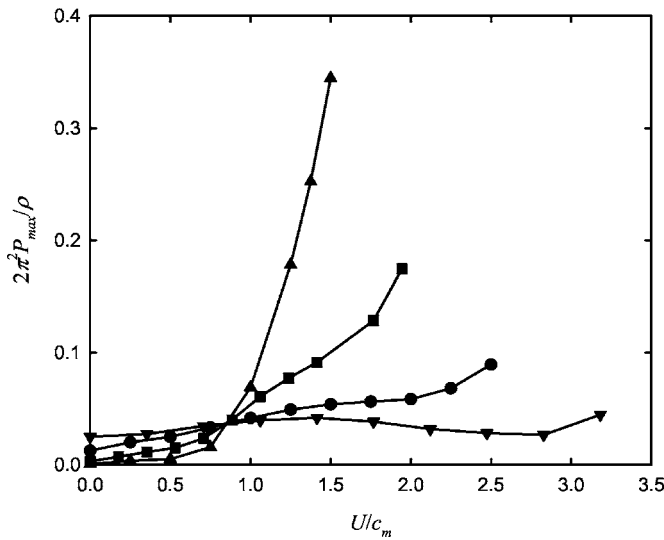


FIG. 10. Maximum radiated acoustic power under different combinations of U and c_m . (∇) $c_m=1/\sqrt{2}$; (\bullet) $c_m=1$; (\blacksquare) $c_m=\sqrt{2}$; and (\blacktriangle) $c_m=2$.

observed as U increases toward 2. Both the membrane vibration and the vortex acceleration are significantly weakened at high U because of the large mean fluid pressure difference between the flow and the cavity. Such weakening is more significant than the amplification of the monopole and the dipole in the presence of the mean flow. There is thus a higher chance of occurrence of a downstream biased radiation at small c_m .

Figure 10 illustrates in detail the maximum acoustic power radiated, P_{\max} , under the combined effects of U and c_m . P_{\max} increases with decreasing c_m for $U/c_m < 0.8$, but the opposite is found otherwise. This $U/c_m \sim 0.8$ is also the location where the negative trough of the dipole strength starts to be more dominant in the overall dipole power radiation. For $U/c_m \leq \sqrt{2}$, the rate of increase of P_{\max} tends to decrease with increasing U/c_m before a maximum or an inflexion point is found on the curves. For a weak membrane-cavity system, a reduction in P_{\max} is observed for a range of U/c_m , but this power increases again when U is kept increasing toward 3 ($M=0.3$), which is the upper limit for the acoustic analogy. The later rise in P_{\max} is due to the strong mean fluid pressure difference between the flow and the cavity leading to the very early but high membrane and vortex upward accelerations (not shown here). It is believed that this phenomenon occurs even for membrane of high wave speed.

It is found that the vortex dipole in the presence of the mean flow has adverse effects on the overall acoustic power radiation (Fig. (11)). Apart from the case for $U=0$ where the vortex dipole adds to the overall radiation, this dipole tends to counteract with the effect of the convective amplification in the presence of the mean flow. Also, the stronger the mean flow, the stronger such cancellation will be regardless of the value of c_m . The contribution from the dipole is insignificant for $U=0$, $c_m=2$ as shown in Fig. 11(b).

IV. CONCLUSIONS

The effects of a low Mach number mean flow on the sound radiated by the unsteady motion of a vortex moving in

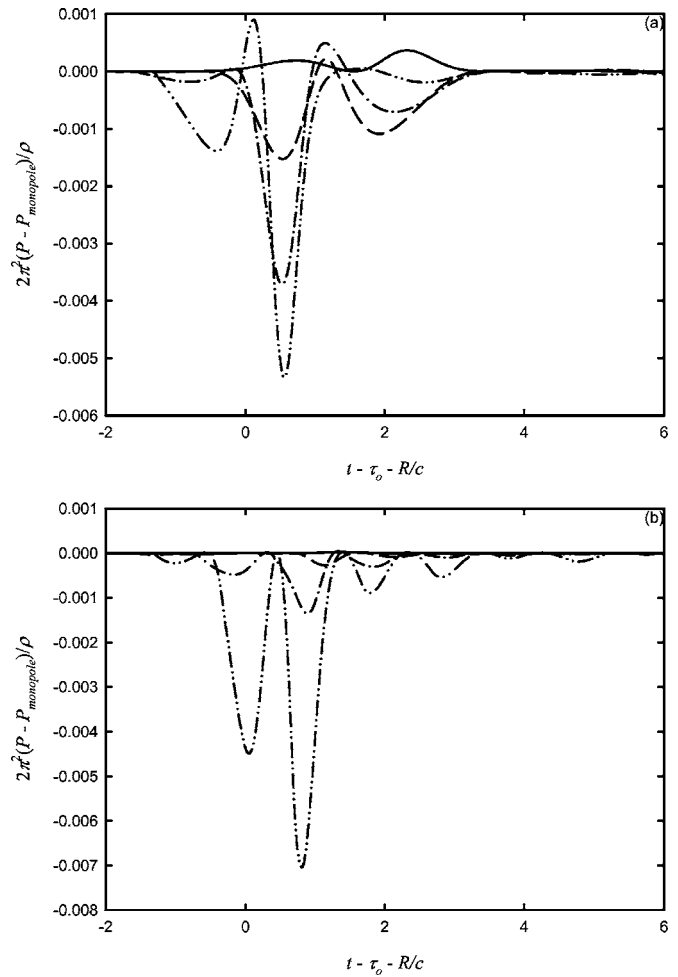


FIG. 11. Contributions from vortex dipole to acoustic power radiation. (a) $c_m=1/\sqrt{2}$ and (b) $c_m=2$. (---) $U=0$; (- - -) $U=1$; (- · - ·) $U=1.5$; and (· · ·) $U=2$.

the proximity of a finite length flexible membrane backed by an airtight cavity on an otherwise rigid plane are investigated in detail theoretically in the present study.

The vortex moves toward the flexible membrane transversely as it propagates over the leading edge of the membrane and then moves upwards as it traverses over the membrane trailing edge when the mean flow is not high. The mean flow results in a higher vortex longitudinal speed which increases the fluid pressure acting on the membrane. However, the presence of the mean flow and the cavity also gives rise to a mean fluid pressure difference between the two sides of the membrane, which tends to reduce the vibration magnitude of the membrane when the mean flow becomes high, especially for a weak membrane. The former factor is dominant in low mean flow speed cases, but the latter takes over when the mean flow exceeds a certain threshold depending on the wave speed of the membrane. The introduction of the cavity in general reduces the magnitude of the vortex induced membrane vibration and thus weakens the sound radiation. However, it is found that such effect does not vary monotonically with the flow speed because of the two counteracting forces stated earlier.

The vortex motion generates a dipole which propagates to the far field together with the monopole created by the

membrane vibration. The former is usually an order weaker than the latter. These sound fields counteract with each other in the upstream radiation directions. The mean flow results in amplifications of both sound fields, but in general, the overall effect is that the monopole receives a higher amplification due to the convective effect and the stronger vibration of the membrane. The dipole is only important in the period when the vortex is in close proximity of the membrane. Under some circumstances, the dipole strength may be strong enough to cause a fluctuation in the overall radiation directivity. However, the increase in the mean flow speed does not strengthen the overall acoustic power radiation when the membrane system is weak compared to the mean flow momentum. It is observed that an attenuation of the acoustic power radiation is possible when the mean flow speed is high compared to the membrane wave speed where the weakening of the membrane vibration cannot be compensated by the amplification resulting from the mean flow in the presence of the counteracting vortex dipole.

ACKNOWLEDGMENT

Funding support under Grant No. PolyU1/02C from the Research Grants Council of the Government of the HKSAR is gratefully acknowledged.

- ¹M. J. Lighthill, "On sound generated aerodynamically I. General theory," *Proc. R. Soc. London, Ser. A* **211**, 564–587 (1952).
- ²N. Curle, "The influence of solid boundaries upon aerodynamic sound," *Proc. R. Soc. London, Ser. A* **231**, 505–514 (1955).
- ³L. L. Beranek, *Noise and Vibration Control, Principles and Applications* (Wiley, New York, 1992).
- ⁴L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).
- ⁵J. E. Ffowcs Williams and D. J. Lovely, "Sound radiation into uniformly flowing fluid by compact surface vibration," *J. Fluid Mech.* **71**, 689–700 (1975).
- ⁶A. Dowling, "Convective amplification of real simple sources," *J. Fluid Mech.* **74**, 529–546 (1976).
- ⁷M. S. Howe, "Mechanism of sound generation by low Mach number flow over a wall cavity," *J. Sound Vib.* **273**, 103–123 (2004).
- ⁸L. Huang, "A theoretical study of duct noise control by flexible panels," *J. Acoust. Soc. Am.* **106**, 1801–1809 (1999).
- ⁹M. R. Visbal and D. V. Gaitonde, "Very high-order spatially implicit schemes for computational acoustics on curvilinear meshes," *J. Comput.*

- Acoust.* **9**, 1259–1286 (2001).
- ¹⁰R. C. K. Leung, X. M. Li, and R. M. C. So, "Comparative study of nonreflecting boundary condition for one-step duct aeroacoustic simulation," *AIAA J.* **44**, 664–667 (2006).
- ¹¹H. G. Davies, "Sound from turbulent boundary layer excited panels," *J. Acoust. Soc. Am.* **49**, 878–889 (1971).
- ¹²S. F. Wu and L. Maestrello, "Responses of finite baffled plate to turbulent flow excitations," *AIAA J.* **33**, 13–19 (1995).
- ¹³E. H. Dowell, "Radiation from panels as a source of airframe noise," *AIAA J.* **13**, 1529–1530 (1975).
- ¹⁴D. G. Crighton, "Radiation from vortex filament motion near a half-plane," *J. Fluid Mech.* **51**, 357–362 (1972).
- ¹⁵M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, Cambridge, 2003).
- ¹⁶R. C. K. Leung and R. M. C. So, "Noise generation of blade-vortex resonance," *J. Sound Vib.* **245**, 217–237 (2001).
- ¹⁷S. K. Tang and C. K. Lau, "Vortex sound in the presence of a wedge with inhomogeneous surface flow impedance," *J. Sound Vib.* **281**, 1077–1091 (2005).
- ¹⁸S. K. Tang, R. C. K. Leung, R. M. C. So, and K. M. Lam, "Acoustic radiation by vortex induced flexible wall vibration," *J. Acoust. Soc. Am.* **118**, 2182–2189 (2005).
- ¹⁹D. G. Crighton, "Basic principles of aerodynamic noise generation," *Prog. Aerosp. Sci.* **16**, 31–96 (1975).
- ²⁰H. R. Vallentine, *Applied Hydrodynamics* (Butterworth, London, 1969).
- ²¹N. Peake, "On the unsteady motion of a long fluid-loaded elastic plate with mean flow," *J. Fluid Mech.* **507**, 335–366 (2004).
- ²²P. J. T. Filippi, O. Lagarrigue, and P.-O. Mattei, "Perturbation method for sound radiation by a vibrating plate in a light fluid: Comparison with the exact solution," *J. Sound Vib.* **177**, 259–275 (1994).
- ²³S. K. Tang and J. E. Ffowcs Williams, "Acoustic radiation from a vortex approaching a circular cylinder with surface suction," *Acust. Acta Acust.* **84**, 1007–1013 (1998).
- ²⁴J. E. Ffowcs Williams and D. L. Hawkings, "Shallow water wave generation by unsteady flow," *J. Fluid Mech.* **31**, 779–788 (1968).
- ²⁵S. K. Tang and N. W. M. Ko, "Sound generation by interaction of two inviscid two-dimensional vortices," *J. Acoust. Soc. Am.* **102**, 1463–1473 (1997).
- ²⁶A. Frendi, L. Maestrello, and A. Bayliss, "Coupling between plate vibration and acoustic radiation," *J. Sound Vib.* **177**, 207–226 (1994).
- ²⁷C. K. Lau and S. K. Tang, "Sound generated by vortices in the presence of a porous half-cylinder mounted on a rigid plane," *J. Acoust. Soc. Am.* **119**, 2084–2095 (2005).
- ²⁸P. G. Saffman, *Vortex Dynamics* (Cambridge University Press, Cambridge, UK, 1992).
- ²⁹Y. S. Choy and L. Huang, "Effect of flow on the drumlike silencer," *J. Acoust. Soc. Am.* **118**, 3077–3085 (2005).
- ³⁰S. K. Tang, "Effects of porous boundaries on the dynamics of an inviscid vortex filament," *Q. J. Mech. Appl. Math.* **54**, 65–84 (2001).
- ³¹L. Huang, "Modal analysis of a drumlike silencer," *J. Acoust. Soc. Am.* **112**, 2014–2025 (2002).

A computational method of evaluating noncompact sound based on vortex sound theory

T. Takaishi^{a)}

Railway Technical Research Institute, Shiga, 521-0013, Japan

M. Miyazawa and C. Kato

Institute of Industrial Science, University of Tokyo, Tokyo, 153-8505, Japan

(Received 10 December 2006; revised 12 December 2006; accepted 12 December 2006)

A numerical investigation is made of the production of sound by turbulence interacting with a noncompact body. The problem is formulated in the frequency domain by extending the theory of vortex sound proposed by Howe. The anomalous “numerical” generation of sound by the sudden termination of Lighthill’s stress tensor at the outer boundary of a finite computational domain is avoided by identification of “scattered” sound sources that generate sound principally by interaction with the solid surface. It is argued that the boundary element method is the most efficient means of computing the aeroacoustic Green’s function for the problem, because it requires a minimum of CPU time, is not prone to numerical errors such as dispersion and dissipation during propagation, and the radiation condition is easily applied at the outer boundary. The method is applied to the problem of sound generation by high Reynolds number flow past a circular cylinder. The “scattered” sources are shown to be confined to the vicinity of the cylinder surface. At low frequencies the radiation has a dipole-like directivity in agreement with the compact approximation. However, the directivity is quite different at high frequencies, where our noncompact method predicts a more complicated “leaf-like” radiation pattern.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2431345]

PACS number(s): 43.28.Ra, 43.50.Nm, 43.50.Lj [MSH]

Pages: 1353–1361

I. INTRODUCTION

Noise generated aerodynamically has become a dominant noise source for high-speed trains that are currently running at a maximum speed of 300 km/h (190 mile/h) in Japan. The aerodynamic noise is produced mainly by projecting devices such as pantographs and cable heads on the roof, gaps between cars, and cavities at the bogie sections.¹ This noise is proportional to the 6–8th power of the wind (vehicle) speed, and at a high speed it exceeds other noises such as a rolling noise from wheels and rails that are normally proportional to the third power of the vehicle speed. Since the Environmental Quality Standards established by the Japanese government stipulate that the peak noise level shall not exceed 75 dB(A) at tracksides, it is a crucial issue to reduce the aerodynamic noise radiated from trains. The overall objectives of this study are to investigate the mechanism through which aerodynamic noise is generated from such devices or irregularities, and to propose countermeasures to reduce the noise.

Although experiments in a low-noise wind tunnel are an effective way to examine countermeasures, various numerical techniques for the prediction of aeroacoustic noise have been proposed for understanding and reducing aeroacoustic sources, principally because of the remarkable improvement in a computer performance in recent years. Direct Navier-Stokes simulation (DNS) with high-order accuracy including compressibility can capture a flow field and the sound simul-

aneously, and will be a promising technique in the future.^{2,3} Hardin *et al.*⁴ presented another method that obtains the time-dependent incompressible flow first, and then solves the compressible Navier-Stokes equation for propagation of acoustic waves. Bailly *et al.*⁵ studied propagation of acoustic waves in a mean flow by using linearized Euler equations (LEE).

Low-Mach-number aeroacoustic problems are solved efficiently by employing Lighthill’s acoustic analogy. Hardin *et al.*⁶ calculated the sound radiation from a cylinder with the compact Green’s function proposed by Howe.⁷ The present authors⁸ proposed a method to evaluate dipole sound sources accurately in a finite computational domain. In their study, the balance of momentum in a finite computational domain was considered strictly, and the effect of those vortices that are outside the computational domain was replaced by a contribution from inner properties by using an auxiliary velocity potential φ_j . From an acoustical point of view, this procedure avoids the sudden termination of Lighthill’s stress tensor at the outer boundary and extracts the net contribution from surface scattered dipole sound sources.

The studies in Refs. 6–8 used the approximation of acoustical compactness—if the characteristic length of the source region is small compared to the wavelengths of the sound waves, differences of retarded time in the source region are negligible. Oberai *et al.*⁹ pointed out that in order to solve the acoustic problem without this restriction, one needs to treat the body surface as “sound-hard.” They determined Lighthill’s tensor by using large-eddy simulation (LES), and obtained the far-field acoustic pressure from a trailing edge

^{a)}Electronic mail: takaishi@rtri.or.jp

by coupling it with finite element method. Howe¹⁰ showed theoretically the possibility of evaluating noncompact problems from the incompressible flow properties.

In this study, we extend our former studies based on vortex sound theory into noncompact problems and propose a computational method for evaluating noncompact sound. In Sec. II, formulas for evaluating noncompact sound are summarized. The scattered sound is distinguished from the direct sound. In Sec. III, the technique of LES and an improvement in the algorithm are presented. The aeroacoustic noise generated by turbulent flow around a circular cylinder is evaluated numerically in Sec. IV. Noncompact Green's functions are obtained by using the boundary element method (BEM), and characteristics of the scattered sound are discussed in detail.

II. FORMULAS FOR EVALUATING NONCOMPACT SOUND

The theory of aerodynamic sound was first developed by Lighthill,¹¹ who reformulated the Navier-Stokes equation into an exact, inhomogeneous wave equation. Howe⁷ simplified Lighthill's equation for a high Reynolds number flow at a low Mach number, obtaining

$$\frac{1}{c_0^2} \frac{\partial^2 B(\mathbf{y}, \tau)}{\partial \tau^2} - \nabla_{\mathbf{y}}^2 B(\mathbf{y}, \tau) = \nabla_{\mathbf{y}} \cdot (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \tau), \quad (1)$$

where B is the total enthalpy

$$B \equiv \frac{p}{\rho_0} + \frac{1}{2} |\mathbf{u}|^2, \quad (2)$$

and $\boldsymbol{\omega} = \nabla_{\mathbf{y}} \times \mathbf{u}$ is the vorticity. Consider the fluid domain Ω_{all} whose inner boundary is the body surface S and outer boundary is the surface Σ_{inf} at infinity, which is large enough to contain all of the vorticity and where the flow is uniform. The body surface is assumed to be fixed, so that the velocity at S must vanish. If the observer at \mathbf{x} is far enough from the source region, Green's theorem yields the sound pressure in the form

$$p_a(\mathbf{x}, t) = -\rho_0 \int_{\Omega_{\text{all}}} (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \tau) \cdot \nabla_{\mathbf{y}} G(\mathbf{x}, \mathbf{y}, t - \tau) d\mathbf{y} d\tau + \eta \int_S \boldsymbol{\omega}(\mathbf{y}, \tau) \times \nabla_{\mathbf{y}} G(\mathbf{x}, \mathbf{y}, t - \tau) \cdot \mathbf{n} dS d\tau, \quad (3)$$

where the Green's function $G(\mathbf{x}, \mathbf{y}, t - \tau)$ is chosen to satisfy

$$\frac{1}{c_0^2} \frac{\partial^2 G(\mathbf{x}, \mathbf{y}, t - \tau)}{\partial \tau^2} - \nabla_{\mathbf{y}}^2 G(\mathbf{x}, \mathbf{y}, t - \tau) = \delta(\mathbf{x} - \mathbf{y}) \delta(t - \tau) \quad (4)$$

in the fluid domain and its normal derivative is 0 on S , namely,

$$\frac{\partial G(\mathbf{x}, \mathbf{y}, t - \tau)}{\partial y_i} n_i = 0 \quad \text{on } S. \quad (5)$$

The Fourier transform with respect to time transforms Eq. (3) into

$$p_a(\mathbf{x}, \omega) = -\rho_0 \int_{\Omega_{\text{all}}} (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \tau) \cdot \nabla_{\mathbf{y}} G(\mathbf{x}, \mathbf{y}, \omega) d\mathbf{y} d\tau + \eta \int_S \boldsymbol{\omega}(\mathbf{y}, \tau) \times \nabla_{\mathbf{y}} G(\mathbf{x}, \mathbf{y}, \omega) \cdot \mathbf{n} dS d\tau, \quad (6)$$

where ω is the angular frequency.

Here, Green's function $G(\mathbf{x}, \mathbf{y}, \omega)$ in the frequency domain satisfies

$$-\frac{\omega^2}{c_0^2} G(\mathbf{x}, \mathbf{y}, \omega) - \nabla_{\mathbf{y}}^2 G(\mathbf{x}, \mathbf{y}, \omega) = \delta(\mathbf{x} - \mathbf{y}), \quad (7)$$

$$\frac{\partial G(\mathbf{x}, \mathbf{y}, \omega)}{\partial y_i} n_i = 0 \quad \text{on } S. \quad (8)$$

The last terms of Eq. (3) and Eq. (6) are the contribution from viscosity, and it is nominally of order $1/\text{Re} \ll 1$ relative to the contribution from the volume vorticity.¹²

If the characteristic length of the source region is small compared to the wavelengths of the sound, differences of the retarded time in the source region are negligible, and the source is said to be acoustically compact. Then the three-dimensional Green's function may be expanded around the source region in the form

$$G_c(\mathbf{x}, \mathbf{y}, t) \equiv \frac{1}{4\pi|\mathbf{x}|} \delta\left(t - \tau - \frac{|\mathbf{x}|}{c_0}\right) + \frac{\mathbf{x} \cdot \mathbf{Y}}{4\pi c_0 |\mathbf{x}|^2} \frac{\partial}{\partial t} \delta\left(t - \tau - \frac{|\mathbf{x}|}{c_0}\right), \quad (9)$$

$$G_c(\mathbf{x}, \mathbf{y}, \omega) \equiv \frac{1}{4\pi|\mathbf{x}|} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right) - i\omega \frac{\mathbf{x} \cdot \mathbf{Y}}{4\pi c_0 |\mathbf{x}|^2} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right). \quad (10)$$

In these equations, Y_j denotes the velocity potential of an imaginary flow around the body that has unit speed in the j direction at large distances from the body. This velocity potential must satisfy

$$\frac{\partial Y_j(\mathbf{y})}{\partial y_i} n_i = 0 \quad (11)$$

on surfaces S and Laplace's equation

$$\nabla_{\mathbf{y}}^2 Y_j = 0 \quad (12)$$

in the fluid domain. By using the relation

$$Y_j = y_j - \varphi_j, \quad (13)$$

where φ_j is the velocity potential of an imaginary flow that would be produced by translational motion of the rigid body at the unit speed in the j direction, Eq. (10) is therefore

$$G_c(\mathbf{x}, \mathbf{y}, \omega) \equiv \frac{1}{4\pi|\mathbf{x}|} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right) - i\omega \frac{\mathbf{x} \cdot \mathbf{y}}{4\pi c_0 |\mathbf{x}|^2} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right) + i\omega \frac{\mathbf{x} \cdot \boldsymbol{\varphi}}{4\pi c_0 |\mathbf{x}|^2} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right). \quad (14)$$

The first and second terms of Eq. (14) represent a low-frequency approximation of the free space Green's function

$$G_0(\mathbf{x}, \mathbf{y}, \omega) \equiv \frac{1}{4\pi|\mathbf{x} - \mathbf{y}|} \exp\left\{i\omega \frac{|\mathbf{x} - \mathbf{y}|}{c_0}\right\} \quad (15)$$

that describes propagation of the direct sound from the turbulent flow. The third term of Eq. (14) represents propagation of the sound scattered at the body surface. Accordingly, we define

$$G_c^i(\mathbf{x}, \mathbf{y}, \omega) \equiv \frac{1}{4\pi|\mathbf{x}|} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right) - i\omega \frac{\mathbf{x} \cdot \mathbf{y}}{4\pi c_0 |\mathbf{x}|^2} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right), \quad (16)$$

$$G_c^s(\mathbf{x}, \mathbf{y}, \omega) \equiv i\omega \frac{\mathbf{x} \cdot \boldsymbol{\varphi}}{4\pi c_0 |\mathbf{x}|^2} \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right). \quad (17)$$

On the other hand, if the approximation of acoustical compactness is not satisfied, Eqs. (7) and (8) must be solved in some manner. Here the boundary element method (BEM) is applied to obtain Green's function for an arbitrary body shape.

Let $G_N(\mathbf{x}, \mathbf{y}, \omega)$ be the solution of Eq. (7) with the boundary condition Eq. (8). Green's function for the free space Eq. (15) is the fundamental solution of Eq. (7) and satisfies

$$-\frac{\omega^2}{c_0^2} G_0(\mathbf{x}, \mathbf{y}, \omega) - \nabla_y^2 G_0(\mathbf{x}, \mathbf{y}, \omega) = \delta(\mathbf{x} - \mathbf{y}). \quad (18)$$

Then Green's theorem yields

$$\begin{aligned} & \int_{\Omega} \{G_0(\mathbf{z}, \mathbf{y}, \omega) \nabla_z^2 G_N(\mathbf{x}, \mathbf{z}, \omega) - G_N(\mathbf{x}, \mathbf{z}, \omega) \nabla_z^2 G_0(\mathbf{z}, \mathbf{y}, \omega)\} d\mathbf{z} \\ &= - \int_S \left\{ G_0(\mathbf{z}, \mathbf{y}, \omega) \frac{\partial G_N(\mathbf{x}, \mathbf{z}, \omega)}{\partial z_i} - G_N(\mathbf{x}, \mathbf{z}, \omega) \frac{\partial G_0(\mathbf{z}, \mathbf{y}, \omega)}{\partial z_i} \right\} n_i dS(\mathbf{z}), \end{aligned} \quad (19)$$

where the unit normal vector n_i on S is directed into the fluid domain Ω_{all} . Substituting Eqs. (7), (8), and (18) into Eq. (19), we obtain

$$G_N(\mathbf{x}, \mathbf{y}, \omega) = G_0(\mathbf{x}, \mathbf{y}, \omega) + \int_S G_N(\mathbf{x}, \mathbf{z}, \omega) \frac{\partial G_0(\mathbf{z}, \mathbf{y}, \omega)}{\partial z_i} n_i dS(\mathbf{z}). \quad (20)$$

By bringing the point \mathbf{y} in the fluid close to the point \mathbf{z}_p on the boundary surface S , Eq. (20) becomes in the limit

$$G_N(\mathbf{x}, \mathbf{z}_p, \omega) \left(1 - \frac{\Phi(\mathbf{z}_p)}{4\pi}\right) = G_0(\mathbf{x}, \mathbf{z}_p, \omega) + \int_{S'} G_N(\mathbf{x}, \mathbf{z}, \omega) \times \frac{\partial G_0(\mathbf{z}, \mathbf{z}_p, \omega)}{\partial z_i} n_i dS(\mathbf{z}), \quad (21)$$

where $\Phi(\mathbf{z}_p)$ is the inner solid angle at \mathbf{z}_p and S' is the boundary surface excluding singular area around \mathbf{z}_p from S . Then the boundary S' in Eq. (21) is discretized into a number of curvilinear elements, and $G_N(\mathbf{x}, \mathbf{z}_p, \omega)$ on S is solved numerically. By substituting $G_N(\mathbf{x}, \mathbf{z}_p, \omega)$ into Eq. (20), the value of $G_N(\mathbf{x}, \mathbf{y}, \omega)$ at the arbitrary point \mathbf{y} in the fluid is obtained.

Equation (20) states that a point source located at an arbitrary point \mathbf{y} in the fluid is scattered at the body surface S , and then propagates to the observer at \mathbf{x} . The first term of Eq. (20) represents the propagation of the direct sound, and the second term represents the propagation of the sound scattered at the body surface.

Define

$$G^i(\mathbf{x}, \mathbf{y}, \omega) \equiv G_0(\mathbf{x}, \mathbf{y}, \omega), \quad (22)$$

$$G^s(\mathbf{x}, \mathbf{y}, \omega) \equiv \int_S G_N(\mathbf{x}, \mathbf{z}, \omega) \frac{\partial G_0(\mathbf{z}, \mathbf{y}, \omega)}{\partial z_i} n_i dS(\mathbf{z}), \quad (23)$$

then Eq. (20) is

$$G_N(\mathbf{x}, \mathbf{y}, \omega) = G^s(\mathbf{x}, \mathbf{y}, \omega) + G^i(\mathbf{x}, \mathbf{y}, \omega). \quad (24)$$

Then Eq. (6) without viscous term becomes

$$p_a(\mathbf{x}, \omega) = -\rho_0 \int_{\Omega_{\text{all}}} (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G^s(\mathbf{x}, \mathbf{y}, \omega) dy - \rho_0 \Omega_{\text{all}} (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G^i(\mathbf{x}, \mathbf{y}, \omega) dy. \quad (25)$$

Similarly, define

$$p_a^s(\mathbf{x}, \omega) \equiv -\rho_0 \int_{\Omega_{\text{all}}} (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G^s(\mathbf{x}, \mathbf{y}, \omega) dy, \quad (26)$$

$$p_a^i(\mathbf{x}, \omega) \equiv -\rho_0 \int_{\Omega_{\text{all}}} (\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G^i(\mathbf{x}, \mathbf{y}, \omega) dy. \quad (27)$$

Then the sound pressure at the observer \mathbf{x} becomes

$$p_a(\mathbf{x}, \omega) = p_a^s(\mathbf{x}, \omega) + p_a^i(\mathbf{x}, \omega), \quad (28)$$

where $p_a^s(\mathbf{x}, \omega)$ is the acoustic sound scattered at the surface S , and $p_a^i(\mathbf{x}, \omega)$ is the direct sound from the turbulence in the fluid.

The integrand in Eq. (26),

$$(\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G^s(\mathbf{x}, \mathbf{y}, \omega), \quad (29)$$

can be regarded as the source of the scattered sound, and the integrand in Eq. (27),

$$(\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G^i(\mathbf{x}, \mathbf{y}, \omega), \quad (30)$$

as the source of the direct sound.

Equation (27) can produce serious errors because of the sudden termination of Lighthill's stress tensor at the outer boundary of a finite computational domain.^{8,13} At low Mach number, since the radiation efficiency of sound scattered at the body surface is larger than the direct sound, Eq. (26) will extract the net contribution from the dominant sound sources. The computational grid should be small enough to resolve the boundary layer and bound vortices around the body, otherwise the nonslip condition at S is not satisfied and artificial sound called "a suction dipole" will be produced under slip condition.¹²

III. GOVERNING EQUATIONS AND NUMERICAL METHODS FOR THE FLOW-FIELD COMPUTATION

The source fluctuations in the flow are first computed by a large-eddy simulation (LES) of incompressible fluid using a dynamic Smagorinsky model,^{14,15} and they are fed as input data to the acoustical computation described in Sec. II.

A. Governing equations

The governing LES equations for an incompressible flow are the spatially filtered continuity equation and Navier-Stokes equations in the Cartesian coordinates

$$\frac{\partial \bar{u}_i}{\partial y_i} = 0, \quad (31)$$

$$\frac{\partial \bar{u}_i}{\partial \tau} + \frac{\partial}{\partial y_j} (\bar{u}_i \bar{u}_j + \tau_{ij}) = -\frac{1}{\rho_0} \frac{\partial \bar{p}}{\partial y_i} + \frac{\partial}{\partial y_j} \left\{ \nu \left(\frac{\partial \bar{u}_i}{\partial y_j} + \frac{\partial \bar{u}_j}{\partial y_i} \right) \right\}, \quad (32)$$

where $\bar{u}_i (i=1, 2, 3)$ is the grid scale velocity component, \bar{p} is the grid-scale static pressure, and ν is the kinematic viscosity. The subgrid-scale (SGS) stress tensor τ_{ij} is defined by

$$\tau_{ij} = \overline{u_i u_j} - \bar{u}_i \bar{u}_j. \quad (33)$$

The Smagorinsky closure is applied to the SGS stress τ_{ij} :

$$\tau_{ij} - \frac{1}{3} \delta_{ij} \tau_{kk} = -2\nu_{\text{SGS}} \bar{S}_{ij}, \quad (34)$$

where

$$\nu_{\text{SGS}} = (C_s \Delta)^2 |\bar{S}|, \quad |\bar{S}| = \sqrt{2 \bar{S}_{ij} \bar{S}_{ij}}, \quad \bar{S}_{ij} = \frac{1}{2} \left(\frac{\partial \bar{u}_i}{\partial y_j} + \frac{\partial \bar{u}_j}{\partial y_i} \right). \quad (35)$$

The quantity C_s is the Smagorinsky coefficient and Δ is the size of the grid filter. In the present study, the dynamic Smagorinsky model with modification due to Lilly¹⁶ is used, where C_s is determined locally in time and space by

$$C_s^2 = \frac{1}{2\Delta^2} \frac{\langle M_{ij} l_{ij} \rangle}{\langle M_{ij} M_{ij} \rangle}, \quad (36)$$

$$M_{ij} = \tilde{\beta}_{ij} - \alpha^2 |\tilde{S}| \tilde{S}_{ij}, \quad \alpha \equiv \frac{\tilde{\Delta}}{\Delta} = 2, \quad (37)$$

$$l_{ij} = (\tilde{\gamma}_{ij} - \tilde{u}_i \tilde{u}_j) - \frac{1}{3} \delta_{ij} (\tilde{\gamma}_{kk} - \tilde{u}_k \tilde{u}_k), \quad (38)$$

where $\beta_{ij} = |\bar{S}| \bar{S}_{ij}$ and $\gamma_{ij} = \bar{u}_i \bar{u}_j$. The symbol " \sim " represents a test-filtering operation, " $\langle \rangle$ " represents an averaging operation, and $\tilde{\Delta}$ is the size of the test-filtered width.

B. Finite element formulations

Both an explicit time-accurate streamline upwind scheme^{17,18} and a fully implicit Crank-Nicolson scheme have been used in the LES code for integrating the momentum equations with respect to time. The finite element method (FEM) is adopted for the spatial discretization. Both schemes have second-order accuracy in terms of time and space. For the Crank-Nicolson implicit scheme three to five Newton iterations are made within a time step in order to maintain second-order accuracy in time.

A fractional step (FS) method is used to solve the pressure Poisson equation. The momentum equations and the pressure Poisson equation are solved separately at each time step. The continuity Eq. (31) is satisfied through the solution of the pressure Poisson equation in the FS method. The rate of convergence of the FS method is higher than that of the arbitrary boundary marker and cell (ABMAC) method.¹⁴ The global linear system of equations that results from the pressure Poisson equation in the FS method is solved by the Bi-CGSTAB (Ref. 19) method combined with residual cutting method²⁰ as its outer loop.

IV. APPLICATION TO THE CIRCULAR CYLINDER

In this section, the aeroacoustic noise generated by turbulent flow past a circular cylinder is evaluated numerically using the method proposed in Secs. II and III. In order to validate the numerical results, an experiment was performed in a small-scale low-noise wind tunnel at the Railway Technical Research Institute (RTRI).²¹

A. Calculation of flow

First, the turbulent flow past a circular cylinder with a diameter (D) of 0.05 m perpendicularly placed in flow with a uniform velocity (U_∞) of 41.67 m/s is computed by the method in Sec. III. The Reynolds number based on D and U_∞ is 1.1×10^5 . Figure 1 shows a schematic diagram of the computational domain Ω_{int} . The center of the cylinder is taken as the origin of the coordinates with the main flow direction as the y_1 axis, lateral direction as the y_2 axis, and spanwise direction as the y_3 axis. A uniform flow U_∞ without turbulence is assumed for the inlet boundary. At the upper and lower boundaries, the slip (symmetry) condition is applied, and the no-slip condition for velocity is applied at the cylinder surface. At the remainder of the boundaries, the traction-free condition

$$n_j \left[(\nu + \nu_{\text{SGS}}) \left(\frac{\partial \bar{u}_i}{\partial y_j} + \frac{\partial \bar{u}_j}{\partial y_i} \right) \right] = 0, \quad (39)$$

is applied. The grid size in the radial direction close to the model surface is $0.005D$, which should be small enough to

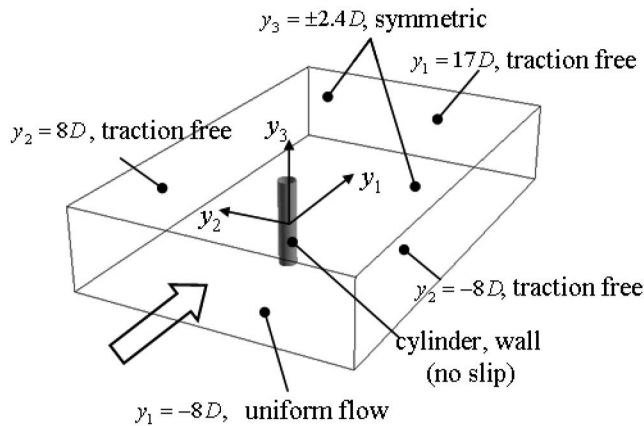


FIG. 1. Schematic diagram of the computational domain around a cylinder.

resolve the boundary layer and bound vortices around the body as pointed out at the end of Sec. II. 180 elements with nonregular intervals are used in the circumferential direction. Twenty-four elements are used in the spanwise direction with a constant spacing of $\Delta y_3 = 0.2D$, which results in a spanwise length of the computational domain of $4.8D$. The total number of computational elements is 138 000 for coarse mesh and 465 600 for fine mesh. The explicit time-accurate streamline upwind scheme was used with a nondimensional time increment of $\Delta t^* = \Delta t \times U_\infty / D = 5 \times 10^{-3}$.

Figure 2 shows the time histories of the force exerted on the cylinder. The force is nondimensionalized by the dynamic pressure of the uniform flow ($=\rho_0 U_\infty^2 / 2$) and the area of the cross section ($=D \times 4.8D$). The calculation is continued up to $t^* = 100$ when the flow becomes fully turbulent and periodical oscillations in y_2 direction appear. Then the flow properties are sampled at every 10 time steps (namely $10 \times \Delta t^* = 5 \times 10^{-2}$), and a total of 4 096 steps from $t^* = 100$ to $t^* = 304.8$ are sampled for the subsequent acoustical computations.

Figure 3 shows the profile of the time-averaged pressure at the cylinder surface where the vertical axis denotes pressure coefficient C_p normalized by the dynamic pressure of

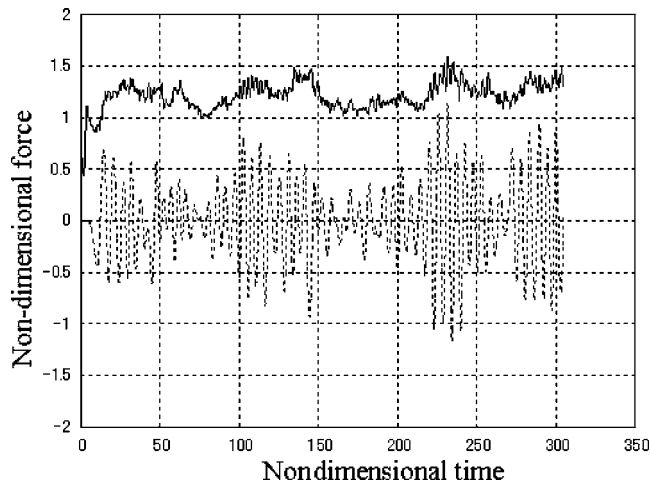


FIG. 2. Time histories of the nondimensional force exerted on a cylinder. Solid line: drag; dotted line: lift.

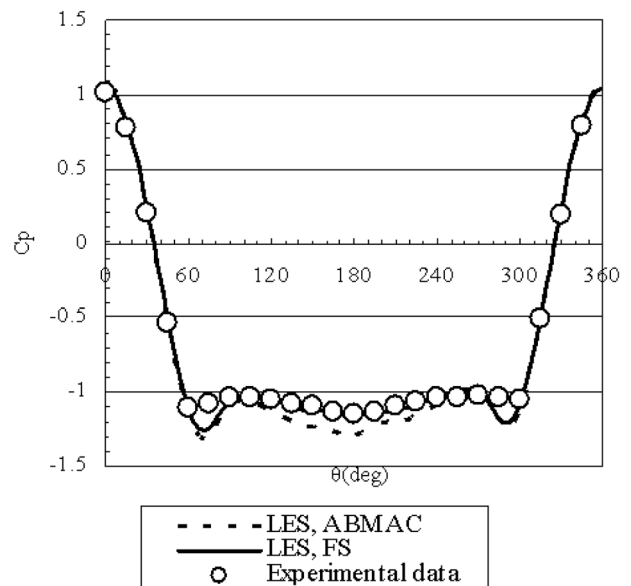


FIG. 3. Profile of time-averaged pressure at a cylinder surface. Point marks in the graph show the experimental data (Ref. 21).

the uniform flow, and the horizontal axis denotes the clockwise angle from the $-y_1$ axis (or the front stagnation point). The circles indicate the experimental data.²¹ Compared with the former results⁸ whose pressure Poisson solver was the ABMAC method, the present results by the FS method agrees well with the experimental results.

B. Calculation of Green's function

In this subsection, both the compact Green's function $G_c(\mathbf{x}, \mathbf{y}, \omega)$ and the noncompact Green's function $G_N(\mathbf{x}, \mathbf{y}, \omega)$ in the frequency domain are predicted numerically.

In the calculation of $G_c(\mathbf{x}, \mathbf{y}, \omega)$, the auxiliary velocity potential φ_j in Eq. (14) must be specified. Although there exists an analytical solution of φ_j for the cylinder, Laplace's equation (12) with Eqs. (11) and (13) is solved numerically by FEM for the sake of future extension to arbitrary body shapes (numerical errors are confirmed to be negligible in a previous study⁸).

Usually, the Helmholtz equation for the acoustic pressure p_a is applied to a direct BEM with p_a or $\partial p_a / \partial n$ given at the body surface S .²² In our case, however, we need $G_N(\mathbf{x}, \mathbf{y}, \omega)$ at each point \mathbf{y} in Ω_{int} , as discussed in Sec. II. Therefore Eq. (7) with Eq. (8) for $G_N(\mathbf{x}, \mathbf{y}, \omega)$ is determined directly by solving the integral equation (21) with a modified direct BEM and then substituting into Eq. (20). The merits of choosing BEM are as follows:

- (1) It is not necessary to solve the discretized wave equation for propagation to the far field. Therefore, no grid is needed in the domain, which reduces the CPU time and also numerical errors such as dispersion and dissipation.
- (2) BEM analytically satisfies the Sommerfeld radiation condition at the outer boundary Σ_{inf} at infinity. No artificial treatment is necessary to avoid the reflection.

Discretization elements at the body surface S need not be identical with those for the flow simulation. This is be-

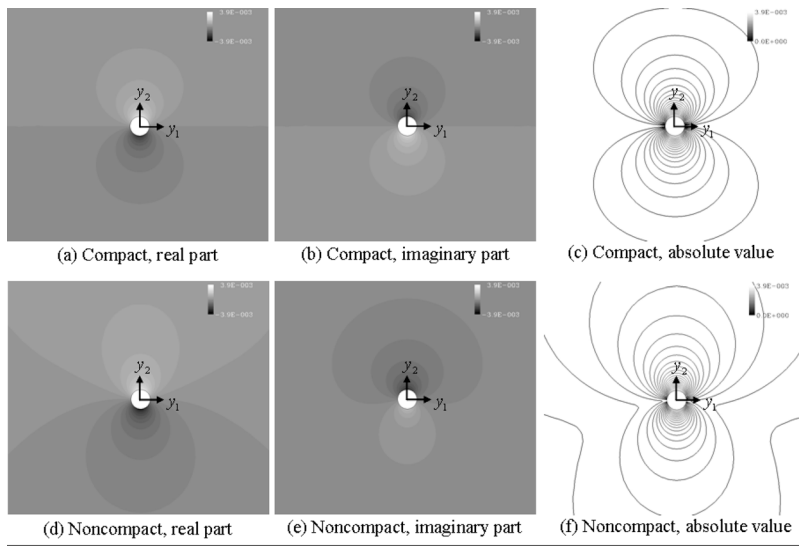


FIG. 4. Distribution of scattered Green's function around a cylinder for $\mathbf{x}=(0, 2 \text{ m}, 0)$, at $\omega/2\pi=195 \text{ Hz}$.

cause the element size for BEM needs only to be small enough to capture the acoustical wave at the frequency ω of interest. In most cases, the element size for the flow simulation is much smaller and it is not efficient to use the fluid mesh for the acoustic computations. Furthermore it is difficult to apply a symmetry condition at both spanwise boundaries for the acoustical computation. Therefore a cylinder with a finite spanwise length of 0.8 m ($16D$) is adopted for the boundary elements. The length should be large enough to avoid the effect of the cylinder's ends in Ω_{int} . The symmetry condition is applied at $y_3=0$ and only half of the cylinder (namely $8D$) is actually used as the boundary. The circumference of the cylinder is equally divided into 36 elements ($D/2 \times \Delta\theta=0.157D$), and 40 elements are used in the spanwise direction with a constant spacing of $\Delta y_3=0.2D$. These mesh sizes are small enough to capture acoustic waves up to 10 kHz with four points, although the number of elements in the circumferential direction is $1/5$ of the fluid simulation.

Although the fluid domain Ω_{all} should be large enough to contain all of the vorticity, the computational domain Ω_{int} is inevitably finite and the flow is not necessarily uniform at the outer boundary. The Lighthill stress terms are still sig-

nificant at the outer boundary, and their sudden termination is known to cause strong, spurious acoustic sources.¹³ The present authors avoided this problem by eliminating the direct sound from the turbulence flow and extracting the net contribution from the scattered sound.⁸ This idea matches with the fact that the acoustical efficiency of the direct sound (quadrupole) is less than that of the scattered sound (dipole) in the case of low Mach number flow. Hence, scattered parts are mainly discussed in the following sections.

Figure 4 shows a comparison of $G_c^s(\mathbf{x}, \mathbf{y}, \omega)$ and $G^s(\mathbf{x}, \mathbf{y}, \omega)$ for $\mathbf{x}=(0, 2 \text{ m}, 0)$ at $\omega/2\pi=195 \text{ Hz}$. The ratio of the wavelength λ to the diameter of the cylinder D is 31.1 . Figure 4 implies that the approximation of acoustically compactness is well satisfied at this frequency. Figure 5 shows a comparison of $G_c^s(\mathbf{x}, \mathbf{y}, \omega)$ and $G^s(\mathbf{x}, \mathbf{y}, \omega)$ for $\mathbf{x}=(0, 2 \text{ m}, 0)$ at $\omega/2\pi=8.3 \text{ kHz}$. The ratio λ/D is 0.823 . In this case the compact approximation is no longer applicable. From the reciprocal theorem, $G^s(\mathbf{x}, \mathbf{y}, \omega)=G^s(\mathbf{y}, \mathbf{x}, \omega)$, which implies that the pressure at \mathbf{x} produced by a point source \mathbf{y} is equal to the pressure at \mathbf{y} produced when the same source is placed at \mathbf{x} .⁷ Hence the distribution of the noncompact Green's func-

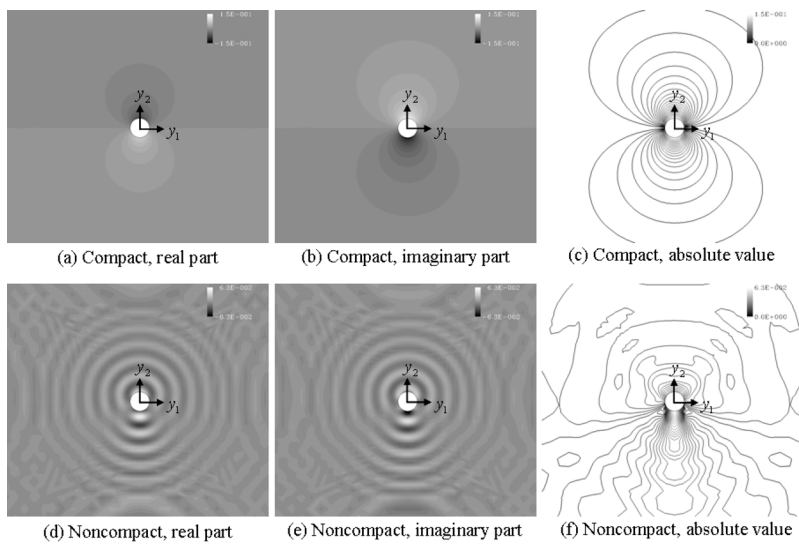


FIG. 5. Distribution of scattered Green's function around a cylinder for $\mathbf{x}=(0, 2 \text{ m}, 0)$, at $\omega/2\pi=8.3 \text{ kHz}$.

tion in Fig. 5 is similar to the forward scattering of electromagnetic wave (i.e., Mie scattering) whose source is located at $\mathbf{x}=(0,2\text{ m},0)$.

C. Estimation of sound pressure at the far field

In the previous sections, the instantaneous flow properties are solved by the LES technique, and Green's function is also obtained numerically. Hence, in this section, the sound pressure in the far field and the distribution of sound sources are predicted by using Eqs. (25)–(28). As mentioned in Sec. IV A, flow properties of 4 096 steps are stored. Then a Fourier transform is applied under the following condition: data length=256, data overlap=0%, frequency resolution (df)=65.1 Hz, window function=Hanning. A total of $4\,096/256=16$ sets of flow properties in the frequency domain is now obtained. Finally, the scattered component of the sound pressure $p_a^s(\mathbf{x},\omega)$ is calculated from Eq. (26). In order to compare with experiment,²¹ the power spectrum density (PSD) level of $p_a^s(\mathbf{x},\omega)$ is evaluated by means of

$$\begin{aligned} & \text{PSD level of } p_a^s(\mathbf{x},\omega) \text{ [dB/Hz]} \\ &= 10 \log_{10}\{|\overline{p_a^s(\mathbf{x},\omega)}| \times 2/(3/8)\} - 10 \log_{10}(df) \\ & \quad - 20 \log_{10}(2.0 \times 10^{-5}) + 10 \log_{10} \\ & \quad \times \left(\frac{\text{spanwise length of the cylinder at the experiment}}{\text{spanwise length of the cylinder at flow simulation}} \right). \end{aligned} \quad (40)$$

The sound pressure is also estimated by Curle's equation.²³ If the sound source is acoustically compact and the observer \mathbf{x} is far enough, the scattered component of sound pressure is

$$p_a^s(\mathbf{x},\omega) = -\frac{x_j}{4\pi c_0 |\mathbf{x}|^2} i\omega \int_S p(\mathbf{y},\omega) n_j dS \exp\left(i\omega \frac{|\mathbf{x}|}{c_0}\right). \quad (41)$$

Equation (41) is mathematically identical to Eq. (26) with $G^s(\mathbf{x},\mathbf{y},\omega) \cong G_c^s(\mathbf{x},\mathbf{y},\omega)$.^{7,8}

Figure 6 shows the sound pressure level at $\mathbf{x}=(0,2\text{ m},0)$. Not only for the discrete tone around 200 Hz dominated by the Kármán vortices, but also for the broadband sound up to 2 kHz, the predicted results agree well with experiment. The result with compact Green's function matches Curle's equation up to 3 kHz. This implies that the momentum equations (32) are strictly satisfied up to 3 kHz.

The results obtained by compact and noncompact methods agree well at a low frequency. Differences between them

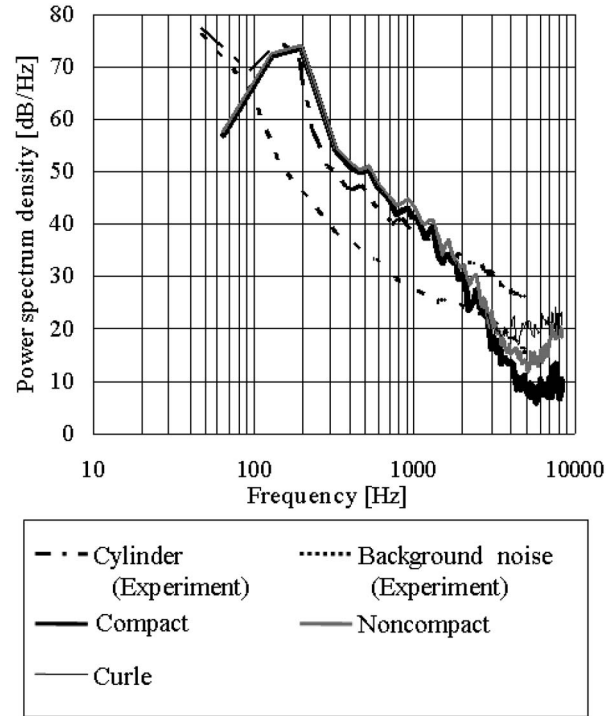


FIG. 6. Sound pressure level at $\mathbf{x}=(0,2\text{ m},0)$. Experimental data are from Ref. 21. “Compact” is calculated by Eqs. (17) and (26), “noncompact” by Eqs. (23) and (26), and “Curle” by Eq. (41).

appear progressively from 500 Hz and seem to increase with frequency, which is a result of the difference in the Green's functions as illustrated in Fig. 5.

Figure 7 shows the distribution of the scattered and direct sound sources for $\mathbf{x}=(0,2\text{ m},0)$ at $\omega/2\pi=195\text{ Hz}$. Figures 7(a) and 7(b) show the distribution of sound sources of the scattered sound expressed as Eq. (30), and Figs. 7(c) and 7(d) show the distribution of sound sources of the direct sound expressed as Eq. (30). Although the direct sound sources are widely distributed in the shear layers, the scattered sound sources are confined to the body surface. This seems to be reasonable because the influence of the body surface S is inversely proportional to the square of the distance between S and the source. From a fluid-dynamical point of view, this procedure is identical to estimating the contribution of bound vortices produced in response to the shedding of vorticity into the fluid from the surface. In Figs. 7(a) and 7(b), the phases of $(\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y},\omega) \cdot \nabla_y G^s(\mathbf{x},\mathbf{y},\omega)$ are almost the same in $y_2 > 0$ and $y_2 < 0$. This might seem strange. However, it can be easily explained by the motion of bound vortices as follows. When clockwise vorticity is shed

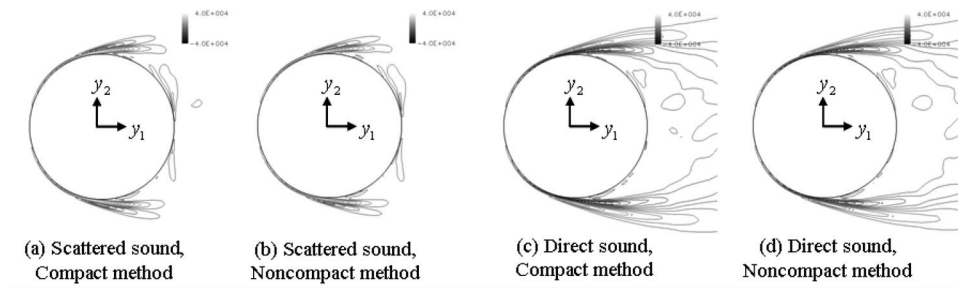


FIG. 7. Distribution of sound sources $(\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y},\omega) \cdot \nabla_y G^s(\mathbf{x},\mathbf{y},\omega)$ for $\mathbf{x}=(0,2\text{ m},0)$ at $\omega/2\pi=195\text{ Hz}$.

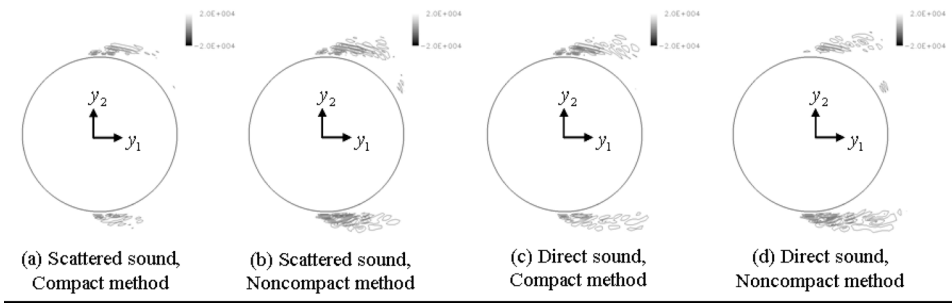


FIG. 8. Distribution of sound sources $(\boldsymbol{\omega} \times \mathbf{u})(\mathbf{y}, \omega) \cdot \nabla_y G(\mathbf{x}, \mathbf{y}, \omega)$ for $\mathbf{x} = (0, 2 \text{ m}, 0)$ at $\omega/2\pi = 8.3 \text{ kHz}$.

into the fluid from $y_2 > 0$ side of the cylinder, circulation around the cylinder increases. At the same time, the counter-clockwise vorticity that was shed from the $y_2 < 0$ side of the cylinder half a period earlier disappears. This also increases the clockwise circulation. Hence, the contribution from both sides of the cylinder are in phase.

Figure 8 shows the distribution of the scattered and direct sound sources for $\mathbf{x} = (0, 2 \text{ m}, 0)$ at $\omega/2\pi = 8.3 \text{ kHz}$. Although the sources are located in the shear layers, their distribution is more random compared with Fig. 7.

In the discussion so far, the observer \mathbf{x} is fixed at $\mathbf{x} = (0, 2 \text{ m}, 0)$, that is just beside the cylinder. Next the directivity of the scattered sound in the plane $x_3 = 0$ is calculated. The distance from the cylinder to the observer ($|\mathbf{x}|$) is fixed at 2 m. Figure 9(a) shows the directivity of the scattered sound $p_a^s(\mathbf{x}, \omega)$ for $\omega/2\pi = 195 \text{ Hz}$. At this low frequency, both the compact and the noncompact methods yield the same dipole-type directivity. Figure 9(b) shows the directivity at $\omega/2\pi = 8.3 \text{ kHz}$. At this high frequency, the directivities are quite different. The compact approximation implies that the scattered sound is emitted preferentially along the x_1 axis, whereas the noncompact method predicts a more complicated leaf-like pattern.

V. CONCLUSIONS

We have extended vortex sound theory to study noncompact problems and proposed a computational method for

evaluating the noncompact sound. The method has been applied to compute the far-field sound generated by flow past a circular cylinder. The following are the major conclusions:

- (1) The pressure in the far field is determined in the frequency domain by extending a theory of the vortex sound proposed by Howe. Scattered and direct components are decomposed, and the extraction of scattered parts avoids the sudden termination of Lighthill's stress tensor at the outer boundary of the finite computational domain.
- (2) The source fluctuations in the flow are first computed by a large-eddy simulation of incompressible flow using the dynamic Smagorinsky model. The fractional step method is used to solve the pressure Poisson equation. This improvement brings a better fit of time-averaged pressure at the cylinder surface with the experimental data.
- (3) Both compact and noncompact Green's functions are determined numerically. The boundary element method is used to compute the noncompact Green's function because of the reduced CPU time and small numerical errors caused by dispersion and dissipation during propagation, and also because of the easier treatment of the radiation condition at the outer boundary.
- (4) The noncompact Green's function and the compact Green's function agree well at low frequencies, but are quite different at high frequencies when the approximation of acoustical compactness is no longer applicable.
- (5) The sound pressure just beside the cylinder consists of a

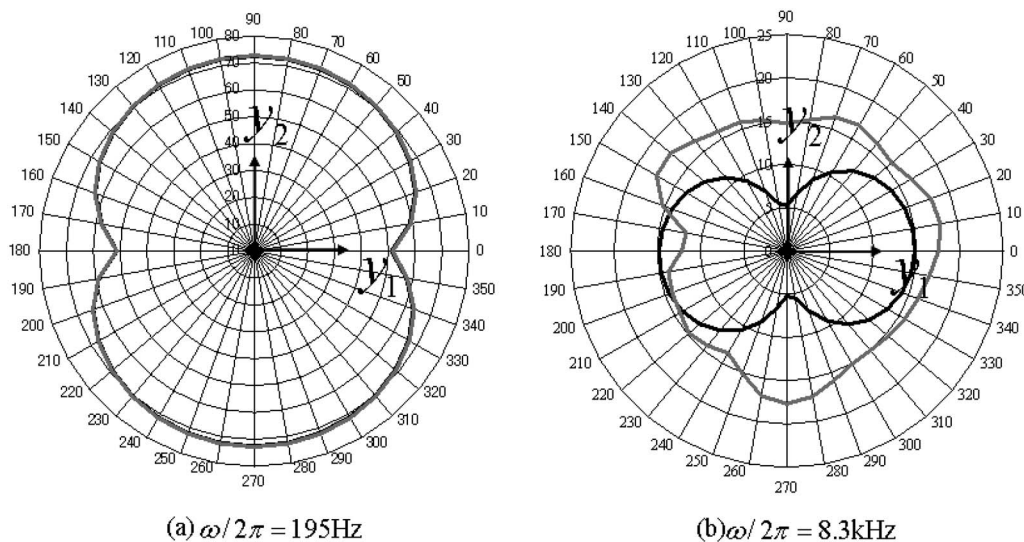


FIG. 9. Directivity of the scattered sound $p_a^s(\mathbf{x}, \omega)$ at $|\mathbf{x}| = 2 \text{ m}$ in $x_3 = 0$. Black line: compact method; gray line: noncompact method.

discrete tone dominated by the Kármán vortices together with broadband sound, and numerical predictions are consistent with experiments performed at frequencies up to 2 kHz. (6) The scattered sound pressure level predicted with compact Green's function matches Curle's equation up to 3 kHz. This implies that the momentum equations are strictly satisfied.

(7) Although the direct sound sources at low frequencies are widely distributed in the shear layers, the scattered sound sources are confined to the vicinity of the body surface. This is reasonable because the influence of the body surface is inversely proportional to the square of the distance between the surface and source.

(8) At low frequencies, both the compact and the noncompact methods yield the same dipole-type directivity. However, the directivity is quite different at high frequencies, and the noncompact method predicts a more complicated leaf-like pattern.

¹T. Takaishi, A. Sagawa, K. Nagakura, and T. Maeda, "Numerical analysis of dipole sound source around high speed trains," *J. Acoust. Soc. Am.* **111**, 2601–2608 (2002).

²C. K. W. Tam and K. A. Kurbatskii, "A wavenumber based extrapolation and interpolation method for use in conjunction with high-order finite difference schemes," *J. Comput. Phys.* **157-2**, 588–617 (2000).

³O. Inoue and N. Hatakeyama, "Sound generation by a two-dimensional circular cylinder in a uniform flow," *J. Fluid Mech.* **471**, 285–314 (2002).

⁴J. C. Hardin and D. S. Pope, "Sound generation by flow over a two-dimensional cavity," *AIAA J.* **33**, 407–412 (1995).

⁵C. Bailly and D. Juvé, "Numerical solution of acoustic propagation problems using linearized Euler equations," *AIAA J.* **38-1**, 22–29 (2000).

⁶J. C. Hardin and S. L. Lamkin, "Aeroacoustic computation of cylinder wake flow," *AIAA J.* **22-1**, 51–57 (1984).

⁷M. S. Howe, *Acoustics of Fluid-Structure Interactions* (Cambridge University Press, Cambridge, 1998).

⁸T. Takaishi, M. Ikeda, and C. Kato, "Method of evaluating dipole sound

source in a finite computational domain," *J. Acoust. Soc. Am.* **116-3**, 1427–1435 (2004).

⁹A. A. Oberai, F. Roknaldin, and T. J. R. Hughes, "Computation of trailing-edge noise due to turbulent flow over an airfoil," *AIAA J.* **40-1**, 2206–2216 (2002).

¹⁰M. S. Howe, "Vorticity and the theory of aerodynamic sound," *J. Eng. Math.* **41-4**, 367–400 (2001).

¹¹M. J. Lighthill, "On sound generated aerodynamically," *Proc. R. Soc. London, Ser. A* **211**, 564–587 (1952).

¹²M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, Cambridge, 2003).

¹³J. C. Hardin and M. Y. Hussaini, *Computational Aeroacoustics* (Springer-Verlag, New York, 1993).

¹⁴C. Kato, Y. Yamade, H. Wang, Y. Guo, M. Miyazawa, and T. Takaishi, "Numerical prediction of sound generated from flows with a low Mach number," *Comput. Fluids* **1-1**, Paper 2, URL: [http://www.sciencedirect.com/\[cited 29 November 2005\]](http://www.sciencedirect.com/[cited 29 November 2005]).

¹⁵M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, "A dynamic subgrid-scale eddy viscosity model," *Phys. Fluids A* **3**, 1760–1765 (1991).

¹⁶D. K. Lilly, "A proposed modification of the Germano subgrid-scale closure method," *Phys. Fluids A* **4**, 633–635 (1992).

¹⁷C. Kato and M. Ikegawa, "Large eddy simulation of unsteady turbulent wake of a circular cylinder using the finite element method," *Proceedings of ASME Fluids Engineering Conference*, **117**, 49–56 (1991).

¹⁸C. Kato, M. Kaiho, and A. Manabe, "An overset finite-element large-eddy simulation method with applications to turbomachinery and aeroacoustics," *J. Appl. Mech.* **70**, 32–43 (2003).

¹⁹A. van der Vorst, "Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems," *J. Sci. Comput.* **13-2**, 631–644 (1992).

²⁰A. Tamura, K. Kikuchi, and T. Takahashi, "Residual cutting method for elliptic boundary value problems," *J. Comput. Phys.* **137-2**, 247–264 (1997).

²¹T. Takaishi, M. Ikeda, and C. Kato, "Effects of periodic holes on the suppression of aeroacoustic noise from a pantograph horn," *Proceedings of 4th ASME/JSME Joint Fluids Engineering Conference (on CD-ROM)*, (2003).

²²*Boundary Element Acoustics*, edited by T. W. Wu (Cambridge WIT Press, Southampton, 2000).

²³N. Curle, "The influence of solid boundaries on aerodynamic sound," *Proc. R. Soc. London, Ser. A* **231**, 505–514 (1955).

Multipath pulse shapes in shallow water: Theory and simulation

Chris H. Harrison^{a)} and Peter L. Nielsen^{b)}

NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, Italy

(Received 27 September 2006; revised 19 December 2006; accepted 19 December 2006)

In shallow water propagation the steeper ray angles are weakened most by boundary losses. Regarding the sound intensity as a continuous function of angle it can be converted into a function of travel time to reveal the multipath pulse shape received from a remote source (one-way path) or a target (two-way path). The closed-form isovelocity pulse shape is extended here to the case of upward or downward refraction. The envelope of the earliest arrivals is roughly trapezoidal with a delayed peak corresponding to the slowest, near horizontal refracted paths. The tail of the pulse falls off exponentially (linearly in decibels) with a decay constant that depends only on the bottom reflection properties and water depth, irrespective of travel time, a useful property for geoacoustic inversion and for sonar design. The nontrivial analytical problem of inverting explicit functions of angle into explicit functions of time is solved by numerical interpolation. Thus exact solutions can be calculated numerically. Explicit closed-form approximations are given for one-way paths. Two-way paths are calculated by numerical convolution. Using the wave model C-SNAP in several broadband cases of interest it is demonstrated that these solutions correspond roughly to a depth average of multipath arrivals. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434691]

PACS number(s): 43.30.Gv, 43.30.Pc [RCG]

Pages: 1362–1373

I. INTRODUCTION

The multipaths of shallow water propagation spoil the resolution of active sonars by introducing a spread in travel times. The broadening of pulse transmission is therefore a nuisance to sonar detection and underwater communications (Urlick, 1967; Atkinson, 1974; Sachs *et al.*, 1968). On the other hand it has been shown (Smith, 1971; Harrison, 2003a) that the pulse shape contains easily extractable environmental information. So for both reasons the shape of the pulse and its dependence on environmental properties are of interest. One could investigate these effects with ray traces, but here the more general behavior is established by studying the pulse shape analytically.

In a multipath shallow water environment acoustic travel times and boundary losses vary according to the trajectories of the rays. In the absence of additional constraints, such as source or receiver beam patterns, or target vertical directionality this results in a calculable spreading of the transmitted pulse shape. If the sonar system has a broad frequency band then individual eigenray arrivals may be seen inside this spread. If it has a narrow band then interference effects make the rays group into modal arrivals. In both cases the arrivals tend to increase their separation as time advances. The mathematical approach here is insensitive to these detailed arrivals because it treats the ray angles as a continuum. Nevertheless it takes account of both their changing amplitudes and their changing separations in such a way that the cumulative time integral of the pulse shape matches the more “steppy” cumulative integral of the true eigenray or mode arrival pulse shape. The pulse envelope calculated here corresponds physically to a depth average or locally range-averaged pulse

shape. The depth average is particularly close for the tail of the pulse where rays interact with both seabed and sea surface.

Smith (1971) and Harrison (2003a) investigated this behavior for isovelocity water and found that for a two-way path the received pulse decayed exponentially (i.e., the roll-off was a fixed number of decibels per unit time) with a time constant that was independent of travel time or range but fixed by the angle dependence of the reflection loss. Subsequently Prior and Harrison (2004) applied the findings to experimental data, and demonstrated that the derived reflection properties were consistent with the literature.

This paper extends the earlier analytical work on pulse shape first by including the critical angle’s truncation of the pulse (trivial for one-way path but not trivial for two-way path), and second by including a uniform vertical gradient refraction using the approach of Harrison (2003b). In both cases one-way and two-way paths are considered. The former would be appropriate for direct blast measurements while the latter would be appropriate for the multipath echo from a point target or an echo repeater. In this respect the approach contrasts with other active sonar inversion techniques based, for instance, on matched field processing of the one-way path (Siderius *et al.*, 2002), or inversion of reverberation (Preston *et al.*, 2005). This time spreading is still expected even if the target has a large horizontal extent which may lead to a “glint” near the horizontal specular direction. However, no such effect will be found if the target has a large vertical extent like, for example, the leg of an oil rig.

Finally some comparisons in three environments with the normal mode model C-SNAP demonstrate the relationship between the closed-form pulse shapes and the sequence of resolvable eigenray arrivals.

^{a)}Electronic mail: harrison@nurc.nato.int

^{b)}Electronic mail: nielsen@nurc.nato.int

II. EIGENRAYS

The approach in Harrison (2003a) was to calculate propagation intensity analytically as an eigenray sum in which there are so many eigenrays that their index can be treated as a continuum in angle. The intensity can be treated equivalently as a continuum of incoherent modes or as energy flux. The solution behaves like a local range average, retaining depth variation but eliminating rapid interference fluctuations and convergence effects. The technique has been applied to bistatic geometry (Harrison, 2005a), a refracting environment with flat seabed (Harrison, 2003b), and a refracting environment with sloping seabed (Harrison, 2005b). The power arriving at a remote receiver can be written as an integral over eigenray index, which can be converted, by change of variable, to an integral over angle (at the source or receiver). In the context of time smearing, for each eigenray the travel time and the angle are known, so the power can also be written as an integral in travel time. Since travel time can be resolved, the multipath pulse shape is the integrand in this time integral. So to find these functions in a refracting environment one can use the formulas already derived for propagation in terms of ray elevation angle θ or inverse cycle distance u in Harrison (2003b). A minor difficulty is in writing the result in terms of time t so that the pulse shape is an explicit function of time.

Generally, knowing a propagation formula as the integral of a quantity $Q(\nu)$ over a parameter ν , and knowing the conversion from ν to t one obtains a pulse shape $I_{1\text{way}}$,

$$I_{1\text{way}}dt = \left\{ Q(\nu(t)) \left| \frac{d\nu(t)}{dt} \right| \right\} dt \quad (1)$$

The modulus sign is required for the case where t is not a single-valued function of ν , for instance ν could be angle. Integrating $I_{1\text{way}}$ in t would, of course, give the same result as integrating $Q(\nu)$ in ν . Both can be interpreted as either the linear form of transmission loss ($10^{-TL/10}$) or the energy per unit area resulting from a unit energy source. Thus $I_{1\text{way}}$ would be an intensity. The isovelocity case will be recapitulated in Sec. III. The pulse shape for the two-way path is given by the convolution of the one-way pulse shape $I_{1\text{way}}$ with itself,

$$I_{2\text{way}}dT = \int_0^T I_{1\text{way}}(t)I_{1\text{way}}(T-t)dtdT. \quad (2)$$

The function $I_{1\text{way}}$ may be discontinuous or split into several regions in each of which there is one continuous function. For a one-way path this is straightforward, but for the two-way path the convolution will contain several cross-term contributions with various integration limits.

III. REVIEW OF PULSE SHAPE WITH ISOVELOCITY WATER

A. One-way path

Following Harrison (2003a) the total time for a ray, tilted at θ , to travel a horizontal distance r at speed c is given by

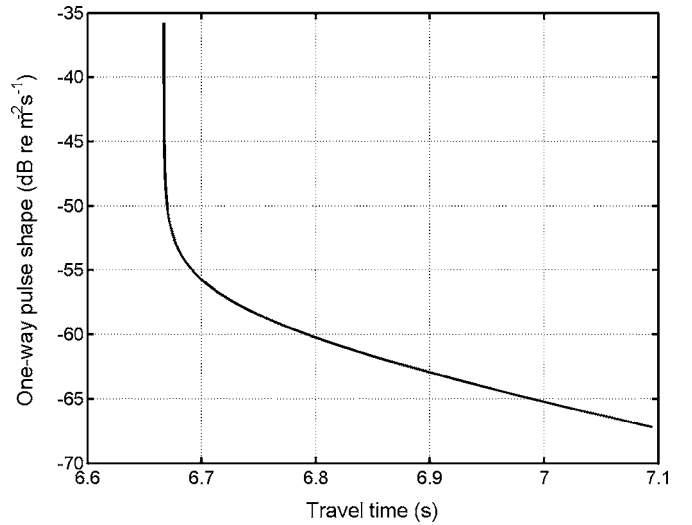


FIG. 1. One-way pulse shape in isovelocity water.

$$t = (r/c)\sec \theta \cong (r/c) + (r/2c)\theta^2. \quad (3)$$

In terms of the delay after first arrival τ this is

$$\tau = t - r/c = r\theta^2/2c, \quad (4)$$

$$d\tau = (r/c)\theta d\theta. \quad (5)$$

The received energy for a unit energy source is

$$E = \frac{2}{rH} \int_0^{\theta_c} \exp\left\{-\frac{\alpha\theta^2 r}{2H}\right\} d\theta, \quad (6)$$

where H , θ , θ_c are water depth, horizontal angle, and critical angle, and $\alpha\theta$ is a linear approximation to $\log(R)$, i.e., bottom reflection loss in dB is $RL = 20 \log_{10}(R) = \alpha_{\text{dB}}\theta$, where $\alpha = \alpha_{\text{dB}}/(10 \log_{10} e)$. This converts to a one-way pulse shape

$$I_{1\text{way}}d\tau = \frac{2}{rH} \frac{\exp\{-\alpha\tau/t_H\}}{\sqrt{2t_0\tau}} d\tau, \quad (7)$$

where $t_0 \equiv r/c$ is the delay time to the first arrival and $t_H = H/c$.

To the earlier derivation is added the condition that this equation is valid for $0 < \tau < \tau_c$ where τ_c corresponds to the critical angle θ_c . Elsewhere $I_{1\text{way}} = 0$. The behavior of this function is shown in Fig. 1. In this example $r = 10$ km, $\alpha_{\text{dB}} = 1$ dB, $\theta_c = 20^\circ$, $H = 100$ m. Units will be discussed in detail in Sec. VI. For now note that the mathematical quantity $I_{1\text{way}}$ in Eq. (7) has explicit dimensions of per unit area per unit time. Thus the y axis is labeled dB re $\text{m}^{-2} \text{s}^{-1}$. Notice that in the angle or travel time continuum case the first arrival is a singularity. This is because in the continuum the interval between eigenray arrivals goes to zero at this point. In reality, of course, eigenrays are discrete so the interval between first and second arrival is finite. Nevertheless the total energy in the pulse (integral $d\tau$) is finite since it has just been transformed from the, more obviously finite, integral in θ .

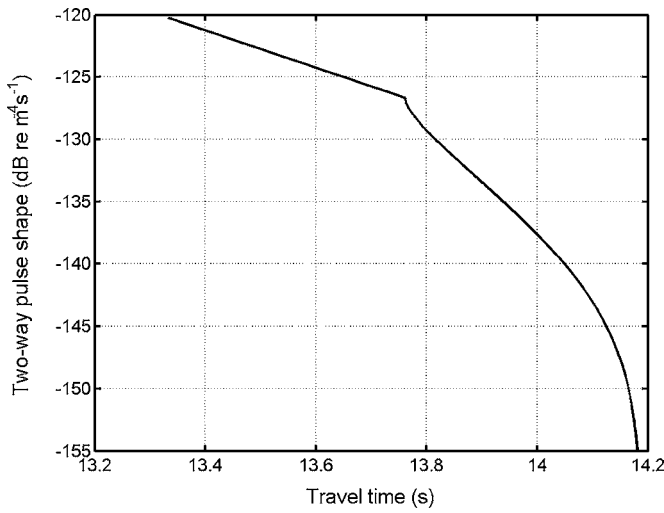


FIG. 2. Two-way pulse shape in isovelocity water. The early part is exponential (linear in dB).

B. Two-way path

The derivation in Harrison (2003a) for a two-way path T is correct in the absence of a critical angle, and is in any case correct for $T < \tau_c$. Otherwise the critical angle's truncation of the one-way pulse introduces a second functional form between delays τ_c and $2\tau_c$, beyond which $I_{2\text{way}}$ is zero. The integral has the same form as Eq. (53) in Harrison (2003a) but with different limits. Thus for $0 < T \leq \tau_c$ it is

$$\begin{aligned}
 I_{2\text{way}} dT &= \frac{2}{r^2 H^2 t_0} \int_0^T \frac{1}{\sqrt{\tau_1(T - \tau_1)}} d\tau_1 \exp\{-\alpha T/t_H\} dT \\
 &= \frac{2}{r^2 H^2 t_0} \int_{\tau'=-T/2}^{\tau'=T/2} \frac{1}{\sqrt{(T/2)^2 - \tau'^2}} d\tau' \\
 &\quad \times \exp\{-\alpha T/t_H\} dT \\
 &= \frac{2}{r^2 H^2 t_0} [\text{asin}(2\tau'/T)]_{-T/2}^{T/2} \exp\{-\alpha T/t_H\} dT \\
 &= \frac{2\pi}{r^2 H^2 t_0} \exp\{-\alpha T/t_H\} dT. \tag{8}
 \end{aligned}$$

For $\tau_c < T < 2\tau_c$ it is

$$\begin{aligned}
 I_{2\text{way}} dT &= \frac{2}{r^2 H^2 t_0} \int_{\tau'=(\tau_c - T/2)}^{\tau'=\tau_c - T/2} \frac{1}{\sqrt{(T/2)^2 - \tau'^2}} d\tau' \\
 &\quad \times \exp\{-\alpha T/t_H\} dT \\
 &= \frac{4}{r^2 H^2 t_0} \text{asin}((2\tau_c/T) - 1) \exp\{-\alpha T/t_H\} dT. \tag{9}
 \end{aligned}$$

The behavior of the complete function is shown in Fig. 2. Parameters are the same as in Fig. 1. There is no singularity as was seen with the one-way path, and the angle discontinuity separating the two functions is clearly visible in the middle. Note that this quantity is the time domain representation of two-way propagation loss with one grand smearing. Its units are therefore per unit area-squared per unit time.

IV. PULSE SHAPE WITH REFRACTION

To calculate dv/dt in Eq. (1) one needs formulas for travel time and cycle distance. In an environment with a uniform sound speed gradient c' between upper and lower sound speed values c_H, c_L (c_H may be physically at the top or the bottom of the water column) the cycle distance for a ray with a turning point velocity c_T is

$$d_0 = \frac{2}{|c'|} (|\sqrt{c_T^2 - c_L^2} - \sqrt{c_T^2 - c_H^2}|) \tag{10}$$

and the cycle travel time can be expressed in three ways:

$$\begin{aligned}
 \tau_0 &= \frac{1}{|c'|} \left(\left| \ln\left(\frac{1 + \sin \theta_L}{1 - \sin \theta_L}\right) - \ln\left(\frac{1 + \sin \theta_H}{1 - \sin \theta_H}\right) \right| \right) \\
 &= \frac{2}{|c'|} (|\text{atanh}(\sin \theta_L) - \text{atanh}(\sin \theta_H)|) \\
 &= \frac{2}{|c'|} (|\text{asinh}(\tan \theta_L) - \text{asinh}(\tan \theta_H)|). \tag{11}
 \end{aligned}$$

Throughout this paper subscripts L and H are attached to various properties to denote their values at the boundary corresponding to low sound speeds (L) or high sound speeds (H). The total one-way travel time can be written in terms of the horizontal cycle distance and the cycle travel time

$$t = \frac{r\tau_0}{d_0}. \tag{12}$$

This can easily be generalized to a piece-wise linear depth dependence (Harrison, 2006), and this method will be used in the last test case of Sec. VI.

Harrison (2003b) gives explicit formulas for propagation intensity with a flat seabed and uniform sound speed gradient. At very high frequencies one expects to find the effects of caustics, manifest in the depth dependence, but at lower frequencies they are expected to be less important. Propagation and reverberation intensities are calculated with, and without, caustic effects; in the main derivations here caustics are ignored although they are reconsidered in the context of the test cases in Sec. VI. Justification for this at frequencies of a few kilohertz or below is given in Appendix A of Harrison (2005b).

Note that the exact formulas from Harrison (2003b) for $\tan \theta_L$ and $\tan \theta_H$ are consistent with the exact formula for cycle distance [Eq. (10)] and cycle travel time [Eq. (11)]. Although, in the isovelocity case, a small angle approximation was introduced, all efforts are made, in the refraction case, to avoid approximations in calculating times (as opposed to intensities) until absolutely necessary because of the more complicated behavior.

There are two refraction regimes: In the first, rays have one refraction turning point and interact with only one boundary (the low speed side); in the second, rays interact with both boundaries.

A. One-way path: Single boundary interaction

Following Sec. 2.2.1 of Harrison (2003b) when rays interact with only one boundary the cycle distance d_0 and its reciprocal u are given exactly by

$$d_0 = u^{-1} = (2c_L/c') \tan \theta_L. \quad (13)$$

To avoid making *ad hoc* approximations at this stage, the mathematically convenient relationship $R = \alpha \tan \theta$ for each reflection is assumed at the outset. Therefore the total boundary loss at a fixed range r is independent of angle (and therefore time)

$$\alpha_L \tan \theta_L r u = \alpha_L \tan \theta_L r c' / (2c_L \tan \theta_L) = \alpha_L r c' / (2c_L). \quad (14)$$

The original equation [Eq. (2.6), Harrison, 2003b] was written as an integral in the parameter u . From Eq. (12) the total travel time t can also be written in terms of u ,

$$\begin{aligned} t &= \frac{r\tau_0}{d_0} = ur\tau_0 = \frac{2ru}{c'} \operatorname{asinh}(\tan \theta_L) = \frac{r \operatorname{asinh}(\tan \theta_L)}{c_L \tan \theta_L} \\ &= \frac{2ru}{c'} \operatorname{asinh}\left(\frac{c'}{2c_L u}\right). \end{aligned} \quad (15)$$

Since the exact maximum and minimum values of u are, respectively,

$$u_{\max} = \sqrt{c' / (4h_{\text{sr}}(c_{\text{sr}} + c_L))}, \quad (16)$$

$$u_{\min} = \sqrt{c' / (4H(c_H + c_L))}, \quad (17)$$

where H , c_L , c_H are water depth and sound speeds at the boundaries (high and low). The gradient is $c' = (c_H - c_L) / H$. The greater of the sound speeds at the source and receiver is c_{sr} , and the corresponding distance of this point from the low speed boundary is h_{sr} . Thus h_{sr} is a depth in an upward refracting environment but a height from the seabed in a downward refracting environment.

The corresponding exact minimum and maximum values of $\tan \theta_L$ are

$$\tan \theta_{L\min} = \sqrt{(c_{\text{sr}}/c_L)^2 - 1}, \quad (18)$$

$$\tan \theta_{L\max} = \sqrt{(c_H/c_L)^2 - 1}. \quad (19)$$

Substitution into Eq. (15) gives exact maximum and minimum travel times,

$$t_{\max} = r \operatorname{asinh}\left(\sqrt{(c_{\text{sr}}/c_L)^2 - 1}\right) / \sqrt{(c_{\text{sr}}^2 - c_L^2)}, \quad (20)$$

$$t_{\min} = r \operatorname{asinh}\left(\sqrt{(c_H/c_L)^2 - 1}\right) / \sqrt{(c_H^2 - c_L^2)}. \quad (21)$$

A straightforward numerical approach to evaluating Eq. (1) is first to differentiate Eq. (15) with respect to u to obtain du/dt as a function of u ,

$$\frac{du}{dt} = u(t - (r/c_L) \operatorname{sech}(tc'/(2ru))), \quad (22)$$

then the function $t(u)$ [in Eq. (15)] can be inverted to $u(t)$ by simply interpolating it on to a linear grid of t .

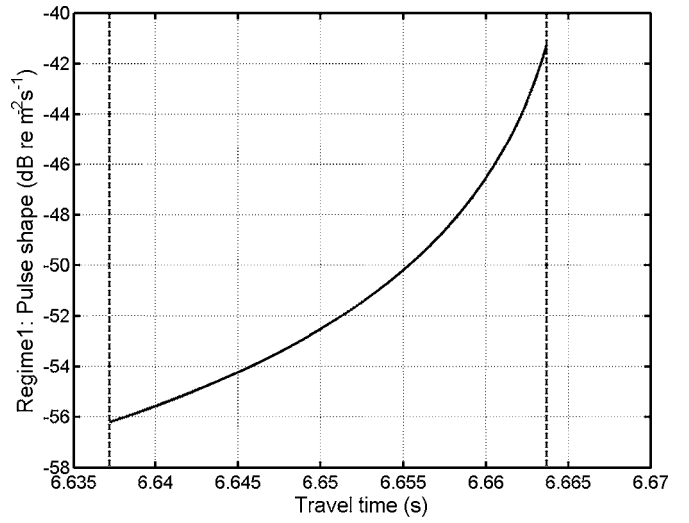


FIG. 3. Contribution to the pulse for refraction regime 1, bounded by the times t_{\min} and t_{\max} . Parameters are $c_L=1500$ m/s, $c_H=1520$ m/s, $r=10$ km, $\alpha_{L\text{dB}}=1$ dB, $\alpha_{H\text{dB}}=0$ dB, $H=100$ m, and $z_{\text{sr}}=90$ m.

Following Eq. (1) and adopting Eq. (22) leads to a formula for the single boundary interaction part (regime 1) of the pulse shape

$$\begin{aligned} I_1 dt &= \frac{4}{r} \exp(-R_L r u) \frac{du}{dt} \\ &= \frac{4}{r} u(t - (r/c_L) \operatorname{sech}(tc'/(2ru))) dt \exp\{-\alpha_L c' r / (2c_L)\} \end{aligned} \quad (23)$$

whose functional form in t can be seen by interpolation using Eq. (15).

In most of this paper it is assumed that bottom loss is linear with tangent of angle, however, the interpolation approach allows one to drop this assumption and take an arbitrary curve, as implied by the first line of Eq. (23) (R_L is the natural logarithm of the power reflection coefficient of the L boundary, i.e., on the low sound speed side). This will be useful in later comparisons with a wave model. An example with parameters $c_L=1500$ m/s, $c_H=1520$ m/s, $r=10$ km, $\alpha_{L\text{dB}}=1$ dB, $\alpha_{H\text{dB}}=0$ dB, $H=100$ m, $z_{\text{sr}}=90$ m is shown in Fig. 3 (z_{sr} is the complement of h_{sr} , i.e., $z_{\text{sr}}=H-h_{\text{sr}}$). The vertical dotted lines denote t_{\min} and t_{\max} .

To find an explicit functional form in t one needs to make an approximation. There are various approaches which are elaborated in Appendix A. The resulting approximate contribution to the pulse shape can be written as an explicit function of t (and $t_L=r/c_L$),

$$I_1 dt = \frac{dt}{(t_L - t)^{3/2}} \left\{ \frac{t_L^{3/2} c'}{\sqrt{6} r^2} \exp[-(\alpha_L c' / (2c_L)) r] \right\}. \quad (24)$$

The factor in curly braces does not affect the pulse shape of this contribution but it does control the amplitude relative to the two-boundary contribution. It is clear from the form of Eq. (24) that a singularity is possible when $t=t_L$. Usually this is precluded by the fact that $t_{\max} < t_L$, however, if both the source and receiver are on the “ L ” boundary then $t_{\max}=t_L$ [see Eq. (20)]. One could superimpose a plot of Eq. (24)

on the exact curve in Fig. 3 [i.e., Eq. (23)], however visual discrimination would be difficult since the mean difference is 0.1% in this case. This figure becomes 0.5% with $z_{sr}=10$ m, and reduces with sound speed contrast.

B. One-way path: Two-boundary interaction

Following Sec. 2.2.2 of Harrison (2003b) the total boundary loss is $(\alpha_L \tan \theta_L + \alpha_H \tan \theta_H)ru$, and the relationship between angles at the low and high speed boundaries θ_L , θ_H and cycle distance d_0 and its reciprocal u are given exactly by

$$\tan \theta_L = 2Hu a_L + c'/(4c_L u), \quad a_L = 1 + Hc'/2c_L, \quad (25)$$

$$\tan \theta_H = 2Hu a_H - c'/(4c_H u), \quad a_H = 1 - Hc'/2c_H, \quad (26)$$

$$t = \frac{2ru}{c'} (\operatorname{asinh}\{2Hu a_L + c'/(4c_L u)\} - \operatorname{asinh}\{2Hu a_H - c'/(4c_H u)\}). \quad (27)$$

The limits on u , θ , and t are as follows taking θ_c to be the smaller critical angle of the two boundaries. Strictly the upper limit on u is given by Eq. (25) with $\theta_L = \theta_c$, but it is approximately

$$u_{\max} \cong \tan \theta_c / 2H, \quad (28)$$

$$u_{\min} = \sqrt{c'/(4H(c_H + c_L))}. \quad (29)$$

Note that this second limit is the same as for regime 1 and that insertion into Eq. (27) leads to the same formula for t since the second term vanishes. The corresponding exact minimum and maximum values of $\tan \theta_L$ are

$$\tan \theta_{L\max} = \tan \theta_c, \quad (30)$$

$$\tan \theta_{L\min} = \sqrt{(c_H/c_L)^2 - 1}. \quad (31)$$

Substitution into Eq. (15) gives exact maximum and minimum travel times

$$t_{\max} = (r/c_L) \operatorname{asinh}(\tan \theta_c) / \tan \theta_c, \quad (32)$$

$$t_{\min} = r \operatorname{asinh}(\sqrt{(c_H/c_L)^2 - 1}) / \sqrt{c_H^2 - c_L^2}. \quad (33)$$

Again, a numerical solution is straightforward. Differentiating Eq. (27) with respect to u leads, after some manipulation, to

$$\frac{du}{dt} = u \left/ \left\{ t - \frac{2r}{\sqrt{\{2Hu(c_H + c_L)\}^2 + (c'/2u)^2 + 2(c_L^2 + c_H^2)}} \right\} \right. \quad (34)$$

From Eq. (2.17) Harrison (2003b) the pulse shape for the two-boundary-interacting component is

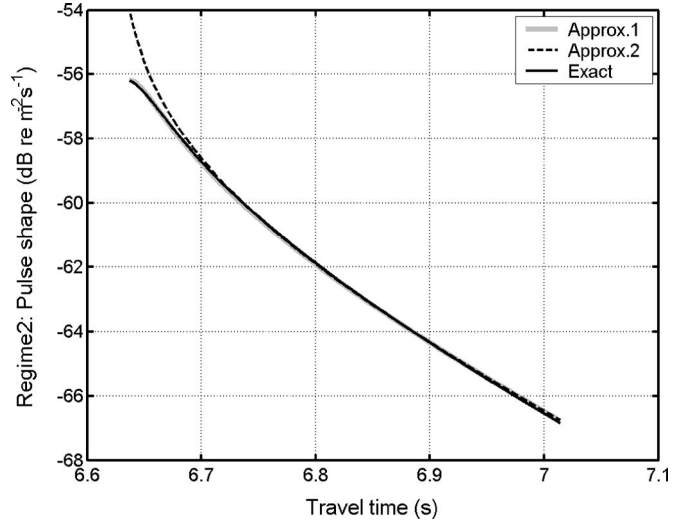


FIG. 4. Contribution to the pulse for refraction regime 2: “Exact” (solid black line); “Approx.1” [Eq. (36)] (thick grey line, partially obscured); “Approx.2” [Eq. (37)] (black dashed line). Parameters are $c_L=1500$ m/s, $c_H=1520$ m/s, $r=10$ km, $\alpha_{LdB}=1$ dB, $\alpha_{HdB}=0$ dB, $H=100$ m, and $z_{sr}=90$ m.

$$\begin{aligned} I_2 dt &= \frac{4}{r} \exp\{-(R_L + R_H)ru\} \frac{du}{dt} dt \\ &= \frac{4}{r} \exp\{-(\alpha_L \tan \theta_L + \alpha_H \tan \theta_H)ru\} \frac{du}{dt} dt \\ &= \exp\{-2rHu^2(\alpha_L a_L + \alpha_H a_H)\} \frac{du}{dt} \\ &\quad \times \left\{ \frac{4}{r} \exp\{-c'r(\alpha_L/c_L - \alpha_H/c_H)/4\} \right\} \quad (35) \end{aligned}$$

with a_L and a_H given by Eqs. (25) and (26). By using Eq. (27) to interpolate to t and Eq. (34) for du/dt the exact result is plotted in Fig. 4 with the same parameters as for regime 1 in Fig. 3. It is already clear that this solution must be close to that for the isovelocity case [Eq. (7)] for large τ . The first line of Eq. (35) is a reminder that the interpolation approach allows one to adopt arbitrary boundary loss R_L , R_H and to drop the assumption of linearity.

To find a pulse shape that is an explicit function of t one needs to make an approximation to Eqs. (27) and (28), but there are many choices, and there are two aims. First one seeks a compromise between accuracy and simplicity, and second, one seeks a solution that is simple enough in form to convolve with itself and with the regime 1 contribution [Eq. (24)] in order to derive a two-way pulse shape. The former succeeds but unfortunately the latter does not. Two contenders are derived in Appendix A using, respectively, Eq. (A16) and Eq. (A17) but both using Eq. (A14), and these are compared with the exact solution in Fig. 4. As functions of t these two approximations are, respectively,

$$\begin{aligned} I_2 dt &= \left(\frac{t}{\sqrt{b_1 t + b_0}} - \frac{1}{\sqrt{c_2 t^2 + c_1 t + c_0}} \right)^{-1} \\ &\quad \times \exp(-a_1 t - a_0) dt 2Hf_0, \quad (36) \end{aligned}$$

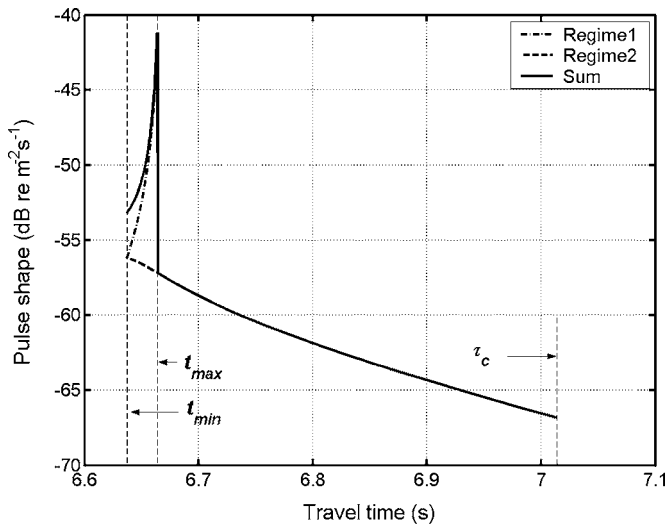


FIG. 5. Contributions to the pulse shape from refraction regime 1 (dash-dot line), regime 2 (dashed line), and their sum during the period in which they overlap (solid line). Parameters are as in Figs. 3 and 4.

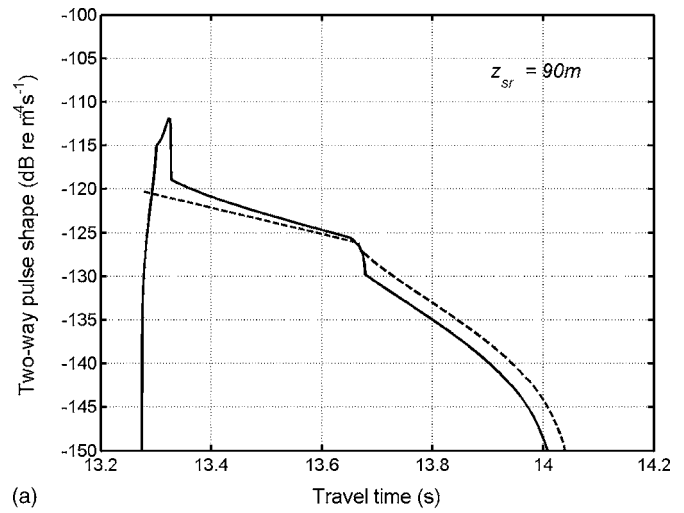
$$I_2 dt = \frac{t}{\sqrt{t^2 - t_0^2}} \exp(-a_1 t - a_0) \frac{1}{t_0} dt f_0, \quad (37)$$

where the constants a_0 , a_1 , b_0 , b_1 , c_0 , c_1 , c_2 , f_0 , t_0 are given in Appendix A.

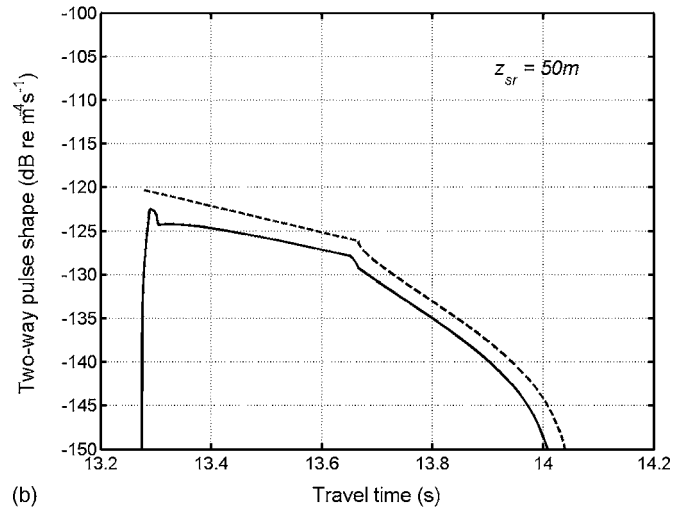
Note that the decay rate a_1 reduces to $(\alpha_L a_L + \alpha_H a_H)/t_H$ through Eq. (A13) if t_H is redefined as travel time across the water depth with an average sound speed. It is therefore still independent of travel time or range as in the isovelocity case. Comparing these formulas with the isovelocity case, Eq. (7), one finds that both equations converge on it as $c_L \rightarrow c_H$.

C. Two-way path: Numerical convolution

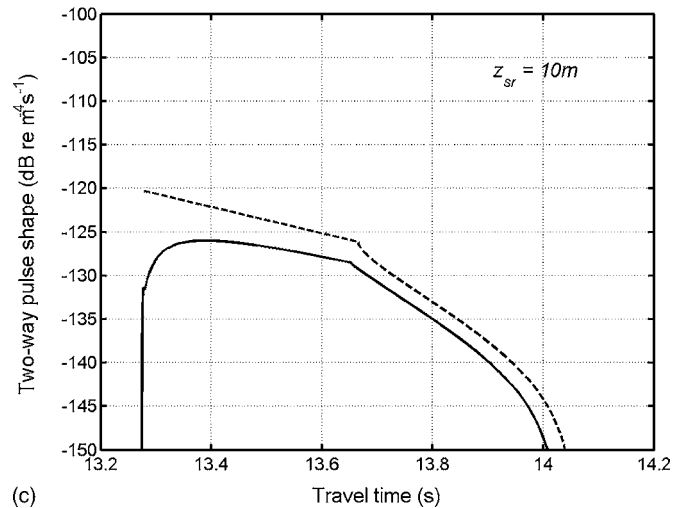
The pulse shapes for the two regimes overlap in time (see Figs. 3 and 4), but since the rays are considered to be incoherent their powers can be added to obtain an overall pulse shape as shown in Fig. 5. The characteristic trapezoidal or sail shape consists of a rise from first arrival at $t=t_{\min}$ with a later peak at $t=t_{\max}$ (rarely a singularity), followed by a sudden drop to an exponential decay which continues until $t=t_c$ at the critical angle. Thus between t_{\min} and t_{\max} the pulse shape is one explicit, continuous function, and between t_{\max} and t_c it is another. From the point of view of convolving the pulse by itself as in Eq. (2) there are three integrals to be formed, each with various limits. Here it has been tacitly assumed that the two-way paths go from source to receiver and back to source location. Clearly many permutations are possible if the source and receiver are separated. The calculation then consists of the convolution of the two different one-way pulse shapes. So far, the search for approximations to these functions that can be integrated analytically has been unsuccessful, so a simpler approach is numerical convolution. Three examples of pulse shapes for two-way paths are shown in Fig. 6 for the same parameters as in Fig. 5 but with $z_{sr}=90, 50$, and 10 m. The dashed lines show the isovelocity equivalent based on the average sound speed for comparison. More examples are given in Harrison (2006).



(a)



(b)



(c)

FIG. 6. Two-way pulse shape (solid) for parameters as in Fig. 5 except that z_{sr} (depth of the shallowest of source and receiver) takes values: (a) 90 m, (b) 50 m, and (c) 10 m. Isovelocity equivalent for average sound speed superimposed (dashed line).

V. RULES OF THUMB

It is possible to describe the one-way trapezoid shape roughly through two parameters. These are the time spread

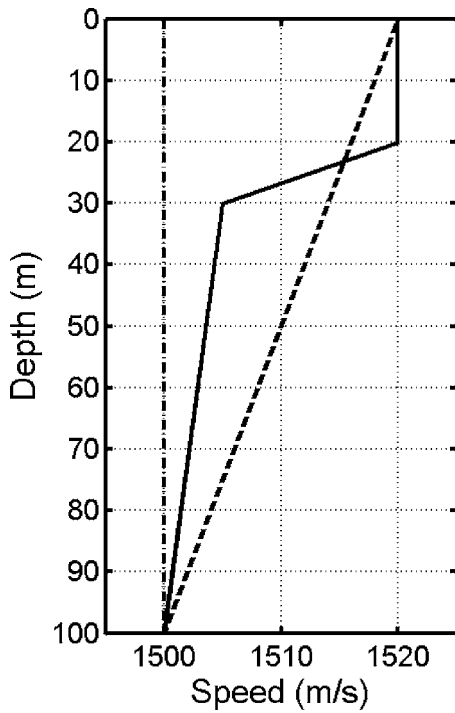


FIG. 7. Sound speed profiles for test cases: Case 1, isovelocity (dash-dot line), case 2, uniform gradient (dashed line), and case 3, “summer” (solid line).

of the regime 1 contribution, T_0 (i.e., the time difference between the first arrival and the peak $T_0 = t_{\max} - t_{\min}$), and the ratio of intensities at these two times F_0 . The right-hand sides of Eqs. (20) and (21) are both of the form $r/c_L \times \text{asinh}(X)/X$, so their difference can be written in terms of the series expansion as

$$\frac{\text{asinh}(X)}{X} - \frac{\text{asinh}(Y)}{Y} = -\frac{X^2 - Y^2}{3!} + \dots \quad (38)$$

So the time difference can be written in terms of the range, sound speed gradient, and $z_{\text{sr}}(z_{\text{sr}} = H - h_{\text{sr}})$, or in terms of sound speed contrast $\delta c_{\text{sr}} = c_H - c_{\text{sr}}$ and travel time $t_L = r/c_L$ as

$$T_0 = t_{\max} - t_{\min} = \frac{r}{6c_L^3}(c_H^2 - c_{\text{sr}}^2) \approx \frac{rz_{\text{sr}}c'}{3c_L^2} \approx \frac{1}{3}t_L \frac{\delta c_{\text{sr}}}{c_L}. \quad (39)$$

As one might expect, putting either source or receiver near the high speed boundary (small z_{sr}) or weakening the sound speed gradient results in a smaller angle range for regime 1 and therefore a shorter time spread.

According to Eq. (24) the ratio of intensities at these two times is

$$F_0 = \left(\frac{t_L - t_{\min}}{t_L - t_{\max}} \right)^{3/2}. \quad (40)$$

Again invoking Eqs. (20) and (21) and expanding their series to first order one finds

$$F_0 = \left(\frac{c_H^2 - c_L^2}{c_{\text{sr}}^2 - c_L^2} \right)^{3/2} = \left(\frac{H(c_H + c_L)}{(H - z_{\text{sr}})(c_{\text{sr}} + c_L)} \right)^{3/2} \approx \left(\frac{H}{(H - z_{\text{sr}})} \right)^{3/2} = \left(\frac{H}{h_{\text{sr}}} \right)^{3/2}. \quad (41)$$

To a good approximation the intensity ratio just depends on the proximity of the (closest of) source and receiver to the high speed boundary. In summary it is possible to control the pulse shape to a certain extent: The sail shape is most pronounced if both source and receiver are close to the low speed boundary; it is least pronounced and the pulse resembles the isovelocity case when one or both are near the high speed boundary. The tail of the pulse, with or without refraction, has an exponential decay that depends on reflection loss and water depth only. The shape of the tail is therefore independent of range.

As mentioned at the beginning of Sec. IV C it has been assumed that two-way paths are from source to receiver and back to the same source location. Of course, in reality there are multipaths between source and target and then between target and receiver, all of which may be at different depths. The above-mentioned rules of thumb are for one-way paths, and one can still combine their results assuming different end point depths on the outward and return paths.

VI. COMPARISON WITH A WAVE MODEL

Section I alluded to the fact that the approach used here is based on a continuum of rays or modes whereas in reality one may be able to discriminate actual rays with a broad band system or actual modal arrivals with a narrow band system. To make this comparison more concrete and to check the results the wave model C-SNAP (Ferla *et al.*, 1993) is run and averages devised to demonstrate the agreement explicitly. In the following, three test cases are considered: isovelocity, uniform downward refraction, and a three layer summer profile with a thermocline.

For the given range and source and receiver depths C-SNAP calculates a normal mode solution at 2000 frequencies between 500 and 1500 Hz. This is then shaded and Fourier transformed to form a received time series. To ensure correspondence with the analytical solutions the time series is normalized by also forming a source time series and calculating its integral of pressure squared over time. Thus there is effectively a unit energy source resulting in a received intensity, and the units of the response are intensity per unit source energy or $\text{m}^{-2} \text{s}^{-1}$, as already noted for the closed-form solutions (see also Hall, 1995; Ainslie and Beerens, 2005).

A further step needs to be taken because the time series resolves eigenrays, given a wide frequency band (or resolves modal arrivals for a narrow band). In principle these can be smoothed out with an average in travel time or range or depth. It is inevitable that in the tail of the smoothed pulse, where rays interact with both boundaries, the eigenray delays vary uniformly and continuously with depth so a depth average is appropriate. In the early part of the pulse for the refraction cases the analytical solution is also a function of depth so one can also compare the depth average of the ana-

TABLE I. Seabed and water half-space parameters.

	Sound speed c (m/s)	Density d (kg/m ³)/1000	Vol. absorption a (dB/wavelength)
Water	1500	1.0	0.0
Sediment	1580	1.5	0.1

lytical solution with the depth average of the normal mode solution. The formulas in the main text ignore focusing and caustic effects, however, following Weston (1980) it is straightforward to include them, as well as depth averaging, as explained in Appendix B.

All cases assume a water depth of 100 m, a source depth of 50 m, and a range of 50 km. Sound speed profiles are shown in Fig. 7, and the seabed is assumed to be a half-space, although this is not a restriction, with parameters given in Table I. The relevant part of the Rayleigh reflection coefficient is shown in Fig. 8. Notice that the linear approximation to reflection loss is very good up to about 16° (the critical angle is 18.3°, α is 0.32 rad⁻¹, and α_{dB} = 1.39 dB/rad). At worst it overestimates the loss by about 0.04 dB near 12°.

A. Isovelocity case

Figure 9 shows the closed-form solutions superimposed on the depth average of the C-SNAP isovelocity solution. The dashed line is the linear reflection loss approximation using Eq. (7) or (35), and the solid line is for Rayleigh reflection loss using the top line of Eq. (35). Considering the large number of eigenrays, and therefore bottom reflections (~80 near the critical angle), the Rayleigh curve is an extremely good fit. The slight misfit of the linear curve at around 34 s is entirely due to the small discrepancy seen in Fig. 8 of about +0.04 dB in 0.25 dB. The value of α quoted for the curve in Fig. 8 is based on the Rayleigh gradient at

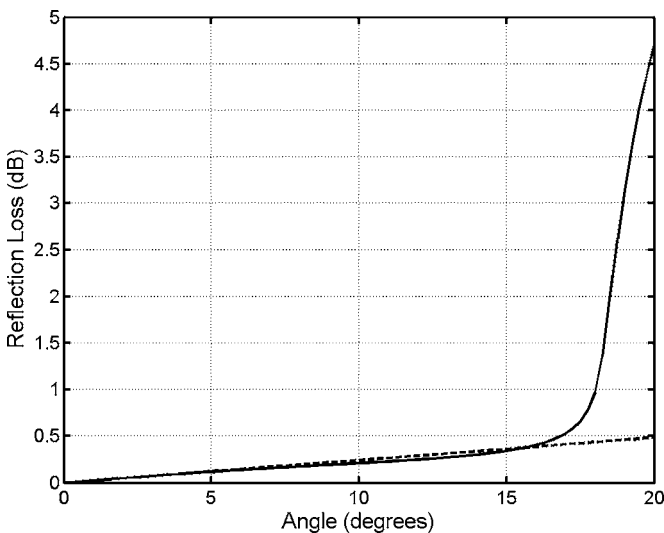


FIG. 8. Relevant part of the power Rayleigh reflection coefficient for half-space parameters as in Table I (solid line), and the linear approximation to it (dashed line). The difference reaches 0.04 dB at about 12°. The value of α is 0.32 (α_{dB} = 1.39).

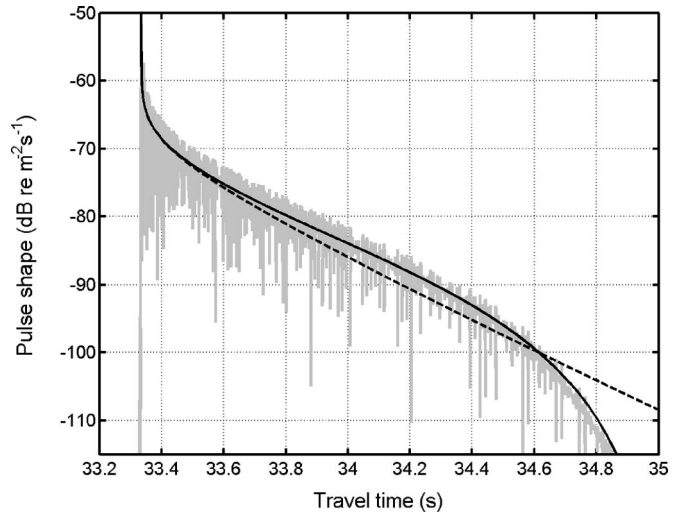


FIG. 9. Depth averaged pulse shape computed by C-SNAP for isovelocity water overlaying a half-space seabed defined by parameters in Table I (grey). Superimposed are the analytical solutions with Rayleigh reflection coefficient (black solid line), and linear approximation (dashed line).

the origin. In spite of this, an experimental measure of the fall-off rate of the pulse would lead to a reasonably accurate estimate of this slope.

B. Uniform gradient

The depth-averaged uniform gradient C-SNAP curve is shown as grey in Fig. 10, and one can see a pronounced refraction peak (regime 1). Superimposed are two curves, both using the Rayleigh reflection loss as in the first lines of Eqs. (23) and (35). To compare like with like the closed-form formulas are depth averaged as described in Appendix B. The solid line excludes caustic effects (a low frequency approximation), and the dashed line includes them (a high fre-

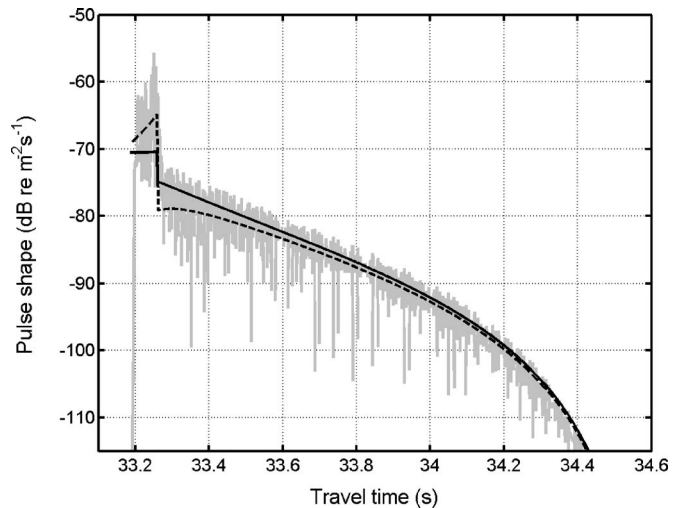


FIG. 10. Depth averaged pulse shape computed by C-SNAP for uniform sound speed gradient overlaying a half-space seabed defined by parameters in Table I (grey). Superimposed are the analytical solutions (also depth averaged) with Rayleigh reflection coefficient, excluding (black solid line), and including focusing effects (dashed line).

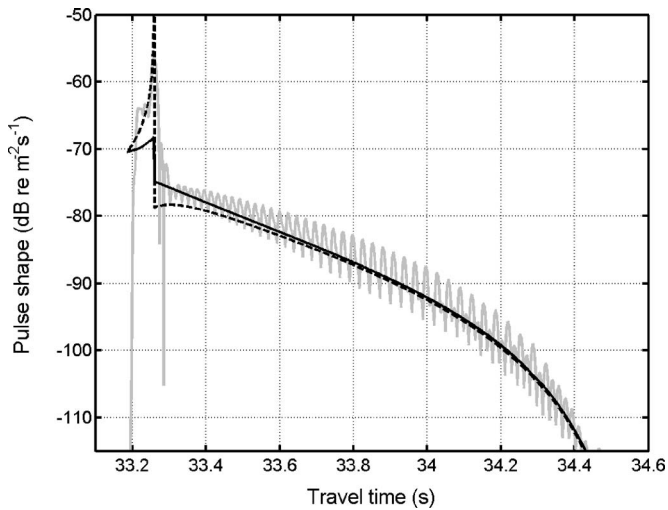


FIG. 11. Temporally smoothed pulse shape computed by C-SNAP for uniform sound speed gradient overlying a half-space seabed defined by parameters in Table I (grey line). Source and receiver depths are both 50 m. Superimposed are the analytical solutions with Rayleigh reflection coefficient, excluding (black solid line), and including focusing effects (dashed line).

quency approximation), making the modifications described in Appendix B. Both curves are very close to C-SNAP and more or less span its spread.

A more ambitious comparison is to exclude the depth averaging in all three cases and fix the source and receiver both at 50 m depth where there is a caustic (Brekhovskikh and Lysanov, 1982). The caustic occurs when the ray angle at the source is horizontal which corresponds to the latest refracting (regime 1) arrival, or the peak in Fig. 11. The C-SNAP result is smoothed in time in order not to interfere with its depth variations. Again closed-form solutions are superimposed including and excluding caustics. Agreement is particularly good when caustics are included, bearing in mind that the mean of the C-SNAP curve is 2 or 3 dB below its peak values.

C. Arbitrary summer profile

Using Eqs. (10) and (11) piece-wise one can construct the components of Eq. (1), and therefore a pulse shape, numerically. This has been done with, and without, caustics, for the summer profile from Fig. 7, and superimposed on the C-SNAP solution in Fig. 12. Before considering the details of the early arrivals it is worth noting that in all cases so far there is a clear point in time beyond which the pulse shape is linear in decibels, or at least a direct mapping of the reflection coefficient. This is important from the point of view of inversion. This delay time is determined by the slowest refracting path (i.e., that which does not reflect from both boundaries). In the uniform gradient case this time is given by t_{\max} [Eq. (20)], which in that case is 33.26 s, as seen in the plots. In the summer profile case the slowest ray is no longer the horizontal ray at the source. Instead it is the one whose turning point is at the top of the thermocline (20 m), since the strong curvature steepens the rays in the relatively slow bottom layer. Rays that remain in the bottom layer must arrive, according to Eqs. (20) and (21), between 33.296 and

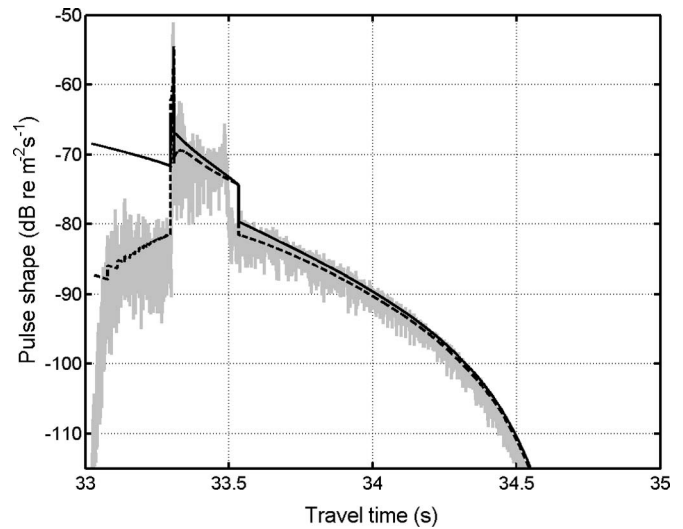


FIG. 12. Depth averaged pulse shape computed by C-SNAP for a summer sound speed profile overlying a half-space seabed defined by parameters in Table I (grey). Superimposed are the analytical solutions (also depth averaged) with Rayleigh reflection coefficient, excluding (black solid line), and including focusing effects (dashed line).

33.307 s, and are seen as the spike in the figure. Earlier delays than this correspond to rays that are just steep enough to enter the top fast layer. Because a very narrow range of angles at the source accounts for a large range of very long cycle distances the neglect of focusing effects becomes important. This explains the misfit of the solid line from 33.0 to 33.3 s but the relatively good fit of the dashed line. Otherwise the fit of both lines is extremely good for delays beyond about 33.3 s.

VII. CONCLUSIONS

This paper has derived formulas for the envelope of a multipath pulse in a shallow water environment with a uniform sound speed gradient. Variants are included for more general profiles. The refracted arrivals that interact only with the low sound speed boundary form a characteristic quadrilateral, or sail shape, near the leading edge of the pulse. In contrast, the steeper rays that interact with both boundaries form a long tail that decays more or less exponentially (linearly in decibels) with a rate dependent only on the reflection loss and the water depth. Furthermore this rate is independent of range or refraction. Some rules of thumb enable one to estimate the duration of the first refracted arrival [Eq. (39)] and the intensity ratio of its maximum and minimum [Eq. (41)].

Two approaches were taken to calculate the complete pulse shape. One was to write an exact expression [Eqs. (23) and (35)] in terms of mixed variables (time, cycle distance, and derivatives) which require interpolation to reveal pulse shape. The other was to make approximations to find explicit functions of time [Eqs. (24), (36), and (37)]. Two-way pulse shapes were obtained by numerical convolution rather than attempting analytical convolution, as was possible in the isovelocity case.

To establish correspondence between these pulse shapes and the expected eigenray or modal arrivals, three compari-

sons were made against the normal mode model C-SNAP, transforming its output into the time domain. The formulas for the three cases, isovelocity, uniform sound speed gradient, and summer profile, show very good agreement with a depth average or a time smoothing. Also the interpolation approach can easily handle the Rayleigh reflection loss, instead of its linear approximation.

The reasons for interest in the multipath pulse shape are threefold. First, the pulse shape is a useful predictor of geoaoustic properties because the tail of the pulse (i.e., the part later than the slowest refracting arrival) decays at a rate that depends only on water depth and angle derivative of reflection loss. In addition it is readily available for any existing active sonar. Equation (35) and the test cases show that there is also potential for extracting the shape of an arbitrary reflection loss curve from experimental data given adequate signal-to-noise-ratio. Second, it is vital to be able to predict multipath time spreading in order to adjust signal processing integration times. Since the shape of the tail is independent of range it can easily be calculated in advance and incorporated in typical signal processing feature extractors. Third, with perfect knowledge of the environment one might be able to extract a point target by deconvolution of the multipath arrivals. With imperfect knowledge one might have a better chance of deconvolution with the incoherent envelope predicted here since it is less sensitive to the detailed shifts and strengths of the eigenrays while retaining the sharp rise at the leading edge and the correct fall-off. In general this kind of filter (effectively an exponentially weighted smoothing) is invertible.

In addition, analytical and simple numerical approaches imply very short computation times so that these solutions can readily be incorporated in real-time systems, whether aimed at detection and classification or at geoaoustic inversion.

ACKNOWLEDGMENT

The authors would like to thank Alberto Baldacci of NURC for stimulating interest in the refraction version of these pulse shape formulas and some early comparisons with ray traces. Without his enthusiasm this work would not have existed.

APPENDIX A: PULSE SHAPE AS AN EXPLICIT FUNCTION OF TIME

There are two refraction regimes. In the first, rays interact only with the boundary on the c_L side; in the second, they interact with both boundaries. By making approximations the components of pulse shape Eqs. (23) and (35) are written as explicit functions of time.

1. Regime 1 contribution

The expression for t [Eq. (15), main text] is expanded as a series in u then differentiated with respect to t . This series is convergent since the argument of asinh (namely $\tan \theta_L$) is smaller than unity in the region of interest. Therefore one can invert the series (Morse and Feshbach, 1953). Equation (15) is expanded as

$$\begin{aligned} t &= \frac{2ru}{c'} \left(\left(\frac{c'}{2c_L u} \right) - \frac{1}{6} \left(\frac{c'}{2c_L u} \right)^3 + \frac{3}{40} \left(\frac{c'}{2c_L u} \right)^5 \right. \\ &\quad \left. - \frac{5}{112} \left(\frac{c'}{2c_L u} \right)^7 + \dots \right) \\ &= \frac{r}{c_L} \left(1 - \frac{1}{6} \left(\frac{c'}{2c_L u} \right)^2 + \frac{3}{40} \left(\frac{c'}{2c_L u} \right)^4 - \frac{5}{112} \left(\frac{c'}{2c_L u} \right)^6 + \dots \right). \end{aligned} \quad (\text{A1})$$

This can be rearranged as

$$\xi = 1 - t c_L / r = \frac{1}{6} X - \frac{3}{40} X^2 + \frac{5}{112} X^3 - \dots, \quad (\text{A2})$$

$$\text{where } X = \left(\frac{c'}{2c_L u} \right)^2. \quad (\text{A3})$$

The inverse series is

$$X = 6\xi + \frac{81}{5}\xi^2 + \frac{5184}{175}\xi^3 - \dots. \quad (\text{A4})$$

Numerically one can easily evaluate $t(u)$ from Eq. (A1) or (15) and $u(t)$ from Eq. (A4) and plot them on the same graph to demonstrate convergence and consistency. The intensity is given in terms of u by Eq. (2.6) from Harrison (2003b), so following Eq. (1), one needs an expression for du/dt . Differentiating Eq. (A3) with respect to u and Eq. (A2) with respect to t leads to

$$\frac{du}{dt} = \frac{du}{dX} \frac{dX}{d\xi} \frac{d\xi}{dt} = \frac{dX}{d\xi} \left(\frac{2c_L}{c'} \right)^2 \frac{u^3 c_L}{2r} = \frac{dX}{d\xi} \frac{c'}{4r X^{3/2}} \quad (\text{A5})$$

then differentiating Eq. (A4) with respect to ξ and substituting yields

$$\frac{du}{dt} = \frac{1}{4\sqrt{6} r \xi^{3/2}} f(\xi), \quad (\text{A6})$$

where

$$\begin{aligned} f(\xi) &= \frac{\left(1 + \frac{2 \times 81}{6 \times 5} \xi + \frac{3 \times 5184}{6 \times 175} \xi^2 + \dots \right)}{\left(1 + \frac{81}{6 \times 5} \xi + \frac{5184}{6 \times 175} \xi^2 + \dots \right)^{3/2}} \\ &= 1 + \frac{27}{20} \xi - \frac{891}{1120} \xi^2 + \dots \\ &= 1 + 1.35\xi - 0.7955\xi^2 + \dots. \end{aligned} \quad (\text{A7})$$

Because $r/c_L \geq t > r/c_H$, ξ is similarly bounded ($\xi_L \leq \xi < \xi_H$). The limiting values of ξ are $\xi_L = 0$, $\xi_H = H c' / c_H = (c_H - c_L) / c_H$. An extreme case might have $c_H - c_L = 30$ m/s so $\xi_H \sim 0.02$. Thus the second term in $f(\xi)$ is very small and the third is quite negligible. Notice that the first-order term in Eq. (A6) (i.e., setting $f=1$) could easily have been obtained directly from Eq. (A1) by taking just the first two terms and differentiating.

The resulting regime 1 contribution to the pulse shape can be written as an explicit function of t (and $t_L = r/c_L$),

$$I_1 dt = \frac{dt}{(t_L - t)^{3/2}} \left\{ \frac{t_L^{3/2} c'}{\sqrt{6} r^2} \exp[-(\alpha_L c' / 2c_L) r] \right\}. \quad (\text{A8})$$

2. Regime 2 contribution

If one expands the asinh terms of Eq. (27)—main text—in a third-order Taylor series about an angle θ_0 intermediate between θ_L and θ_H , where $\tan \theta_0 = 2Hu$, one arrives at

$$t = \frac{r}{2} \sqrt{1 + (2Hu)^2 (A + B/u^2)}, \quad (\text{A9})$$

where

$$A = \left(\frac{1}{c_L} + \frac{1}{c_H} \right) - \frac{(Hc')^2}{12} \left(3 \left(\frac{c_H + c_L}{c_H^2 c_L^2} \right) - \left(\frac{1}{c_L^3} + \frac{1}{c_H^3} \right) \right), \quad (\text{A10})$$

$$B = -\frac{(c')^2}{6 \times 16} \left(\frac{1}{c_L^3} + \frac{1}{c_H^3} \right). \quad (\text{A11})$$

$$\frac{du}{dt} = 1 \left/ \left\{ \frac{t}{u} - \frac{2r}{\sqrt{\{2H(c_H + c_L)\}^2 u^4 + 2(c_L^2 + c_H^2)u^2 + (c'/2)^2}} \right\} \right. \quad (\text{A16})$$

and substituting Eq. (A14) or (A15) in this makes a good approximation which can be used in Eq. (35)—main text—for the pulse shape, but the form is not useful for convolution. A poorer approximation is obtained by differentiating the approximate form, Eq. (A12), to obtain

$$\frac{du}{dt} = \frac{t}{(2Ht_0)^2 u} = \frac{t}{2Ht_0 \sqrt{t^2 - t_0^2}}. \quad (\text{A17})$$

As functions of t these two approximations are, respectively,

$$I_2 dt = \left(\frac{t}{\sqrt{b_1 t + b_0}} - \frac{1}{\sqrt{c_2 t^2 + c_1 t + c_0}} \right)^{-1} \times \exp(-a_1 t - a_0) dt \, 2Hf_0, \quad (\text{A18})$$

$$I_2 dt = \frac{t}{\sqrt{t^2 - t_0^2}} \exp(-a_1 t - a_0) \frac{1}{t_0} dt f_0, \quad (\text{A19})$$

where the constants $a_0, a_1, b_0, b_1, c_0, c_1, c_2, f_0$ are given by

$$a_0 = r(\alpha_L a_L + \alpha_H a_H)(u_{\min}^2 2t_0 H^2 - t_{\min})/(t_0 H),$$

$$a_1 = r(\alpha_L a_L + \alpha_H a_H)/(t_0 H),$$

$$b_0 = u_{\min}^2 - \frac{t_{\min}}{2t_0 H^2} \approx \frac{-1}{2H^2},$$

$$b_1 = \frac{1}{2t_0 H^2},$$

Even retaining only the first term of A this is an extremely good approximation. Neglect of B results in a slight discrepancy for early arrivals, but under these conditions the formula can be inverted to give

$$u^2 = ((t/t_0)^2 - 1)/(2H)^2 = \left(\frac{t - t_0}{t_0} \right) \left(\frac{t + t_0}{2t_0} \right) \frac{1}{2H^2}, \quad (\text{A12})$$

where

$$t_0 = \frac{r}{2} \left(\frac{1}{c_L} + \frac{1}{c_H} \right). \quad (\text{A13})$$

This suggests the alternative

$$u^2 = \left(\frac{t - t_{\min}}{t_0} \right) \frac{1}{2H^2} + u_{\min}^2 \quad (\text{A14})$$

or even

$$u^2 = \left(\frac{t - t_{\min}}{t_{\min}} \right) \frac{1}{2H^2} + u_{\min}^2 \quad (\text{A15})$$

Equation (34)—main text—can be written as

$$c_0 = \left(\frac{H}{r} \right)^2 (c_L + c_H)^2 \left(u_{\min}^2 - \frac{t_{\min}}{2t_0 H^2} \right)^2 + \frac{(c_L^2 + c_H^2)}{2r^2} \left(u_{\min}^2 - \frac{t_{\min}}{2t_0 H^2} \right) + \frac{c'^2}{16r^2},$$

$$c_1 = \frac{1}{r^2 t_0 H^2} \left\{ (c_L + c_H)^2 \left(H^2 u_{\min}^2 - \frac{t_{\min}}{2t_0} \right) + \frac{(c_L^2 + c_H^2)}{4} \right\},$$

$$c_2 = \frac{(c_L + c_H)^2}{(2rt_0 H)^2},$$

$$f_0 = \frac{2}{rH} \exp\{-c' r(\alpha_L/c_L - \alpha_H/c_H)/4\}$$

and t_0 is given by Eq. (A13).

APPENDIX B: CAUSTICS, FOCUSING, AND DEPTH AVERAGING

At high frequencies in shallow water it is well known (Brekhovskikh and Lysanov, 1982) that there is a pattern of caustics in depth-range space. The effects of taking a range average of intensity in such a field was investigated by Weston (1980). It is possible to extend his formulas to the two regimes considered here by inserting Wentzel–Kramers–Brillouin (WKB) modes into an incoherent mode sum and then writing modes and wave numbers in terms of a continuum of angles at source, receiver, and seabed (Harrison,

2003b). The relationships among mode normalization, cycle distance, and wave number K come from differentiating the WKB “phase integral” with respect to wave number. The result is

$$E = 2\pi \int \phi_s^2 \phi_r^2 Adn/(Kr) = \int \phi_s^2 \phi_r^2 AdK/(uKr)$$

$$= \frac{4}{r} \int \frac{uAdK}{K \tan \theta_s \tan \theta_r} = \frac{4}{r} \int \frac{u \tan \theta_L Ad \theta_L}{\tan \theta_s \tan \theta_r}, \quad (B1)$$

where $\phi_{s,r}$ are the modes at source and receiver, A is the Gaussian attenuation term already calculated in Secs. III and IV, and the integral limits are understood to span the part of the duct for which angles are real. This is converted to pulse shape by writing

$$Idt = \frac{4}{r} \int \frac{u \tan \theta_L A}{\tan \theta_s \tan \theta_r} \frac{d\theta_L}{dt} dt. \quad (B2)$$

Comparing this with the regime 1 [Eq. (23), main text] and regime 2 [Eq. (35)] formulas, having used the appropriate relationships between θ_L and u , [Eqs. (13) and (25)] it is found that they differ by “focusing factors” of, respectively,

$$f_1 = \frac{\sin^2 \theta_L}{\tan \theta_s \tan \theta_r}, \quad (B3)$$

$$f_2 = \frac{\sin \theta_L \cos \theta_L}{\tan \theta_s \tan \theta_r} \left(2Hu a_L - \frac{c'}{4c_L u} \right)$$

$$\approx \frac{\tan \theta_L \tan \theta_H \cos^2 \theta_L}{\tan \theta_s \tan \theta_r}. \quad (B4)$$

The first term has the effect of introducing a singularity in intensity whenever the angle at source, or receiver, or both is zero. Both terms tend to $\cos^2 \theta$ for fairly large angles, i.e., the tail of the pulse. It can be shown, by an eigenray derivation, that this is due to the approximation of slant range by horizontal range. Notice that in the application to pulse shape these terms are merely multipliers, so the problems of integration encountered by Weston (1980) are avoided. Inclusion or exclusion of these terms can be regarded as defining a high or low frequency limit.

By making use of the modes’ orthogonality, namely $\int \phi_r^2 dz = 1$, the depth average of Eq. (B1), is simply

$$E = 2\pi \int \phi_s^2 Adn/(KrH) = \int \phi_s^2 AdK/(uKrH)$$

$$= \frac{2}{rH} \int \frac{AdK}{K \tan \theta_s} = \frac{2}{rH} \int \frac{\tan \theta_L Ad \theta_L}{\tan \theta_s}. \quad (B5)$$

In the case of the simplified formulas of Sec. IV the turning point depth h for each ray determines the relevant weighting

h/H in the depth average. For a uniform gradient it can be shown that h is related exactly to u by

$$h = \frac{c_L}{c'} \left(\sqrt{1 + \frac{c'^2}{4u^2 c_L^2}} - 1 \right), \quad (B6)$$

which clearly has the property that it is exactly H when $u = u_{\min}$ [see Eq. (17), main text], it is h_s when $u = u_{\max}$, [Eq. (16)], and it is zero if $\theta_L = 0$ since u would be infinity.

- Ainslie, M. A., and Beerens, S. P. (2005). Sonar equation terms, in “A compendium of definitions for sonar performance measurements and modelling,” edited by M. A. Ainslie, Version 1, TNO Booklet No. AOG-v9, TNO Underwater Technology Dept., The Hague, The Netherlands, pp. 8–23.
- Atkinson, J. E. (1974). “Computer-implemented acoustic channel model,” *J. Acoust. Soc. Am.* **56**, S1–S18.
- Brekhovskikh, L., and Lysanov, Yu. (1982). *Fundamentals of Ocean Acoustics* (Springer, New York).
- Ferla, C. M., Porter, M. B., and Jensen, F. B. (1993). “C-SNAP: Coupled SACLANTCEN normal mode propagation loss model,” SACLANT Undersea Research Centre, La Spezia, Italy, Memorandum Rep. No. SM-274.
- Hall, M. V. (1995). “Dimensions and units of underwater acoustic parameters,” *J. Acoust. Soc. Am.* **97**, 3887–3889.
- Harrison, C. H. (2003a). “Closed-form solutions for reverberation and signal-excess with mode-stripping and Lambert’s law,” *J. Acoust. Soc. Am.* **114**, 2744–2756.
- Harrison, C. H. (2003b). “Signal and reverberation formulae including refraction,” SACLANT Undersea Research Centre, La Spezia, Italy, Rep. No. SR-370 (obtainable through national distribution centers, see <http://www.nurc.nato.int/publications/index.html>). Last viewed 2/7/07.
- Harrison, C. H. (2005a). “Fast bistatic signal-to-reverberation-ratio calculation,” *J. Comput. Acoust.* **13**, 317–340.
- Harrison, C. H. (2005b). “Closed form bistatic reverberation and target echoes with variable bathymetry and sound speed,” *IEEE J. Ocean. Eng.* **30**, 660–675.
- Harrison, C. H. (2006). “Multipath pulse shapes with refraction,” NATO Undersea Research Centre, La Spezia, Italy, Rep. No. NURC-FR-2006-005 (obtainable through national distribution centers, see <http://www.nurc.nato.int/pubs/index.html>). Last viewed 2/7/07.
- Morse, P. M., and Feshbach, H. (1953). *Methods of Theoretical Physics* (McGraw-Hill, New York), pp. 411–413.
- Preston, J. R., Ellis, D. D., and Gauss, R. C. (2005). “Geoacoustic parameter extraction using reverberation data from the 2000 Boundary Characterization Experiment on the Malta Plateau,” *IEEE J. Ocean. Eng.* **30**, 709–732.
- Prior, M., and Harrison, C. H. (2004). “Estimation of seabed reflection properties from direct blast pulse shape,” *J. Acoust. Soc. Am.* **116**, 1341–1344.
- Sachs, M. B., Sergeant, R. L., and Russotti, J. S. (1968). “Speech intelligibility in a stationary multipath channel,” *J. Acoust. Soc. Am.* **44**, 385.
- Siderius, M., Nielsen, P. L., and Gerstoft, P. (2002). “Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array,” *J. Acoust. Soc. Am.* **112**, 1523–1535.
- Smith, P. J. (1971). “The averaged impulse response of a shallow-water channel,” *J. Acoust. Soc. Am.* **50**, 332–336.
- Urick, R. J. (1967). *Principles of Underwater Sound* (McGraw-Hill, New York).
- Weston, D. E. (1980). “Wave-theory peaks in range-averaged channels of uniform sound velocity,” *J. Acoust. Soc. Am.* **68**, 282–286.

Propagation in an elastic wedge using the virtual source technique

Ahmad T. Abawi and Michael B. Porter

Heat, Light, and Sound Research, Inc., San Diego, California 92130

(Received 3 August 2006; revised 4 December 2006; accepted 8 December 2006)

The virtual source technique, which is based on the boundary integral method, provides the means to impose boundary conditions on arbitrarily shaped boundaries by replacing them by a collection of sources whose amplitudes are determined from the boundary conditions. In this paper the virtual source technique is used to model propagation of waves in a range-dependent ocean overlying an elastic bottom with arbitrarily shaped ocean-bottom interface. The method is applied to propagation in an elastic Pekeris waveguide, an acoustic wedge, and an elastic wedge. In the case of propagation in an elastic Pekeris waveguide, the results agree very well with those obtained from the wavenumber integral technique, as they do with the solution of the parabolic equation (PE) technique in the case of propagation in an acoustic wedge. The results for propagation in an elastic wedge qualitatively agree with those obtained from an elastic PE solution. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431336]

PACS number(s): 43.30.Ma [AIT]

Pages: 1374–1382

I. INTRODUCTION

An exact solution of the wave equation is possible only for a range-independent waveguide with plane parallel boundaries. In this case the wave equation is separable in a curvilinear coordinate system and the solution can be represented rigorously by a wavenumber integral. This solution can also be effectively expressed as a sum of propagating normal modes and a spectral integral representing the continuum. One of the distinct features of propagation in a range-independent waveguide is that, a mode, once excited, propagates without coupling with other modes. This forms the basis of the normal mode technique for the solution of the wave equation in a range-independent waveguide. The separability of the wave equation in a range-independent waveguide lends itself to solutions by a number of other techniques, most notably the wavenumber integration, the normal mode method, and the parabolic equation (PE) method.¹ The normal mode and the wavenumber integration techniques can also rigorously treat propagation of waves in an ocean overlying an elastic bottom.

In a range-dependent ocean environment, where the water depth and/or the ocean parameters vary with range, the wave equation is generally not separable. In this case the description of the solution in terms of previously uncoupled modes introduces mode coupling, as was shown by the coupled mode theory developed by Pierce² and Milder.³ If this coupling is ignored, an adiabatic mode solution is obtained, which is accurate for gently varying water depth. However, adiabatic mode theory is unable to describe the transition of modes from trapped, where they are conventionally defined, through cutoff to the leaky state. Since the original work of Pierce and Milder a number of authors have contributed to this technique.^{4–9} The technique is essentially based on expressing the field in a range-dependent waveguide in terms of local modes with range-dependent mode amplitudes. The application of the continuity of pressure and

the normal component of particle velocity allows a partial separation of the depth and range variables and yields a system of coupled differential equations for the mode amplitudes. However, recent advances in the parabolic equation technique^{10,11} have made it the method of choice for modeling propagation in a range-dependent ocean. The parabolic equation technique provides a one-way solution to the wave equation by factoring the range operator into outgoing and incoming operators and ignoring the factor, which contains the incoming operator. While this factorization is exact in a range-independent waveguide, it is an approximation in a range-dependent waveguide and only valid for small bottom slopes.

In most ocean acoustic applications the ocean bottom is modeled as a fluid. Modeling the ocean bottom as an elastic medium presents a number of problems in range-dependent modeling. In the case of the parabolic equation, there is not a clear choice of the dependent variable for which a parabolic equation can be derived. For some choices, the range operator cannot be factored and for some other choices the variables are not continuous across the interfaces, which causes numerical difficulties. Similarly, even though coupled mode solutions for a fluid overlying an elastic bottom have been formulated,^{12,13} computation of modes in the vicinity of the region where they transition from trapped to leaky presents numerical difficulties. These difficulties arise mainly because there is no obvious way to keep track of a mode's identity as it goes through cutoff, resulting in misinterpreting the exchange of energy between a set of modes and causing discontinuities to appear in the values of the computed field quantities.

In this paper we use the method of virtual sources to compute propagation in a range-dependent waveguide. This method has widely been used in target scattering computations, particularly when the target is located in a waveguide,^{14–17} but its use in modeling propagation in a waveguide is relatively recent.¹⁸ The method is based on

modeling a boundary or an interface by a collection of sources with unknown complex amplitudes. The functions representing the sources must satisfy the wave equation and the radiation condition in the far field. The amplitudes of the sources are determined in a manner by which the field produced by them and the incident field satisfy all the required boundary and interface conditions. This, in principle, provides an exact solution of the wave equation in a waveguide of arbitrary cross section. The accuracy of the solution is a function of the number of sources, and in most cases the use of five to ten sources per wavelength produces sufficiently accurate results. The sources are usually placed a fraction of a wavelength from a contour following the boundary. The formulation presented in this paper is specific to a waveguide composed of an isovelocity water layer over an acoustic or elastic half-space bottom. However, the water depth can be an arbitrary function of range. The case of a waveguide with a variable sound speed profile overlying a bottom composed of multiple layers will be addressed in a future paper. Even though the technique does not place any restrictions on the shape of the boundary, we apply this technique to a penetrable oceanic wedge. The reason we have chosen the wedge is that it has a simple geometry yet it possesses most of the interesting physical properties of a range-dependent waveguide. Furthermore, the wedge has been studied by numerous authors,¹⁹⁻²² whose works provide valuable benchmark solutions. The main purpose of this paper is to provide a benchmark solution for an oceanic wedge of arbitrary bottom slope overlying a fluid bottom or an elastic bottom (henceforth referred to as an acoustic or an elastic wedge, respectively). For simplicity of presentation, we model a wedge in a two-dimensional ocean, realizing that modification to a three-dimensional ocean is straightforward. We will, however, discuss the use of the virtual source technique in modeling propagation in an azimuthally symmetric ocean in Appendix A.

This paper is organized as follows: In Sec. II propagation in range-dependent waveguide is formulated for an isovelocity fluid layer overlying an isovelocity fluid or elastic half space. In Sec. III the method is applied to an acoustic and elastic Pekeris waveguide and to an acoustic and an elastic wedge and the results are compared to available published results. The paper concludes with a discussion in Sec. IV.

II. FORMULATION

A. The acoustic case

This formulation is carried out in a two-dimensional Cartesian coordinate system (x, z) . The z axis is pointing downward and the plane $z=0$ represents the ocean surface. The Helmholtz equation for a line source at (x', z') is given by

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2 \right] G_p(\mathbf{r}, \mathbf{r}') = \delta(x - x') \delta(z - z'), \quad (1)$$

where

$$\mathbf{r} = (x, z), \quad \mathbf{r}' = (x', z').$$

The Green's function, $G_p(\mathbf{r}, \mathbf{r}')$ is given by

$$G_p(\mathbf{r}, \mathbf{r}') = \frac{i}{4} H_0^{(1)}(k|\mathbf{r} - \mathbf{r}'|), \quad (2)$$

where the subscript p is used to indicate that G_p is a pressure Green's function. Later we will also introduce G_u , the displacement Green's function. The Green's function given by Eq. (2) is a fundamental solution of Eq. (1) and satisfies the Sommerfeld radiation condition in the far field. These two conditions make $G(\mathbf{r}, \mathbf{r}')$ suitable for use as basis functions in the virtual source technique.

Let us represent the bottom boundary by $h(x)$ and the unit normal to the boundary by

$$\hat{\mathbf{n}} = \frac{(-\nabla h, 1)}{\sqrt{1 + (\nabla h)^2}}.$$

We intend to solve the Helmholtz equation for the pressure $p(x, z)$,

$$[\nabla^2 + k^2]p(x, z) = 0, \quad (3)$$

for the following boundary conditions:

$$p(x, 0) = 0,$$

$$[p]_{\pm}^+ = 0, \quad [\hat{\mathbf{n}} \cdot \mathbf{u}]_{\pm}^+ = 0. \quad (4)$$

The above equations express the pressure-release boundary condition at the ocean surface and the continuity of pressure and normal displacement at the water-bottom interface. Since the basis functions given by Eq. (2) satisfy Eq. (3), they will be used by the virtual source technique to express the solution as a sum of basis functions. The boundary conditions given by Eq. (4) will subsequently be imposed. Since two sets of boundary conditions need to be satisfied at the interface, two sets of virtual sources are needed. To do this, a set of sources is placed just below the interface to produce the field in water and another set of sources is placed just above the interface to produce the field in the bottom. Let N denote the number of sources, and let the position of source i be denoted by the position vector $r_i = (x_i, z_i)$. Let \mathbf{r}_1 represent the location of the sources that produce the field in the water and \mathbf{r}_2 represent the location of the sources that produce the field in the bottom. Let us also represent by \mathbf{r}_a the location of points on the interface or nodes where the boundary conditions will be imposed. The pressure-release boundary condition at the flat ocean surface can be constructed by the method of images. Therefore, the Green's function for the field in the water produced by the sources located at \mathbf{r}_p is

$$\mathbf{G}_p^{(1)}(\mathbf{r}; \mathbf{r}_1) = \frac{i}{4} \mathbf{H}_0^{(1)}(k_1|\mathbf{r} - \mathbf{r}_1|) - \frac{i}{4} \mathbf{H}_0^{(1)}(k_1|\mathbf{r} - \mathbf{r}_1^i|),$$

where \mathbf{r}_1^i represents the location of the image sources. In the above equation all the quantities represented by capital bold letters are matrices. For example, the entries of $\mathbf{H}_0^{(1)}(k_1|\mathbf{r} - \mathbf{r}'|)$ are $H_0^{(1)}(k_1|r_i - r'_j|)$, where r_i is the location of receiver i and r'_j is the location of source j . In this paper we have adopted the convention of representing vectors by lower-case bold letters and matrices by upper-case bold letters. The Green's function for the field in the bottom produced by the sources located at \mathbf{r}_q is

$$\mathbf{G}_p^{(2)}(\mathbf{r}; \mathbf{r}_2) = \frac{i}{4} H_0^{(1)}(k_2 |\mathbf{r} - \mathbf{r}_2|).$$

In the above equations $k_1 = \omega/c_w$, $k_2 = \omega/c_b$, where ω is the angular frequency, c_w is the water sound speed and c_b is the bottom sound speed, the subscripts 1 and 2 specify the water and bottom layers, and the position vector \mathbf{r} represents the location of the field points and has the same dimension as \mathbf{r}_1 and \mathbf{r}_2 . Note that the Green's functions in the water automatically satisfies the pressure-release boundary condition at the water surface, $z=0$. The field due to the virtual sources of amplitude s_1, s_2, \dots, s_N located at r'_1, r'_2, \dots, r'_N at a field point r can be written

$$p(r) = \sum_{j=1}^N G_p(r; r'_j) s_j.$$

In vector notation, this can be expressed as

$$p(r) = \mathbf{G}_p(r; \mathbf{r}') \mathbf{s},$$

and for a vector of field points r_1, r_2, \dots, r_N , vector-matrix notation may be used to write

$$\mathbf{p}(\mathbf{r}) = \mathbf{G}_p(\mathbf{r}; \mathbf{r}') \mathbf{s},$$

where in the above \mathbf{p} and \mathbf{s} are column vectors of length N and \mathbf{G}_p is a $N \times N$ matrix. Using the same notation, the field in the water and in the bottom can be expressed as a sum of virtual sources of source amplitude \mathbf{s}_1 and \mathbf{s}_2 ,

$$\begin{aligned} \mathbf{p}^{(1,2)}(\mathbf{r}) &= \mathbf{G}_p^{(1,2)}(\mathbf{r}; \mathbf{r}_{1,2}) \mathbf{s}_{1,2}, \\ \mathbf{u}^{(1,2)}(\mathbf{r}) &= \mathbf{G}_u^{(1,2)}(\mathbf{r}; \mathbf{r}_{1,2}) \mathbf{s}_{1,2}. \end{aligned} \quad (5)$$

where the normal displacement, \mathbf{u} , is a column vector whose elements are $u_i = \hat{\mathbf{n}} \cdot \nabla p_i / (\rho \omega^2)$ and \mathbf{G}_u is a matrix whose elements are $G_{u(i,j)} = \hat{\mathbf{n}} \cdot \nabla G_{p(i,j)} / (\rho \omega^2)$. Applying the continuity of pressure and normal displacement at the nodes \mathbf{r}_a gives

$$\begin{aligned} \mathbf{p}_{inc}(\mathbf{r}_a) + \mathbf{G}_p^{(1)}(\mathbf{r}_a; \mathbf{r}_1) \mathbf{s}_1 &= \mathbf{G}_p^{(2)}(\mathbf{r}_a; \mathbf{r}_2) \mathbf{s}_2, \\ \mathbf{u}_{inc}(\mathbf{r}_a) + \mathbf{G}_u^{(1)}(\mathbf{r}_a; \mathbf{r}_1) \mathbf{s}_1 &= \mathbf{G}_u^{(2)}(\mathbf{r}_a; \mathbf{r}_2) \mathbf{s}_2. \end{aligned} \quad (6)$$

In the above equations \mathbf{p}_{inc} is the incident field due to a source in the water and \mathbf{u}_{inc} is a column vector whose elements are $u_{inc(i)} = \hat{\mathbf{n}} \cdot \nabla p_{inc(i)} / (\rho \omega^2)$. The source amplitudes can be obtained from the solution of the above equations,

$$\begin{aligned} \mathbf{s}_1 &= (\mathbf{G}_p^{(1)} - \mathbf{K} \mathbf{G}_u^{(1)})^{-1} (\mathbf{K} \mathbf{u}_{inc} - \mathbf{p}_{inc}), \\ \mathbf{s}_2 &= (\mathbf{L} \mathbf{G}_u^{(2)} - \mathbf{G}_p^{(2)})^{-1} (\mathbf{L} \mathbf{u}_{inc} - \mathbf{p}_{inc}), \end{aligned} \quad (7)$$

where

$$\mathbf{K} = \mathbf{G}_p^{(2)} (\mathbf{G}_u^{(2)})^{-1} \quad \text{and} \quad \mathbf{L} = \mathbf{G}_p^{(1)} (\mathbf{G}_u^{(1)})^{-1}.$$

The arguments of the functions in Eq. (7) are those in Eq. (6). The source amplitudes determined from Eq. (7) in conjunction with Eq. (5) give the field anywhere in the waveguide.

B. The elastic case

For an elastic bottom, in addition to the pressure-release boundary conditions at the ocean surface, the following boundary conditions at the water-bottom interface must be satisfied:

$$\begin{aligned} [\hat{\mathbf{n}} \cdot \mathbf{u}]_+^+ &= 0, \\ \hat{\mathbf{n}} \cdot \boldsymbol{\tau} \cdot \hat{\mathbf{n}} &= -p, \\ \hat{\mathbf{n}} \times (\boldsymbol{\tau} \cdot \hat{\mathbf{n}}) &= 0. \end{aligned} \quad (8)$$

The first equation expresses the continuity of normal displacement and the bottom two express the continuity of the normal component of the stress tensor, $\boldsymbol{\tau}$. We express the field in the bottom as a sum of virtual sources

$$\begin{aligned} \phi^{(2)}(\mathbf{r}) &= \mathbf{G}_\phi^{(2)}(\mathbf{r}; \mathbf{r}_2) \mathbf{s}_2, \\ \psi^{(2)}(\mathbf{r}) &= \mathbf{G}_\psi^{(2)}(\mathbf{r}; \mathbf{r}_3) \mathbf{s}_3. \end{aligned} \quad (9)$$

In the above equations ϕ and ψ are the scalar and shear potentials, satisfying

$$\begin{aligned} [\nabla^2 + k_p^2] \phi^{(2)} &= 0, \\ [\nabla^2 + k_s^2] \psi^{(2)} &= 0, \end{aligned}$$

where $k_p = \omega/c_p$ and $k_s = \omega/c_s$, and c_p and c_s are the compressional and shear speeds in the bottom, respectively. The position vectors, \mathbf{r}_2 and \mathbf{r}_3 , represent the location of sources producing the compressional and shear fields in the bottom. The compressional and shear potential Green's functions, $G_\phi^{(2)}$ and $G_\psi^{(2)}$, satisfy

$$\begin{aligned} [\nabla^2 + k_p^2] \mathbf{G}_\phi^{(2)}(\mathbf{r}, \mathbf{r}') &= \delta(\mathbf{x} - \mathbf{x}') \delta(\mathbf{z} - \mathbf{z}'), \\ [\nabla^2 + k_s^2] \mathbf{G}_\psi^{(2)}(\mathbf{r}, \mathbf{r}') &= \delta(\mathbf{x} - \mathbf{x}') \delta(\mathbf{z} - \mathbf{z}'), \end{aligned}$$

whose solutions are

$$\begin{aligned} \mathbf{G}_\phi^{(2)}(\mathbf{r}, \mathbf{r}') &= \frac{i}{4} \mathbf{H}_0^{(1)}(k_p |\mathbf{r} - \mathbf{r}'|), \\ \mathbf{G}_\psi^{(2)}(\mathbf{r}, \mathbf{r}') &= \frac{i}{4} \mathbf{H}_0^{(1)}(k_s |\mathbf{r} - \mathbf{r}'|). \end{aligned} \quad (10)$$

The displacement vector can be written in terms of two scalar potentials,

$$\mathbf{u} = \nabla \phi + \nabla \times (0, \psi, 0),$$

and the stress tensor in the bottom are related to the displacement according to

$$\tau_{ij} = \lambda \delta_{ij} \left(\frac{\partial \mathbf{u}_1}{\partial x_1} + \frac{\partial \mathbf{u}_2}{\partial x_2} + \frac{\partial \mathbf{u}_3}{\partial x_3} \right) + \mu \left(\frac{\partial \mathbf{u}_i}{\partial x_j} + \frac{\partial \mathbf{u}_j}{\partial x_i} \right).$$

In the above δ_{ij} is the Kronecker delta function and λ and μ are the Lamé constants,

$$\lambda = \rho(c_p^2 - 2c_s^2), \quad \mu = \rho c_s^2.$$

Note that in the above equations the arguments of the components of the fields, \mathbf{r} , have been suppressed; instead they

have been expressed using bold face character to indicate that they are vectors.

In two dimensions the components of the stress tensor can be written in terms of the displacement vector as

$$\tau_{xx} = (\lambda + 2\mu) \frac{\partial \mathbf{u}_x^{(2)}}{\partial x} + \lambda \frac{\partial \mathbf{u}_z^{(2)}}{\partial z},$$

$$\tau_{xz} = \mu \left(\frac{\partial \mathbf{u}_x^{(2)}}{\partial z} + \frac{\partial \mathbf{u}_z^{(2)}}{\partial x} \right),$$

$$\tau_{zz} = (\lambda + 2\mu) \frac{\partial \mathbf{u}_z^{(2)}}{\partial z} + \lambda \frac{\partial \mathbf{u}_x^{(2)}}{\partial x},$$

and the displacement vector can be expressed in terms of the potentials

$$\mathbf{u}_x^{(2)} = \frac{\partial \phi^{(2)}}{\partial x} - \frac{\partial \psi^{(2)}}{\partial z}, \quad \mathbf{u}_z^{(2)} = \frac{\partial \phi^{(2)}}{\partial z} + \frac{\partial \psi^{(2)}}{\partial x}.$$

To express the field at points (\mathbf{r}) due to sources located at (\mathbf{r}'), Eqs. (9) are used, which results in

$$\mathbf{u}_z^{(1)} = \mathbf{E}_z \mathbf{s}_1, \quad \mathbf{u}_x^{(1)} = \mathbf{E}_x \mathbf{s}_1,$$

$$\mathbf{u}_z^{(2)} = \mathbf{U}_{z_1} \mathbf{s}_2 + \mathbf{U}_{z_2} \mathbf{s}_3,$$

$$\mathbf{u}_x^{(2)} = \mathbf{U}_{x_1} \mathbf{s}_2 + \mathbf{U}_{x_2} \mathbf{s}_3,$$

where

$$\mathbf{E}_z = \frac{1}{\rho_1 \omega^2} \frac{\partial \mathbf{G}_p^{(1)}}{\partial z}, \quad \mathbf{E}_x = \frac{1}{\rho_1 \omega^2} \frac{\partial \mathbf{G}_p^{(1)}}{\partial x},$$

$$\mathbf{U}_{z_1} = \frac{\partial \mathbf{G}_\phi^{(2)}}{\partial z}, \quad \mathbf{U}_{z_2} = \frac{\partial \mathbf{G}_\psi^{(2)}}{\partial x},$$

$$\mathbf{U}_{x_1} = \frac{\partial \mathbf{G}_\phi^{(2)}}{\partial x}, \quad \mathbf{U}_{x_2} = -\frac{\partial \mathbf{G}_\psi^{(2)}}{\partial z}.$$

Similarly,

$$\tau_{xx} = [(\lambda + 2\mu) \mathbf{U}_{xx_1} + \lambda \mathbf{U}_{zz_1}] \mathbf{s}_2 + [(\lambda + 2\mu) \mathbf{U}_{xx_2} + \lambda \mathbf{U}_{zz_2}] \mathbf{s}_3,$$

$$\tau_{xz} = \mu [\mathbf{U}_{xz_1} + \mathbf{U}_{zx_1}] \mathbf{s}_2 + \mu [\mathbf{U}_{xz_2} + \mathbf{U}_{zx_2}] \mathbf{s}_3,$$

$$\tau_{zz} = [(\lambda + 2\mu) \mathbf{U}_{zz_1} + \lambda \mathbf{U}_{xx_1}] \mathbf{s}_2 + [(\lambda + 2\mu) \mathbf{U}_{zz_2} + \lambda \mathbf{U}_{xx_2}] \mathbf{s}_3.$$

In the above,

$$\mathbf{U}_{xx_1} = \frac{\partial^2 \mathbf{G}_\phi^{(2)}}{\partial x^2}, \quad \mathbf{U}_{zz_1} = \frac{\partial^2 \mathbf{G}_\phi^{(2)}}{\partial z^2},$$

$$\mathbf{U}_{xx_2} = -\frac{\partial^2 \mathbf{G}_\psi^{(2)}}{\partial x \partial z}, \quad \mathbf{U}_{zz_2} = \frac{\partial^2 \mathbf{G}_\psi^{(2)}}{\partial z \partial x},$$

$$\mathbf{U}_{xz_1} = -\frac{\partial^2 \mathbf{G}_\phi^{(2)}}{\partial x \partial z}, \quad \mathbf{U}_{zx_1} = \mathbf{U}_{xz_1},$$

$$\mathbf{U}_{xz_2} = -\frac{\partial^2 \mathbf{G}_\psi^{(2)}}{\partial z^2}, \quad \mathbf{U}_{zx_2} = \frac{\partial^2 \mathbf{G}_\psi^{(2)}}{\partial x^2}.$$

Let us write the unit normal to the interface as

$$\hat{\mathbf{n}} = n_x \hat{\mathbf{x}} + n_z \hat{\mathbf{z}},$$

and further define

$$\mathbf{A}_1 = (\lambda + 2\mu) \mathbf{U}_{xx_1} + \lambda \mathbf{U}_{zz_1},$$

$$\mathbf{A}_2 = (\lambda + 2\mu) \mathbf{U}_{xx_2} + \lambda \mathbf{U}_{zz_2},$$

$$\mathbf{B}_1 = (\lambda + 2\mu) \mathbf{U}_{zz_1} + \lambda \mathbf{U}_{xx_1},$$

$$\mathbf{B}_2 = (\lambda + 2\mu) \mathbf{U}_{zz_2} + \lambda \mathbf{U}_{xx_2},$$

$$\mathbf{C}_1 = \mathbf{U}_{xz_1} + \mathbf{U}_{zx_1},$$

$$\mathbf{C}_2 = \mathbf{U}_{xz_2} + \mathbf{U}_{zx_2}.$$

The boundary conditions given by Eq. (8) can be written

$$(n_x \partial \mathbf{p} / \partial x + n_z \partial \mathbf{p} / \partial z) / \rho \omega^2 = n_x \mathbf{u}_x^{(2)} + n_z \mathbf{u}_z^{(2)},$$

$$-\mathbf{p} = n_x^2 \tau_{xx} + n_z^2 \tau_{zz} + 2n_x n_z \tau_{xz},$$

$$(n_x^2 - n_z^2) \tau_{zx} + n_x n_z (\tau_{zz} - \tau_{xx}) = 0.$$

In the above equation \mathbf{p} is the total pressure in the water. Substituting for the pressure, the components of displacement, and the stress tensor in terms of the source amplitudes, we get

$$(n_x \partial \mathbf{p}_{inc} / \partial x + n_z \partial \mathbf{p}_{inc} / \partial z) / \rho \omega^2 + (n_x \mathbf{E}_x + n_z \mathbf{E}_z) \mathbf{s}_1 = (n_x \mathbf{U}_{x_1} + n_z \mathbf{U}_{z_1}) \mathbf{s}_2 + (n_x \mathbf{U}_{x_2} + n_z \mathbf{U}_{z_2}) \mathbf{s}_3,$$

$$-\mathbf{p}_{inc} - \mathbf{G}_p^{(1)} \mathbf{s}_1 = (n_x^2 \mathbf{A}_1 + n_z^2 \mathbf{B}_1 + 2n_x n_z \mathbf{C}_1) \mathbf{s}_2 + (n_x^2 \mathbf{A}_2 + n_z^2 \mathbf{B}_2 + 2n_x n_z \mathbf{C}_2) \mathbf{s}_3,$$

$$[(n_x^2 - n_z^2) \mathbf{C}_1 + n_x n_z (\mathbf{B}_1 - \mathbf{A}_1)] \mathbf{s}_2 + [(n_x^2 - n_z^2) \mathbf{C}_2 + n_x n_z (\mathbf{B}_2 - \mathbf{A}_2)] \mathbf{s}_3 = 0.$$

The above equations can be solved to give

$$\mathbf{s}_1 = (-\mathbf{G}_p^{(1)} - \mathbf{K} \mathbf{G}_u^{(1)})^{-1} (\mathbf{K} \mathbf{u}_{inc} + \mathbf{p}_{inc}),$$

$$\mathbf{s}_2 = (-\mathbf{L} \tilde{\mathbf{G}}_u^{(2)} - \tilde{\mathbf{G}}_p^{(2)})^{-1} (\mathbf{L} \mathbf{u}_{inc} + \mathbf{p}_{inc}),$$

$$\mathbf{s}_3 = -\mathbf{\Lambda}_3 \mathbf{\Lambda}_2 \mathbf{s}_2, \tag{11}$$

where

$$\mathbf{\Lambda}_2 = (n_x^2 - n_z^2) \mathbf{C}_1 + n_x n_z (\mathbf{B}_1 - \mathbf{A}_1),$$

$$\mathbf{\Lambda}_3 = (n_x^2 - n_z^2) \mathbf{C}_2 + n_x n_z (\mathbf{B}_2 - \mathbf{A}_2),$$

$$\tilde{\mathbf{G}}_p^{(2)} = \mathbf{\Gamma}_2 - \mathbf{\Gamma}_3 \mathbf{\Lambda}_3^{-1} \mathbf{\Lambda}_2,$$

$$\tilde{\mathbf{G}}_u^{(2)} = \mathbf{Y}_2 - \mathbf{Y}_3 \mathbf{\Lambda}_3^{-1} \mathbf{\Lambda}_2,$$

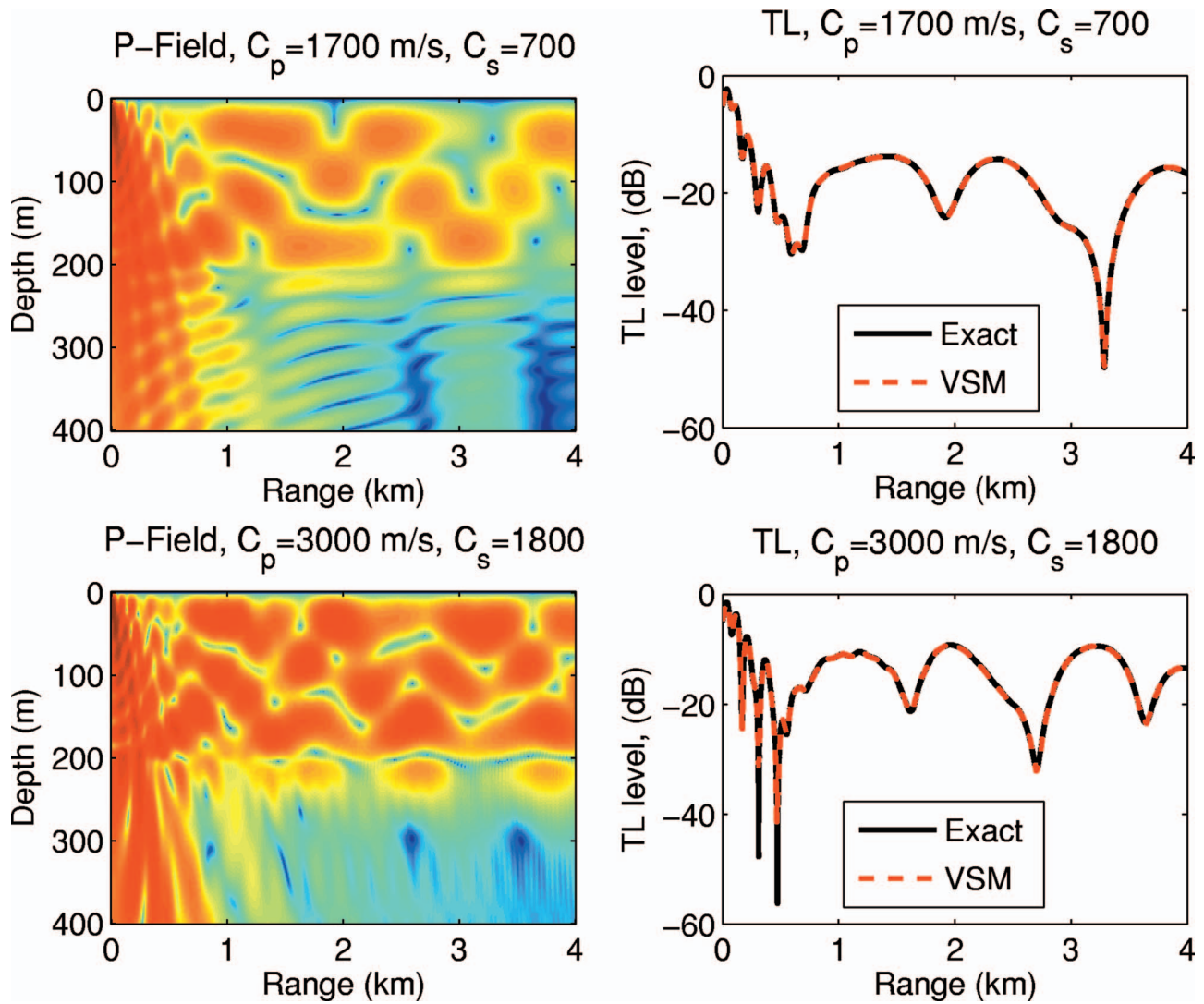


FIG. 1. Propagation in an elastic Pekeris waveguide: The top left panel shows the acoustic field for a 200 m deep water layer over an elastic bottom with compressional sound speed, $c_p=1700$ m/s, and shear sound speed, $c_s=700$ m/s. The bottom left panel shows the same for $c_p=3000$ m/s and $c_s=1800$ m/s. The right panels show a comparison of transmission loss obtained using the virtual source technique and an exact, wavenumber integration solution. The dynamic range in the field plots is from -60 to 0 dB.

$$\Gamma_2 = n_x^2 \mathbf{A}_1 + n_z^2 \mathbf{B}_1 + 2n_x n_z \mathbf{C}_1,$$

$$\Gamma_3 = n_x^2 \mathbf{A}_2 + n_z^2 \mathbf{B}_2 + 2n_x n_z \mathbf{C}_2,$$

$$\mathbf{Y}_2 = n_x \mathbf{U}_{x_1} + n_z \mathbf{U}_{z_1},$$

$$\mathbf{Y}_3 = n_x \mathbf{U}_{x_2} + n_z \mathbf{U}_{z_2},$$

$$\mathbf{K} = \tilde{\mathbf{G}}_p^{(2)} (\tilde{\mathbf{G}}_u^{(2)})^{-1},$$

$$\mathbf{L} = \mathbf{G}_p^{(1)} (\mathbf{G}_u^{(1)})^{-1}.$$

The source amplitudes given by the above equations completely specify the components of the field anywhere in the waveguide.

III. RESULTS

In this section we apply the model developed in the previous section to various propagation problems. As a first

example, we compute the pressure field in a range-independent waveguide composed of an isovelocity water layer over an elastic half-space bottom. Since this problem can be solved exactly by other methods, the purpose of presenting it is to validate the virtual source solution. In this example, a 25-Hz line source is placed at a depth of 30 m. The problem is solved for two sets of bottom parameters. In case 1 the bottom compressional sound speed $c_p=1700$ m/s and its shear sound speed $c_s=700$ m/s; in case 2, $c_p=3000$ m/s and $c_s=1800$ m/s. In both cases, the bottom compressional attenuation is 0.5 dB/ λ and its shear attenuation is 0.25 dB/ λ . Attenuation enters the computations through the complex wavenumbers according to $k_{p,s} = \omega/c_{p,s}(1 + \lambda_{p,s}/54.58i)$.²³ The results are shown in Fig. 1. The two panels on the left show the compressional pressure field in the waveguide for the two cases and the two panels on the right show transmission loss as a function of range for a receiver at 50 m. The transmission loss obtained from the virtual source technique for the two cases is compared with

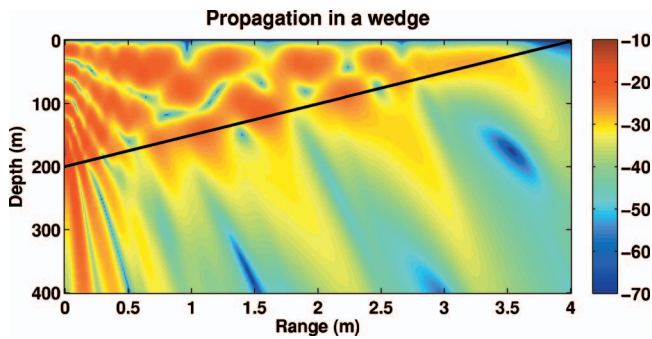


FIG. 2. Propagation in an oceanic wedge using the virtual source technique: In the above figure a 25 Hz source is located at 180 m. The water depth, which is 200 m at the source location, decreases to zero at 4000-m range, resulting in a wedge angle of approximately 2.86 deg. As can be seen, the virtual source technique correctly accounts for mode cutoff.

the exact solution. It should be pointed out that in all examples in this paper we used ten virtual sources per wavelength: It can be seen that the virtual source solutions agree very well with the exact solution. It should be noted that the virtual source technique is designed to force boundary conditions at an arbitrarily shaped boundary or interface. In this example, the method is validated for a simple boundary only because it offers an exact solution. However, this validation confirms that the method is based on sound foundations and gives reason to believe that it can produce equally accurate results for more complicated boundaries. The exact solution, which is used in the validation, will be described in Appendix B.

As a second example, we apply the model to propagation in an oceanic wedge. The wedge is modeled exactly as the ASA Benchmark problem.²⁴ It is composed of a water layer over an acoustic (no shear) half-space bottom. The water depth decreases as a function of range from 200 m at zero range to 0 at 4000 m range, resulting in a wedge angle of approximately 2.86 deg. The source is a 25-Hz line source,

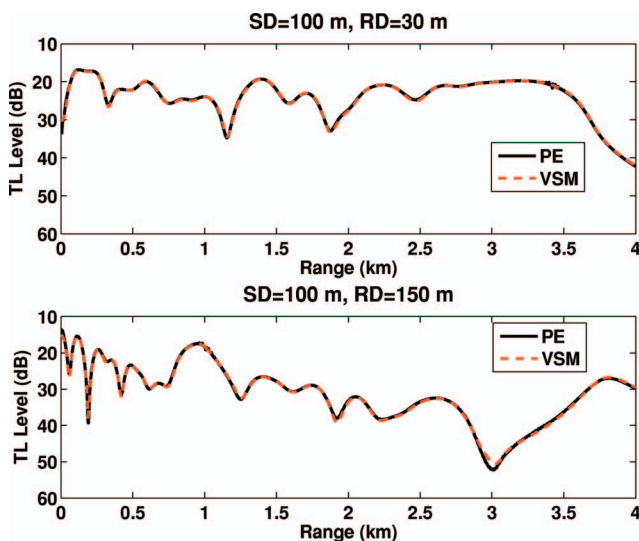


FIG. 3. The above figure shows a comparison of transmission loss obtained using the virtual source technique and the parabolic equation (PE) in an oceanic wedge. The line source is located at 100 m. The receiver in the top figure is at 30 m and in the bottom figure is at 150 m.

the bottom sound speed is 1700 m/s, and its attenuation is 0.5 dB/λ. The acoustic field in the wedge is shown in Fig. 2. The black line in Fig. 2 indicates the water/bottom interface. As the water depth decreases as a function of range, the three propagating modes cut off one after the other. As can be seen in the figure, the model correctly accounts for this process. Figure 3 shows a comparison of transmission loss for two receiver depths, 100 and 150 m, computed using our model and the parabolic equation (PE) model for a line source. The source in both cases was at 100 m. The small differences between the virtual source solution and the PE solution may be associated with the fact that the PE solution is one-way, accurate within the accuracy of the paraxial approximation, where the virtual source solution is an exact solution.

As a final example, we computed the compressional and shear potentials for an elastic wedge. The results are shown in Figs. 4 and 5. In Fig. 4 the bottom compressional sound speed $c_p=1700$ m/s and its shear sound speed $c_s=700$ m/s. The top panel shows the compressional potential field and the bottom panel shows the shear potential field for a 25-Hz source located at 180 m. Figure 5 shows the same for the case when $c_p=3000$ m/s and $c_s=1800$ m/s. The results for this example compare well with those obtained using the parabolic equation method in Fig. 3 of Ref. 25. The source distribution for this example is shown in Fig. 6.

IV. SUMMARY

Boundary integral methods essentially take advantage of the divergence theorem to replace a partial differential equation in a volume with an integral equation along a boundary, with the free-space Green's function expressing how points on the boundary mutually affect each other. The virtual source method is really just a boundary integral method that emphasizes the view of points on the boundary as discrete sources and receivers.

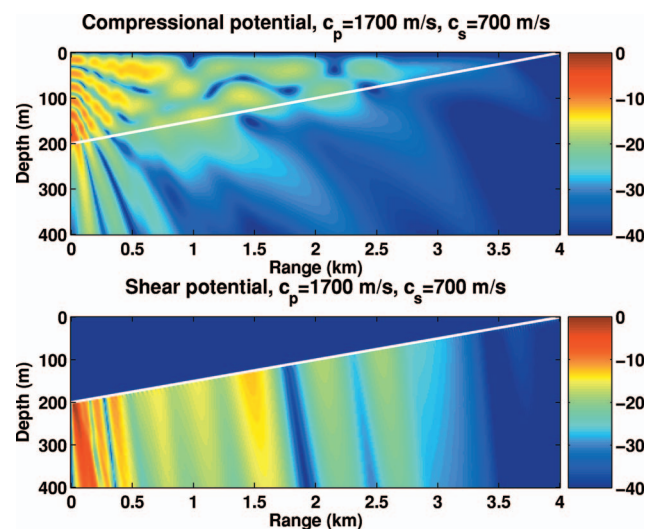


FIG. 4. The above figure shows the compressional and shear potentials in an elastic wedge in the case where the compressional sound speed $c_p=1700$ m/s and the shear sound speed $c_s=700$ m/s in the bottom. A 25 Hz source is located at 180 m and the compressional and shear attenuation are 0.5 dB/λ and 0.25 dB/λ, respectively.

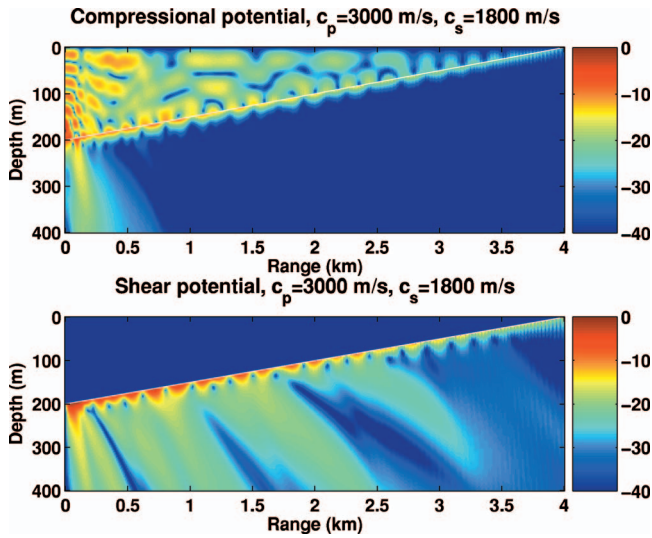


FIG. 5. The above figure shows the compressional and shear potentials in an elastic wedge in the case where the compressional sound speed $c_p = 3000$ m/s and the shear sound speed $c_s = 1800$ m/s in the bottom. A 25-Hz source is located at 180 m and the compressional and shear attenuation are 0.5 dB/ λ and 0.25 dB/ λ , respectively. These results compare well with those obtained using the PE method in Fig. 3 of Ref. 25.

To date, the virtual source method has been used principally for target scattering problems in ocean waveguides—the target then provides a natural boundary. However, the method has much broader applicability. The work of Fawcett¹⁸ for a rough surface patch provides an important example of that. In this paper we have shown that the method can be used for complicated elastic range-dependent waveguides, providing an essentially exact solution. The application to elastic problems is particularly important in providing benchmark-quality solutions for other emerging methods for elastic problems based on the parabolic equation, finite elements, and coupled modes. While simple geometries, such as the wedge, have been shown here, we emphasize that the method is not restricted to flat boundaries. Further, problems with sound-speed gradients may be handled by substituting the appropriate Green's function, which in turn may be computed in various ways. This latter generalization will be the subject of a future paper.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors would like to thank the reviewers for their thorough evaluation of this paper and their constructive comments.

APPENDIX A: THE VIRTUAL SOURCE TECHNIQUE IN AN AZIMUTHALLY SYMMETRIC OCEAN

In an azimuthally symmetric ocean any source not located at the origin is a ring source centered at the origin. To derive the Green's function for a ring source of source strength S_w in free space, consider the wave equation

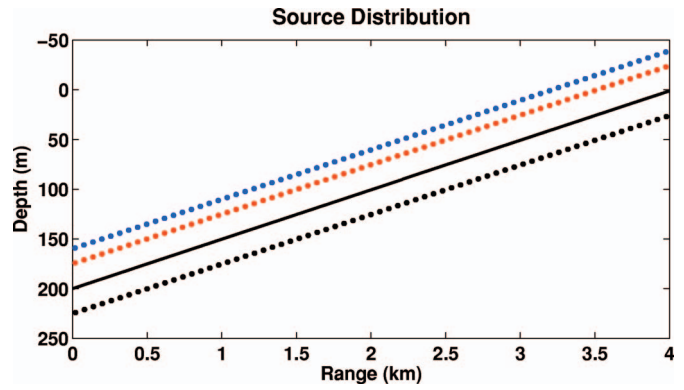


FIG. 6. This figure shows the source distribution for the example of propagation in an elastic wedge. To be able to see the sources, only one in every ten have been shown. The solid line shows the water/bottom interface. The sources represented by the black dots are those that produce the field in the water and those represented by the red and the blue dots are the ones that produce the compressional and the shear potentials in the bottom, respectively.

$$[\nabla^2 + k^2]\psi(r, z) = S_w \delta(z - z_s) \frac{\delta(r - r_s)}{2\pi r}. \quad (\text{A1})$$

Performing the forward Hankel transform,

$$f(k_r, z) = \int_0^\infty \tilde{f}(r, z) J_0(k_r r) r dr,$$

on the above equation gives

$$\left(\frac{d^2}{dz^2} + k_z^2\right)\tilde{\psi}(k_r, z) = \frac{S_w}{2\pi} \delta(z - z_s) J_0(k_r r_s),$$

where $k_z^2 = (k^2 - k_r^2)$. The solution of the above equation is given by²⁶

$$\tilde{\psi}(k_r, z) = -S_w J_0(k_r r_s) \frac{e^{ik_z|z-z_s|}}{4\pi i k_z}. \quad (\text{A2})$$

By performing the inverse Hankel transform on Eq. (A2), the Green's function for a ring source in an azimuthally symmetric ocean can be obtained:

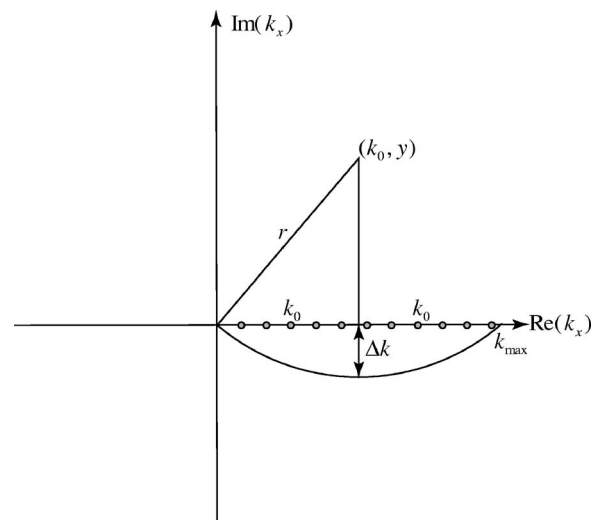


FIG. 7. The arc contour used in the integration of Eq. (B2).

$$G(r, z; r_s, z_s) = -\frac{S_w}{4\pi i} \int_0^\infty \frac{e^{ik_z|z-z_s|}}{k_z} J_0(k_r r) J_0(k_r r_s) k_r dk_r. \quad (\text{A3})$$

To be able to treat propagation for a point source using the virtual source technique in the wedge described in Secs. II A and II B, assuming that the wedge is azimuthally symmetric, the free space Green's function given by Eq. (A3) should be used instead of the free space Green's function for a line source.

APPENDIX B: AN EXACT WAVENUMBER INTEGRAL SOLUTION

Consider a waveguide composed of an isovelocity water layer of depth h over an elastic half space bottom with density ρ_b and compressional and shear sound speeds c_p and c_s . The spectral components of the potential field for a line source located at $z=z_s$ satisfy the depth-separated wave equation

$$\left[\frac{d^2}{dz^2} + (k^2 - k_x^2) \right] \phi(k_x, z) = S_w \frac{\delta(z - z_s)}{2\pi},$$

where

$$\begin{pmatrix} 1 & e^{ik_{z_1} h} & 0 \\ e^{ik_{z_1} h} & 1 & -\frac{\rho_b}{\rho_w} \left(1 - 2 \frac{k_x^2}{k_s^2} \right) \\ ik_{z_1} e^{ik_{z_1} h} & -ik_{z_1} & -ik_{z_2} \\ 0 & 0 & 2k_x k_{z_2} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} S_w \frac{ie^{ik_{z_1} z_s}}{4\pi k_{z_1}} \\ S_w \frac{ie^{ik_{z_1} |h-z_s|}}{4\pi k_{z_1}} \\ -S_w \frac{ie^{ik_{z_1} |h-z_s|}}{4\pi} \\ 0 \end{pmatrix}.$$

To convert the solution back to the spatial domain, Eq. (B1) is used. However, since the kernel of the integral is even with respect to k_x , Eq. (B1) can be written

$$\phi(x, z) = 2 \int_0^{k_{\max}} \tilde{\phi}(k_x, x) \cos(k_x x) dk_x, \quad (\text{B2})$$

where k_{\max} is the maximum value of k_x at which the integral is truncated. In this paper, k_{\max} was chosen to be $2k_w$, where $k_w = \omega/c_w$. Since the integrand has poles on the real axis, which correspond to the eigenvalues of the waveguide, we integrated Eq. (B2) along the arc contour shown in Fig. 7. The contour, which is a section of a circle of radius r , centered at (k_0, y) , is designed such that its maximum distance from the real axis, Δk , is always equal to the wavenumber sampling, Δk_x . Based on this consideration, the center of the circle, (k_0, y) , is given by

$$\phi(x, z) = \int_{-\infty}^{\infty} \tilde{\phi}(k_x, x) e^{ik_x x} dk_x. \quad (\text{B1})$$

In the spectral domain, the potential in the water due to a source at $z=z_s$ can be written

$$\tilde{\phi}_1(k_x, z) = S_w \frac{e^{ik_{z_1}|z-z_s|}}{4\pi i k_{z_1}} + A e^{ik_{z_1} z} + B e^{-ik_{z_1}(z-h)}.$$

In the above equation the first term is the contribution due to the source and the second and the third terms are downward and upward propagating waves, with $k_{z_1} = \sqrt{k_1^2 - k_x^2}$, where $k_1 = \omega/c_1$. In the bottom the upward propagating wave must vanish due to the radiation boundary condition at infinity. The compressional and shear potentials thus can be written as

$$\tilde{\phi}_2(k_x, z) = C e^{ik_{z_2}(z-h)},$$

$$\tilde{\psi}_2(k_x, z) = D e^{i\kappa_{z_2}(z-h)},$$

where $k_{z_2} = \sqrt{k_p^2 - k_x^2}$ and $\kappa_{z_2} = \sqrt{k_s^2 - k_x^2}$, with $k_p = \omega/c_p$ and $k_s = \omega/c_s$. The coefficients A , B , C , and D are determined by satisfying the pressure release boundary condition at $z=0$ and by demanding the continuity of the normal and tangential components of the stress tensor and the normal component of the displacement vector at $z=h$. The result is

$$k_0 = k_{\max}/2,$$

$$y = (k_0^2 - \Delta k_x^2)/2\Delta k_x.$$

The points on the contour satisfy

$$k = k_x + ik_i,$$

where

$$k_i = y - \sqrt{r^2 - (k_0 - k_x)^2}.$$

¹F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, in *Computational Ocean Acoustics* (AIP, New York, 1994), Chaps. 4–6.

²A. D. Pierce, "Extension of the method of normal modes to sound propagation in an almost-stratified medium," *J. Acoust. Soc. Am.* **37**, 19–27 (1965).

³D. M. Milder, "Ray and wave invariants for Sofar channel propagation," *J. Acoust. Soc. Am.* **46**, 1259–1263 (1969).

- ⁴S. R. Rutherford and K. E. Hawker, "Consistent coupled mode theory of sound propagation for a class of nonseparable problems," *J. Acoust. Soc. Am.* **70**(2), 554–564 (1981).
- ⁵J. A. Fawcett, "A derivation of the differential equations of coupled-mode propagation," *J. Acoust. Soc. Am.* **92**(1), 290–295 (1992).
- ⁶R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," *J. Acoust. Soc. Am.* **74**(1), 188–195 (1983).
- ⁷B. E. McDonald, "Bathymetric and volumetric contributions to ocean acoustic mode coupling," *J. Acoust. Soc. Am.* **100**(1), 219–224 (1996).
- ⁸A. T. Abawi, W. A. Kuperman, and M. D. Collins, "The coupled mode parabolic equation," *J. Acoust. Soc. Am.* **102**(1), 233–238 (1997).
- ⁹A. T. Abawi, "An energy-conserving one-way coupled mode propagation model," *J. Acoust. Soc. Am.* **111**(1), Pt. 1, 160–167 (2002).
- ¹⁰F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Under water Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, New York, 1977), pp. 224–281.
- ¹¹M. D. Collins, "A split-step padé solution for the parabolic equation method," *J. Acoust. Soc. Am.* **93**(4), 1736–1742 (1993).
- ¹²R. I. Odom, "A coupled mode examination of irregular waveguides including the continuum spectrum," *Geophys. J. R. Astron. Soc.* **86**, 425–453 (1986).
- ¹³J. Tromp, "A coupled local-mode analysis of surface-wave propagation in a laterally heterogeneous waveguide," *Geophys. J. Int.* **117**, 153–161 (1994).
- ¹⁴P. R. Stepanishen, "A generalized internal source density method for the forward and backward projection of harmonic pressure fields from complex bodies," *J. Acoust. Soc. Am.* **101**(6), 3270–3277 (1997).
- ¹⁵G. H. Koopman, L. Song, and J. B. Fahline, "A method for computing acoustic fields based on the principle of wave superposition," *J. Acoust. Soc. Am.* **86**(6), 2433–2438 (1989).
- ¹⁶J. B. Fahline and G. H. Koopman, "A numerical solution for the general radiation problem based on the combined methods of superposition and singular-value decomposition," *J. Acoust. Soc. Am.* **90**(5), 2808–2819 (1991).
- ¹⁷M. Ochmann, "The source simulation technique for acoustic radiation problems," *Acustica* **81**, 512–527 (1995).
- ¹⁸J. A. Fawcett, "A scattering-chamber approach for solving finite rough surface scattering problems," *J. Acoust. Soc. Am.* **118**(3), 1348–1357 (2005).
- ¹⁹F. B. Jensen and W. A. Kuperman, "Sound propagation in a wedge shaped ocean with penetrable bottom," *J. Acoust. Soc. Am.* **67**, 1564–1566 (1980).
- ²⁰A. B. Coppens and J. V. Sanders, "Transmission of sound into a fast fluid bottom from an overlaying fluid wedge," in *Proceedings of Workshop on Seismic Propagation in Shallow Water* (Office of Naval Research, Arlington, VA, 1978).
- ²¹J. M. Arnold and L. B. Felsen, "Rays and local modes in a wedge-shaped ocean," *J. Acoust. Soc. Am.* **73**(4), 1105–1118 (1983).
- ²²J. M. Arnold and L. B. Felsen, "Coupled mode theory of intrinsic modes in a wedge," *J. Acoust. Soc. Am.* **79**(1), 31–40 (1986).
- ²³F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, in *Computational Ocean Acoustics* (AIP, New York, 1994), Chap. 1.
- ²⁴F. B. Jensen and C. M. Ferla, "Numerical solution of range-dependent benchmark problems in ocean acoustics," *J. Acoust. Soc. Am.* **87**(4), 1499–1510 (1990).
- ²⁵M. D. Collins, "An energy-conserving parabolic equation for elastic media," *J. Acoust. Soc. Am.* **94**(2), Pt. 1, 975–982 (1993).
- ²⁶F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, in *Computational Ocean Acoustics* (AIP, New York, 1994), Chap. 4.

Gray whale target strength measurements and the analysis of the backscattered response

Irena Lucifredi^{a)} and Peter J. Stein
Scientific Solutions Inc., Nashua, New Hampshire 03063

(Received 21 August 2006; revised 11 December 2006; accepted 2 January 2007)

One of the current Integrated Marine Mammal Monitoring and Protection System (IMPAS) directions is concentrated on the design and development of the active sonar modality representing just one component of the global system. The active sonar was designed, built, and tested during the Marine Mammal Active Sonar Test (MAST 04), producing whale detections and whale tracks. The experiment was conducted in January 2004 off the coast of California. One of the objectives of the current work is to distinguish whale backscattered responses from the ones generated by the environmental clutter in a waveguide. Furthermore, the work aims to identify and analyze the target signature features that are necessary for enhanced active sonar detection and classification of marine mammals. Over the years there have been very few documented attempts to capture and analyze the backscattering response of whales using an active sonar system. Nevertheless, whales, mostly owing to their size, their motion, and the aspect dependence of their backscattered field, possess desirable properties that help distinguish their scattered response from clutter and other environment related false alarms. As an initial step, data collected during the MAST 04 experiment are presented, and gray whale target strength measurements are obtained. Results are compared to the previously published whale target strengths. Additionally, an investigation is conducted in an effort to provide whale feature identification points suitable for automated detection and classification, as means of relating gray whale active acoustic signatures to their inherent characteristics and their motion.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2436643]

PACS number(s): 43.30.Sf, 43.40.Fz [WWA]

Pages: 1383–1391

I. INTRODUCTION

Over the years, there have been very few documented attempts to capture and analyze the backscattering response of whales using an active sonar system. Nevertheless, whales, mostly owing to their size and their motion, possess desirable properties that help distinguish their scattered response from clutter and other environment related false alarms. While so far in literature there has not been a comprehensive analysis of the properties and identification features of whale backscattered response, there have been a couple of documented measurements of different whale species target strengths. In the 1960s Dunn¹ reported a target strength measurement of -8 dB at 1 kHz of a sperm whale, *Physeter catodon*, of unknown length insonified at head aspect by an explosive signal. A few years later, Love² made target strength measurements of humpback whales from a tower in the Florida Straits, coming up with a target strength of $+8$ dB at broadside for a 15-m-long whale insonified at 20 kHz. More recently, Miller³ measured the target strength of right whales as a function of aspect using an 86.25 kHz phased array sonar. He found that the target strength of a 15-m-long whale at broadside varied from -4 to -8 dB. In addition, he established that the head on target strength of a juvenile 8-m-long right whale varies from about -7 to -13 dB. Miller also offered a possible explanation for the difference in obtained measurements between humpback

whales and right whales, mostly attributing the lower target strength measurements to the presence of a thick lossy blubber layer on the right whale. A short while after Love published his results, Levenson⁴ measured the bistatic target strength of sperm whales at similar frequencies to Love's, obtaining the maximum target strength of $+14.4$ dB.

The broad objective of the Integrated Marine Mammal Monitoring and Protection System (IMAPS) is to establish a global solution for systematic assessment and protection of the marine environment for the Navy, offshore industry, and scientific research community, through the development and implementation of a comprehensive, modular tool. There is a need for all the sources of information related to the detection and classification of marine mammals to be fused and to operate in real-time in order to provide a robust protection, assessment, and research system. One of the central objectives of the current work is to distinguish whale backscattered responses from the ones generated by the environmental clutter in a waveguide, and to identify and analyze the target signature features that are necessary for enhanced detection and classification of marine mammals. In this regard, the use of active sonar acoustic target signatures for detection and classification has a potential of being a realistic complement to whale passive sonar detections in particular in the medium to short ranges of up to 2 km, with the advantages of high active sonar resolution, the fact that whale vocalizations are not necessary, and the backscattered responses that are rich in information about the target size and the composition. To develop active acoustic signature classification, it is clearly necessary that the relation between the

^{a)}Electronic mail: euler001@yahoo.com

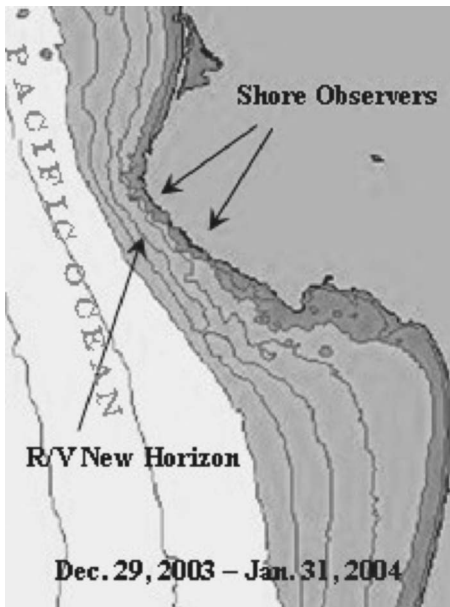


FIG. 1. MAST 04 experiment was conducted in January 2004 one mile off the coast of California, in 70 to 100-m deep water. The active sonar was bottom-stationed with the upward pointing projector beam to insonify the water column. The visual observers were placed on the shore as well as on the R/V New Horizon.

target characteristics and the acoustic scattering is well understood. This paper presents the data collected during the Marine Mammal Active Sonar Test (MAST 04), describes the signal processing tools used in the analysis, and provides some initial whale feature identification points as means of relating whale active acoustic signatures to their inherent characteristics. Rather than concentrating on the theoretical background of marine mammal detection, it is the rarity of the measurement and the novelty of the approach that is the current focus of this work.

II. MAST 04 EXPERIMENT

A series of MAST experiments is being conducted with a long-term objective of developing new sonar and integrated sensor concepts which exploit the information about target characteristics available in the three-dimensional multistatic field, and to attain improved detection, tracking, and classification of marine mammals. As shown in Fig. 1, MAST 04 experiment was conducted in January 2004 one mile off the coast of California by San Luis Obispo, in 70 to 100-m-deep water. The visual observers were placed on the shore as well as on the R/V New Horizon. A phase of the MAST 04 experiment consisted of a monostatic, bottom-stationed active sonar system deployed from the R/V New Horizon, with the upward pointing projector beam to insonify the water column. The active sonar consisted of an electronically steered vertical line array source with a frequency range of 21–25 kHz which was operated at a source level of up to 214 dB re μPa at 1 m producing a vertical directional beam of 6° beamwidth beamsteered 6° from the horizontal, and a cylindrical 60 element receiver array of 27 cm radius, used to beamform the received signals in the horizontal plane. Figure 2 shows the cylindrical receiver ar-

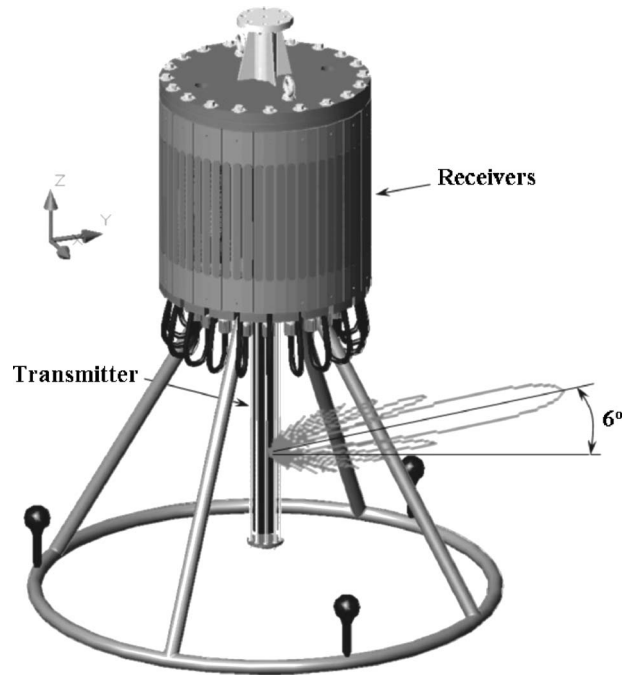


FIG. 2. The cylindrical receiver array with the vertical line array source mounted below it.

ray with the vertical line array source mounted below it. The operation of the system consisted of the transmitter projecting the vertically directional and horizontally omnidirectional beam in the water column at a ping rate of 5 s. The whale backscattered response was captured by the cylindrical receiver in a monostatic position on all 60 channels which can be later used to beamform the raw data in the horizontal direction using the plane wave beamforming approach⁵ and to automatically detect and track the whales using the tracking detection and estimation algorithms⁶ in real time. As the sound propagates through the water column it interacts with the surface, bottom, kelp, and other nontarget clutter and is also backscattered and captured by the system as shown in the following sections by other high level responses present in the data. Simultaneously with the active sonar operation, whale observers both on shore and on the ship were monitoring the whale migration providing positive identification, time, range, and bearing of the surfacing pods or individual mammals.

III. DATA ANALYSIS AND RESULTS

Measurements of the backscattered response of gray whales were made during their January migration off the coast of California. The analysis presented here included the following data processing: plane wave beamforming of the raw data in the horizontal plane, match-filtering of the beamformed response, projection of the beamformed match-filtered output to the range-ping and to the range-range space, backscattered time series analysis, whale target strength determination as a function of aspect, time-frequency analysis, and some whale behavioral analysis. In addition to providing the active sonar maximum detection range and resolution, the target strength measurements, as well as capturing physical attributes that help discriminate

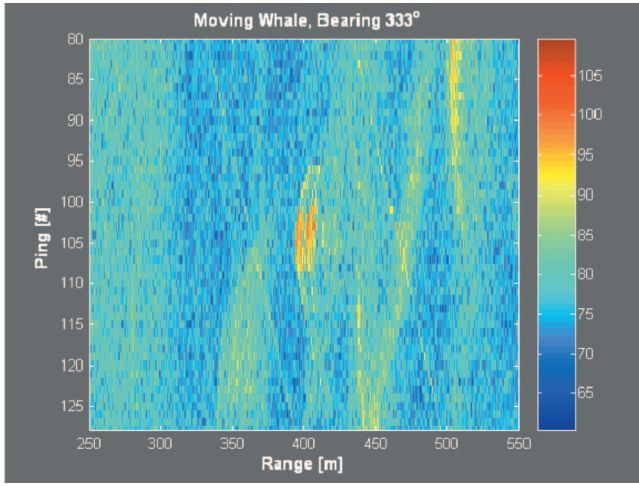


FIG. 3. A detected whale pod represented in the range-ping space at the bearing of 333°, range of 400 m, and across ping numbers 103–108. The result shows that the whales are clearly detectable among the surface, bottom, and water-column clutter, as they pass by the sonar, and that our active sonar implementation is capable of resolving the two distinct whales in the pod.

between target and clutter returns, the analysis also provides a basis for the verification of a potential whale target scattering model. Having a goal of achieving robust recognition performance, successful wave form classification algorithms strongly depend on target feature extraction and feature optimization. Some of the quality feature attributes consist of being independent of other features, being mathematically definable, being explainable in physical terms, and being computationally inexpensive. In order to provide physical inputs for integrated feature extraction and optimization processes, the analysis presented here transforms the raw data onto appropriate projection spaces in which target and clutter characteristics can be better captured and are less sensitive to noise and environmental perturbations.

A. Range-ping space

Most of the whale backscattered measurements were made from animals 400–1000 m away representing the long ranges that are usually of interest in mitigation operations. In Fig. 3 the beamformed, match-filtered result is shown in range-ping space as a function of the bearing indicated on the top of the figure as 333°. To obtain this result the raw data were quadrature demodulated⁷ using the sampling frequency of 82 kHz, plane wave beamformed⁵ in the azimuth along a circular array consisting of 60 receivers 2.87 cm apart, and finally match-filtered⁸ using a replica of the transmitted hyperbolic frequency modulated signal with the frequency band of 21–25 kHz.

Quadrature demodulation of a bandpass, real valued signal $x(t)$ effectively represents the signal by an equivalent low pass signal $\tilde{x}(t)$, and is given by

$$\tilde{x}(t) = u_c(t) + iu_s(t), \quad (1)$$

where

$$x(t) = u_c(t)\cos 2\pi F_c t - u_s(t)\sin 2\pi F_c t, \quad (2)$$

F_c being the center frequency of the bandpass signal corresponding to 23 kHz. In addition, the demodulated data have been beamformed in the frequency domain using a plane wave beamformer as a dot product of a weight vector and the vector of inputs of length N according to

$$\hat{Y}_m(k) = \sum_{i=0}^M a_i e^{-j2\pi k(i-1)\Delta_m/N} \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j2\pi kn/N}, \quad (3)$$

where the sensor weighing factor is given by a_i and k is the discrete fourier transform (DFT) bin number. To perform beamforming over the frequency band, beamformer output at the range of frequencies is generated at each of the fixed-beam directions given by a beam number m . The delay for the elements of the array is given by Δ_m according to the circular array configuration as defined in Ref. 9.

Finally, the output of the beamformer has been match filtered as given by the optimum filter⁸ to obtain the result shown in Fig. 3. The two whales are detected, and shown here at pings 103–108, as they pass through the closest point of approach, around 400 m away from the sonar. This result shows that the whales are clearly detectable among the surface, bottom and water-column clutter, as they pass by the sonar. Some of the shown features useful in discriminating the whales from the clutter are: the strength of their response as compared to the rest of the received field, their specific speed and motion, and their short spatiotemporal characteristics that differ from the ones related to clutter, which is characteristically streaky, as it typically shows across many pings at a relatively fixed range away from the sonar. This particular data set in the range-ping space shows that the magnitude of the whale response is 8–19 dB higher than the magnitude of the surrounding clutter observed in time, as the whales travel along their trajectory from 315° to 357° in bearing. Also, due to the appreciable whale speed and their discrete size, spatiotemporal characteristics of whale response are compact, spanning only across 8 pings for a certain bearing, while clutter response often stretches across 30–50 pings for a fixed bearing snapshot, as shown in Fig. 3 at 500 m range and ping numbers 80–103. In addition to the spatiotemporal characteristics, Fig. 3 shows that our active sonar implementation is capable of resolving the two distinct whales, as later independently confirmed by the visual observers who reported two whales in this particular pod.

B. Backscattered time series

Another space useful for whale backscattered response analysis is the normalized time series of the beamformed, match-filtered data. In this representation, the data set is shown as time series for a set of different bearings changing as a function of time represented by the ping number on the top of the figures. The X axis represents the range and the Y axis represents the backscattered magnitude. In Fig. 4, the high peak responses of the two whales passing through different bearings at 400 m range are shown. The conversion factor to obtain the y axis in dB//1 μ Pa representing the system received level is 0.784. Here the snapshot of the

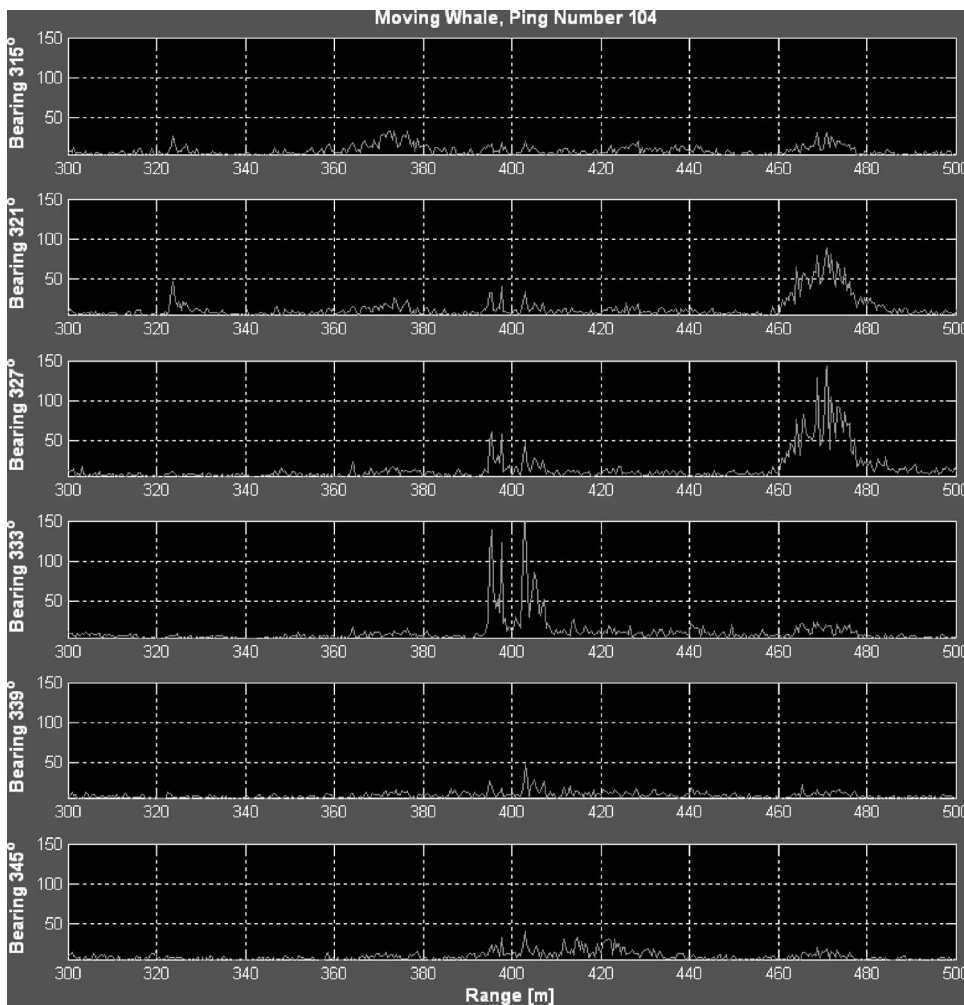


FIG. 4. A snapshot of the movie representing the beamformed, match-filtered time series for a set of different bearings changing as a function of time, represented by the ping number on the top. High peak responses of the two whales passing through different bearings are shown at 400 m range and the bearing of 333°. The whales can be clearly detected in this manner out to 920 m.

movie shows the strongest whale response at the bearing of 333°. At later times, even in the high-clutter environment, the whales can be clearly detected in this manner out to 920 m, and using tracking algorithms they are detectable well beyond 1 km.

In addition to determining the maximum detection range, and the magnitude of the whale response, a temporal target identification and classification feature that can be explored in this data representation is represented by the target range extent. Using the measured whale range extent and the identifiable local multipath structure, it is possible to enhance the detection and classification capabilities of the real time wave form processing algorithms by rejecting the long or short, duration clutter. In order to obtain the necessary information for this type of processing, the measurements of the target extent are essential. By measuring the maximum width of the beamformed, match-filtered response of a single whale and averaging it over 36 pings, the whale range extent of 12.5 m was determined from this data set.

C. Whale target strength measurements

Over the years there have been very few documented attempts to capture and analyze the target strength of whales using an active sonar system. Here, gray whale target strength measurements are obtained at long ranges, based on the data collected during the MAST 04 experiment. The re-

sults are analyzed and compared to the previously published whale target strengths. Target strength, TS of a scattering object as defined in Ref. 10 in terms of the received level RL, transmission loss TL, and source level SL is given by

$$TS = RL - SL + 2TL. \quad (4)$$

Therefore, from Eq. (4), a monostatic target strength of an object is a function of the received level produced by the scattered field, the source level, and the transmission loss to-and-from the target. All of these variables in Eq. (4) necessary for the whale target strength calculation were known to the authors, except for the accurate transmission loss as produced by the propagation of sound through a waveguide. It should be noted that the transmission loss determined in the following section accounts for all of the mechanisms of propagation coherently, including the spreading, attenuation, multipath, elastic bottom, and waveguide interference effects, as opposed to the previous target strength determination methods which only account for the spherical spreading and attenuation not taking into account the waveguide effects.

1. The transmission loss

In order to accurately determine the transmission loss in a waveguide at long ranges from the source, OASES wave number integration, sound propagation model was used to

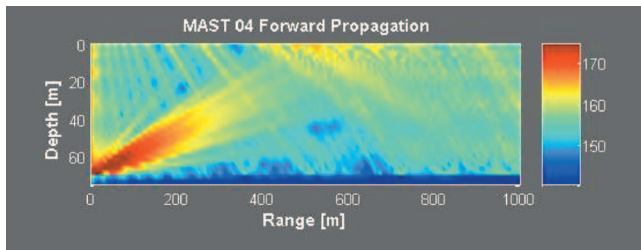


FIG. 5. OASES model result simulating the incident field in the water column for the system characteristics using the environmental parameters that correspond to the MAST 04 experiment.

simulate the forward propagation loss for our system characteristics using the environmental parameters that correspond to the MAST 04 experiment. OASES is a three-dimensional fast field program model for wave propagation in stratified environments, written by Schmidt¹¹ based on the global matrix method. It allows for a more efficient, and more numerically stable wave propagation computation than the traditional propagator matrix methods. As shown in Fig. 5, the system was located on the bottom of the 70-m-deep water column, insonifying it using a 6° wide beam electronically steered 6° up from the bottom. The XBT measured sound velocity profile, obtained during the experiment, was used in the model. Figure 5 shows the resulting modeled incident field in the water column generated at the center frequency of 23 kHz, with the color scale indicating the magnitude of the field. As expected, it is evident that the magnitude of the incident field in the water column varies as a function of range and depth. The roughness of the sea surface of 0.4 m was included in the computation, while the bathymetry was assumed constant at 70 m depth. The following geophysical properties were used in the model: the frequency $f=23$ kHz corresponding to the center frequency of the incident field, the water sound speed c_w varied from 1497.5 to 1495.0 m/s for depths 0 to 70 m, the water density $\rho_w=1000$ kg/m³, the sediment compressional sound speed $c_b=1640$ m/s the sediment attenuation $\lambda_b=0.5$ dB/ λ , the sediment density $\rho_b=1900$ kg/m³, while the shear properties were not taken into account as the sandy bottom does not support shear. Substantiated by the generally sandy bottom of this region, the values for the geoacoustic properties used correspond to the parameters appropriate to the predominantly sandy bottom type.¹² The previously measured¹⁰ Pacific Ocean seawater absorption value of 2 dB/kiloyard at 23 kHz was used. Based on the ray path representation of the scattered field, the overall propagation loss to-and-from the target for the system in a monostatic configuration is given by two times the forward propagation loss.¹⁰ In addition, it has been established in Refs. 13 and 14 that gray whales in the migratory pattern, as is the case here, spend over 90% of their time diving at constant depths 20–35 m below the surface. Figure 5 indicates that for ranges 400–900 m, which are relevant for our analysis, the incident field varies with 20–35 m depth by less than 3 dB.

In addition, some forward propagation measurements were taken during the experiment and are shown here in Fig. 6. A single hydrophone was placed at four different ranges at the depth of 30 m in the water column. The measured re-

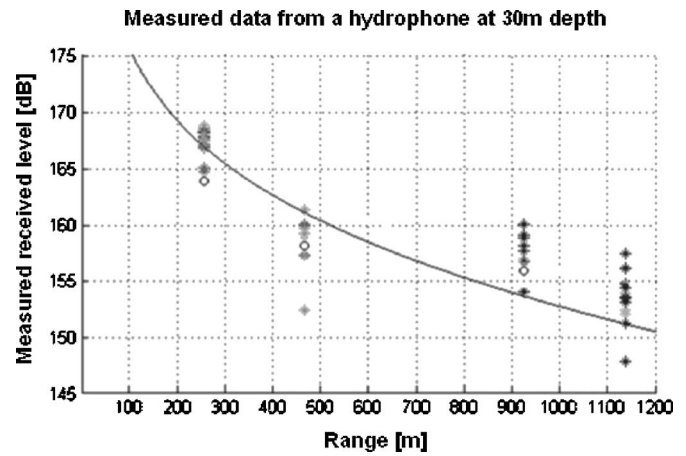


FIG. 6. MAST 04 measured received forward propagating field from a single hydrophone at 35 m depth and the ranges of 247, 467, 926, and 1140 m marked with different shade stars for different pings. OASES modeled received levels for these ranges and 30 m depth are marked with gray circles and are in good agreement with the measured values.

ceived level is marked with different shade stars for different sets of pings. A variation in the received level about 4–7 dB between different pings can be observed. The line falling off as a function of range represents the expected received level based on the spherical spreading and attenuation effects only. OASES modeled received levels for the hydrophone ranges and 30 m depth are marked on the figure with gray circles. Even though some simplifications in modeling, such as the bathymetry and range dependence of the sound velocity profile were employed, the measured incident values and the modeled values at this depth agree well, giving confidence in the accuracy of the modeled result, and rendering it usable for the transmission loss values. The correct estimation and measurement of the transmission loss is important to the target strength determination, as much of the scatter and diversity of the existing historical target strength data is attributed to the erroneous transmission loss assumption, often assumed to be only due to spherical spreading without correcting for the waveguide interaction and sound interference effects whenever appropriate.

2. Whale aspect dependence

Like other, more conventional scattering objects, such as cylinders or spheres, whale scattered response is aspect dependent. Here, the convention is accepted where the 0° azimuthal incidence corresponds to head on, 90° incidence is broadside, and 180° incidence corresponds to tail on. In order to determine the azimuthal orientation of the whales with respect to the incident field, MAST 04 data have been transformed and shown in the range-range space in Fig. 7. Figure 7 is basically the top view of a snapshot of a movie showing the range-range transformation of the beamformed, match-filtered received field as a function of time, marked with the ping number on the top of the figure. The whales are shown here at 370 m range X and 150 m range Y , as they pass by the system at 0 m range X and 0 m range Y , marked by the icon in the lower left corner. Based on a series of such snapshots, the whale orientation with respect to the sonar location was determined incrementally for each period of two con-

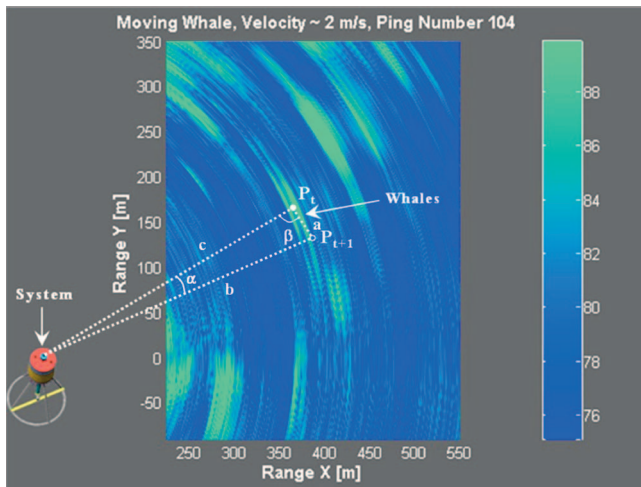


FIG. 7. A snapshot of a movie showing the range-range transformation of the beam-formed, match-filtered received field as a function of time, marked with the ping number on the top. The whales are shown here at the position P_t as they pass by the system. After two consecutive pings the whale position changes to P_{t+1} , based on which the whales azimuthal orientation β , with respect to the incident field, can be determined.

secutive pings. For each two ping segment the whale starting position P_t and ending position P_{t+1} are known. If an assumption is made, that for each one of these time periods of 10 s the whale moved in a straight line, which is not an unreasonable assumption, than the whales' azimuthal orientation β with respect to the direction of the incident field can be determined for each time interval, based on the known ranges at P_t , marked by a letter c in Fig. 7 and at P_{t+1} marked by a letter b , and the azimuthal angle α between them:

$$a = \sqrt{b^2 + c^2 - 2bc \cos \alpha}, \quad \beta = \arcsin \left[\frac{b}{a} \sin \alpha \right].$$

As mentioned previously, it has been established that gray whales in a migratory pattern spend over 90% of their time diving at depths 20–35 m below the surface. Therefore, in agreement with these established whale depths, the transmission loss in a waveguide was taken at 35 m depth. In addition, as shown by Fig. 6, the depth-range area within the 400–900 range and 0–35 m depth region, where the incident field deviates by more than 3 dB from the one established at 35 m depth is 11% of the overall region. Multiplying that 11% region by the 10% time interval that gray whales spend at those depths, gives a probability of 1.1% that the whales were exposed to the incident field that varied by more than 3 dB from the level at 35 m depth at ranges 400–900 m. Therefore, taking the values of the incident field at 35 m depth in the water column are correct 98.9% of the time within 3 dB. It should also be noted that the previously mentioned publications do not provide the water-column incident field, or consider the whale depth for the duration of their measurements. In contrast to that, in addition to providing waveguide propagation corrections and therefore increasing the accuracy of the incident field, we provide the above time-position error analysis establishing the validity of our method. A possible determination of whale locations in range, bearing, and depth can potentially be achieved in the future by

TABLE I. Target strength measurements of gray whales.

f (kHz)	Avg range (m)	Aspect angle β (deg)	Avg target strength (dB)	σ of target strength (dB)
23	746	175–173	-2.9	± 0.6
23	612	161–147	3.2	± 1.2
23	507	145–142	7.8	± 2.1
23	580	140–134	4.8	± 3.6
23	511	127–123	10.4	± 0.6
23	468	117–103	10.9	± 1.7
23	410	97–93	12.7	± 0.7
23	405	92–88	12.8	± 0.3

deploying an additional vertical line array in the water column and applying matched field processing¹⁵ in the vertical, in addition to beamforming in the horizontal direction. Consistent with the characteristics of whale motion they are assumed to travel head on from point P_t to point P_{t+1} for the 10 s time interval, as shown in Fig. 7. In addition to the whale backscattered response marked in Fig. 7, other surrounding high level returns are attributed to the surface, bottom, and water-column clutter being backscattered toward the receiver. For clutter rejection purposes, the origin, the characteristics, and the behavior of the surrounding clutter are considered of importance. Here, we note that some of the nontarget high level returns shown in Fig. 7 that are stationary are attributed to bottom clutter such as rocks and seabed reverberation, while the other nontarget high level returns are due to kelp and sea-surface reflections. The seemingly semicircular shape of these returns is due to the XY projection of the original range-bearing space in which the returns have a more omnidirectional shape.

3. Whale target strengths

Table I shows the measured, ping averaged gray whale target strengths and the corresponding frequency, average range, aspect angle β , and the related standard deviation of the measured target strength σ . Due to the ping rate of 5 s, the number of pings available was not high, and therefore the averages were made based on 4 to 6 pings. Nevertheless, as expected, the results show low target strengths of -2.9 dB at orientations close to the tail on, i.e., $175^\circ - 173^\circ$, and as the orientation changes the target strength goes up to +12.8 dB, at the broadside incidence of $88^\circ - 97^\circ$. Most of the standard deviations of target strengths were around 1 dB, which is small compared to the target strength standard deviations that other authors have reported. While the broadside target strength is somewhat higher here, the trend of these results compares favorably with the target strengths of humpback whales measured by Love² at similar frequencies to ours. They measured the target strength of +8 dB around the broadside aspect followed by the decrease of up to -4 dB for head on orientation. A year after Love published his results, Levenson⁴ measured the bistatic target strength of sperm whales at similar frequencies, obtaining the maximum target strength of +14.4 dB, a result that is also compatible to ours. All of the above-mentioned whale target strength measure-

TABLE II. Intermediate values for gray whale target strength calculation.

Range (m)	Bearing (deg)	Aspect angle β (deg)	Received level (dB)	Transmission loss (dB)	Target strength (dB)
396	342	91	117.5	54.5	12.5
462	12	145	106.6	58.8	10.2
577	27	134	101.8	59.4	6.6
668	33	158	102.0	57.2	2.4

ment values are higher than the previously mentioned Miller's measurements which assign right whale target strength values from -4 to -8 dB. Miller has offered a possible explanation for the difference based on the right whale internal body composition, but it should also be noted that Miller's measurements were taken at significantly higher frequencies than the rest, confirming the fact that the target strength of whales, like other man- or non-man-made objects, is frequency dependent.

To allow reconstruction of the above shown results, a condensed Table II has been added which lists the intermediate values of the range, bearing, aspect angle, received level, modeled transmission loss at 35 m depth, and the resulting TS for several representative pings from our data.

Figure 8 represents a polar plot of the measured target strength data as a function of aspect. It is, in fact, a scattering beam-pattern of a whale, where the target strength values are given by the magnitude of the polar plot and the angles correspond to the whale azimuthal orientation. Whales, like any other object, have a nonuniform directional response. At and around broadside incidence the response forms a main lobe of high backscattered response, it goes down with the increasing angle, turning into a sidelobe at 120° , and eventually falling off to significantly lower values at the tail-on direction. As far as the authors are aware, this is the first time in literature that an aspect-dependent gray whale target strength has been obtained.

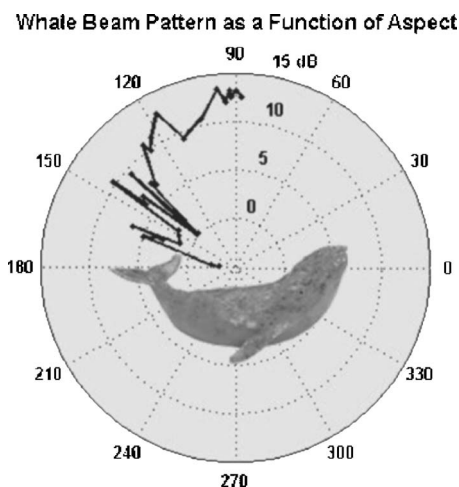


FIG. 8. A scattering beam-pattern of a gray whale as a function of azimuth, consisting of a main lobe around broadside incidence, a sidelobe at 120° , and low response at the tail-on incidence.

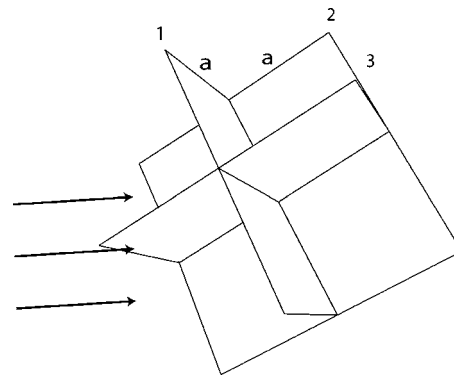


FIG. 9. A tetrahedral corner reflector of 1 m diameter moored in the MAST 04 water column and used for the measurements comparison with an object of known target strength.

4. Corner reflector target strengths

To confirm the validity of the implemented procedure independently, and therefore of the obtained whale target strength results, a comparison of the above-described target strength measurements of objects with the well-defined, calibrated target strength is provided here. A tetrahedral corner reflector with a 1 m diameter was moored in the water column during the MAST 04 experiment at the depth of 36 m and a range of 864 m. As show in Fig. 9, the corner reflector was moored in the water column by its top three corners marked with 1, 2, and 3 in order to control the exposed area of the target. Due to the fact that the target could rotate about its vertical axis it produced backscattered response of variable magnitude from ping to ping depending on the orientation of the corner reflector.

Here, a comparison of the maximum and the minimum measured corner reflector target strengths with the already established maximum and minimum target strengths of the object in the same configuration is given. During the MAST 04 experiment the target strength of the corner reflector backscattered response was measured and the target strength established using exactly the same method as the one used for the whales described in the previous sections. This air-backed corner reflector was manufactured by Scientific Solution Inc. and its target strength, in this particular configuration and these frequencies, was previously established in Ref. 16 by a calibration measurement conducted in BAE Systems Inc., quarry in Nashua, NH. Table III shows the high target strength corresponding to the highest of the 60 pings of 5 s ping rate and the lowest of the same 60 pings for both the MAST 04 measured target strengths, and the already known corner reflector target strengths. The MAST 04 measured high and low target strengths are recorded as $+7.7$ and -3.5 dB while the calibrated target strengths are $+7$ and -2 dB. The target strengths compare well producing an

TABLE III. Corner reflector target strength measurements.

	MAST 04 measured TS (dB)	Known TS (dB)
High	$+7.7$	$+7$
Low	-3.5	-2

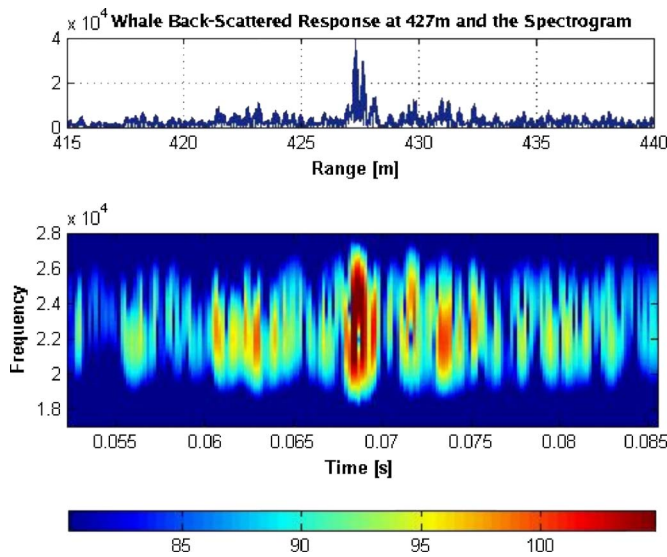


FIG. 10. (Color online) Whale backscattered response at 427 m, and its spectrogram below. Whale structural response appears to high-pass filter the scattered field with respect to the surrounding clutter.

agreement within 0.7 dB at the high end and 1.5 dB difference at the low end confirming the validity of the above-implemented procedure independently, and establishing a confidence measure of the obtained whale target strength results within a ± 1.5 dB margin.

D. Frequency content

In order to capture physical attributes that can help discriminate between target and clutter returns, we briefly exploit spectral and temporal variabilities of the whale backscattered response in the time-frequency domain.

Figure 10 shows the beamformed match-filtered whale backscattered response at 427 m, with the spectrogram of the whale backscattered response shown below. The spectrogram was obtained from the raw time-series data by applying a short time Fourier transform¹⁷ using a Hanning window of 50 samples size, and a 25 sample window overlap. As expected, at far off ranges, the whale response is surrounded by clutter. An interesting characteristic that has been observed is that the maximum value of the whale response on the spectrogram comes back high at 25 kHz even though the strength of the transmitted incident field was uniform across the frequency band of 21–25 kHz. Furthermore, the clutter returns surrounding the whale return at 427 m, arrive at the lower frequency of 23 kHz corresponding to the center frequency of the transmitted signal. This result is repeatable for other pings, indicating the possibility that the whale structural response high-pass filters the backscattered field. It is evident that due to the propagation losses in the waveguide combined with the high levels of clutter, robust tracking of whales using an active sonar system based only on the strength of the backscattered is range limited. For broadband systems, frequency content clues such as this one have a potential of providing an additional classification input, making the identification of whales possible at further ranges. While it is believed that this shift in frequency occurs due to the interaction of the incident field with the internal fluid-

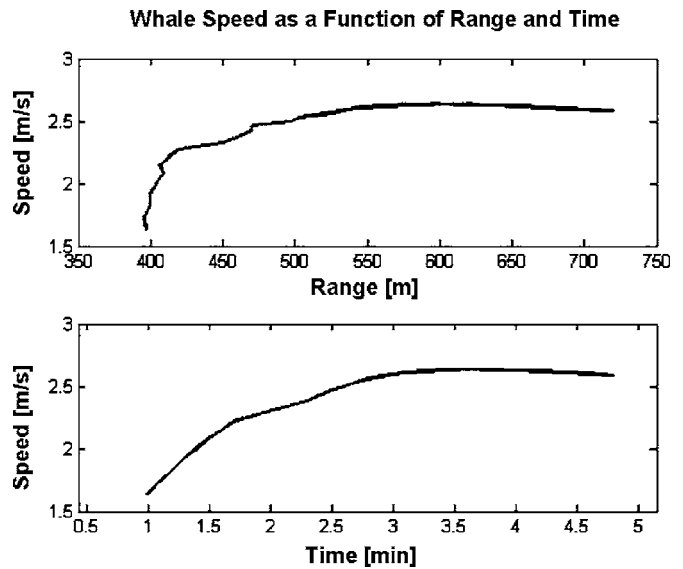


FIG. 11. Whale speed as a function of range on the top, and as a function of time on the bottom. The speed varied from 1.6 m/s at closer ranges higher to 2.6 m/s at ranges further away.

solid structure of the whale, this phenomenon will be further investigated in future work. However, it is worth noting that for man-made objects frequency content results such as those have been shown by Lucifredi,¹⁸ demonstrate wave-based physical mechanisms taking place, providing in return much needed target classification clues.

E. Whale motion

In addition to the frequency content of the whale response, another important aspect in active sonar detection and classification of whales is their behavior and their speed. Based on the whale position determined using the active sonar returns at every two ping interval described in Sec. III C 2, the incremental whale speed was obtained as they travel along their migration trajectory. Figure 11 is the plot of the whale speed as a function of range in the top part of the figure, and as a function of time on the bottom. The top part of Fig. 11 is useful in analyzing the whale speed with respect to how close the whales got to the system, while the slope of the curve on the bottom represents their acceleration or deceleration and can also be related to their proximity to the active sonar system. The whale speed varied from 1.6 m/s at closer ranges and increased to 2.6 m/s at ranges further away. However, the whale acceleration seems to follow the opposite trend, where it is higher at closer ranges where the whales moved at slower speeds, and then falls off for ranges further away where the whales traveled at higher speeds. This behavior has some similarities to the behavior that Love² observed during his experiment, where the whales approached the transmit tower, slowed down or even stopped listening for a while, and then moved on with an increased speed.

IV. CONCLUSIONS

With a goal of identifying and analyzing whale signature features necessary for enhanced active sonar detection and

classification of marine mammals, this paper presents the data collected during the MAST 04 experiment and conducts the analysis which transforms the collected data onto appropriate projection spaces where target and clutter characteristics can be better captured, and are less sensitive to noise and environmental perturbations. For the first time in literature aspect-dependent gray whale target strength measurements are obtained and the results are compared to the previously published whale target strengths (Table I). Gray whale target strengths of +12.8 dB and higher at the broadside orientation, and -2.9 dB at whale orientations close to the tail on, were measured. These results compare favorably with the target strengths of humpback whales measured by Love at similar frequencies, and to the target strengths of sperm whales measured by Levenson. In addition to providing waveguide propagation corrections and therefore increasing the accuracy of the incident field, we provide the time-position error analysis establishing the validity of our method. In addition to the target strength, other whale feature identification points suitable for automated detection and classification were investigated, as means of relating gray whale active acoustic signatures to their inherent characteristics and motion. Temporal and spectral physical attributes, such as whale target extent of 12.5 m and high-pass filtered frequency content were determined, as well as whale speed of 1.6–2.6 m/s, and some aspects of the observed whale behavior. In addition, the maximum detection range of roughly 1 km, and the resolving capabilities of the active sonar system operating in a high clutter environment were demonstrated. Concerning the active sonar system, tracking and classification algorithms are being currently enhanced and they will be implemented and tested during the upcoming experiment featuring an autonomous system, implementing real-time marine mammal detection, tracking, and classification.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

- ¹J. L. Dunn, "Airborne measurements of the acoustic characteristics of a sperm whale," *J. Acoust. Soc. Am.* **46**, 1052–1054 (1969).
- ²R. H. Love, "Target strengths of humpback whales *megaptera novaeangliae*," *J. Acoust. Soc. Am.* **54**, 1312–1315 (1973).
- ³J. H. Miller and D. C. Potter, "Active high frequency phased-array sonar for whale ship-strike avoidance: Target strength measurements," in *Proceedings of the IEEE/MTS Oceans 2001*, Honolulu, pp. 2104–2107.
- ⁴C. Levenson, "Source level and bistatic target strength of the sperm whale (*physeter catodon*) measured from an oceanographic aircraft," *J. Acoust. Soc. Am.* **55**, 1100–1103 (1974).
- ⁵R. O. Nielsen, *Sonar Signal Processing* (Artech House, Boston, 1991).
- ⁶H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part III* (Wiley, New York, 2001).
- ⁷J. G. Proakis and D. G. Manolakis, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1996).
- ⁸A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1977).
- ⁹H. L. Van Trees, *Optimum Array Processing* (Wiley, New York, 2002).
- ¹⁰R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1993).
- ¹¹H. Schmidt and J. Glattetre, "A fast field model for three-dimensional wave propagation in stratified environments based on the global matrix method," *J. Acoust. Soc. Am.* **78**, 2105–2114 (1985).
- ¹²F. Jensen *et al.*, *Computational Ocean Acoustics* (Springer, New York, 2000).
- ¹³C. D. Malcom and D. A. Duffus, "Comparison of subjective and statistical methods of dive classification using data from a time-depth recorder attached to a gray whale (*eschrictius robustus*)," *J. Cetacean Res. Manage.* **2**, 177–182 (2000).
- ¹⁴A. A. Wolman, "Gray whale *Eschrictius robustus*," in *Sirenians and the Baleen Whales, Handbook of Marine Mammals Vol. 3*, edited by S. H. Ridgway and R. Harrison (Academic, London, 1985).
- ¹⁵A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," *J. Acoust. Soc. Am.* **83**, 571–587 (1988).
- ¹⁶P. Stein, *Technical Report: Corner Reflector Target Strength Measurements* (Scientific Solutions Inc., Nashua, NH, 2002).
- ¹⁷V. C. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis* (Artech House, Boston, 2002).
- ¹⁸I. Lucifredi and H. Schmidt, "Sub-critical scattering from buried elastic shells," *J. Acoust. Soc. Am.* **120**, 3566–3583 (2006).

Miniaturized mandrel-based fiber optic hydrophone

Gregory H. Ames and Jason M. Maguire

Naval Undersea Warfare Center, Division Newport, 1176 Howell Street, Newport, Rhode Island, 02841

(Received 30 June 2006; revised 3 November 2006; accepted 11 December 2006)

Fiber optic hydrophones typically use a long sensor fiber wrapped around an air-filled mandrel to achieve high responsivity. Minimum mandrel diameter is determined by fiber reliability considerations. A miniature hydrophone, with diameter of about a third that of previous hydrophones, is reported. Such a hydrophone has applications for very thin towed arrays. Small diameter fiber with high strength Bragg gratings written through the coating are key enablers.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2431340]

PACS number(s): 43.30.Yj, 43.38.Ar [RAS]

Pages: 1392–1395

I. INTRODUCTION

Fiber optic hydrophones have been developed for a variety of sonar applications.^{1,2} Typically, these interferometric sensors achieve high responsivity by winding significant lengths of optical sensing fiber around a compliant mandrel, usually a sealed air-filled metal or plastic tube.³ Typical mandrel diameters are 10–23 mm in diameter.^{2,3} In some applications small size is not particularly important and mandrel diameter may be chosen to maximize the storage of sensing fiber or to achieve desired mandrel compliance. (The compliance of the mandrel is proportional to the ratio of mandrel diameter to mandrel wall thickness.) In applications where small diameter is desired, sensor diameters are limited by the long term reliability of the fiber wrapped on the mandrel. Fiber wrapped around a small diameter mandrel sees strain on the outside of the fiber bend. It is well known that long term strain on optical fibers leads eventually to fiber failure.^{4,5} The total fiber stress is given by the following:

$$\sigma(\theta) = E \left[\left(\frac{d}{D \cos \phi} \right) \cos \theta + \varepsilon_{\text{winding}} \right], \quad (1)$$

where E is the glass's Young's modulus, d is the fiber glass diameter, and D is the mandrel diameter. ϕ is the winding pitch angle and θ is the angle around the circumference of the fiber, with 0 deg at the outside of the bend. These two angles are illustrated in Fig. 1. $\varepsilon_{\text{winding}}$ is the longitudinal strain on the fiber from the winding tension or from changes in mandrel diameter due to pressure or temperature. For close packed winding, where the mandrel diameter is much greater than the fiber diameter, it is apparent that the first term (bending stress) in Eq. (1) is approximately proportional to the ratio d/D . Therefore, one method of achieving a reduction in mandrel diameter without sacrificing fiber lifetime is to reduce the fiber diameter by the same proportion.

The development and testing of a miniature fiber optic hydrophone using a 3.8-mm mandrel will be described. This design was based on using a new fiber with a 50- μm glass outer diameter and a single 110- μm outer diameter acrylate coating.

II. HYDROPHONE DESIGN

A. Fiber

A custom fiber was developed for this project by Fiber-core Ltd. The fiber had a 50- μm outer diameter glass core and 110- μm outer diameter acrylate coating with a single layer. The fiber had a numerical aperture of 0.29 to enable low bend loss at the small bending diameters to be used on the hydrophone. Despite this high numerical aperture, low intrinsic loss of 0.7 dB/km was achieved. The fiber was proof tested to 3%.

To test the fiber bend loss, 90 m of fiber were wound close packed on a 3.81-mm outer diameter mandrel. The fiber was interrogated with an optical time domain reflectometer both before and after winding. The difference indicates a 0.015-dB bending loss for the 90 m. This loss is acceptable for hydrophone arrays that may have a kilometer of sensor fiber.

B. Grating based, path-matched, differential interferometer

The path-matched differential interferometer is a common form of fiber optic interferometric sensor.⁶ Figure 2 shows a particularly advantageous form of this type of sensor array. The reflective Fabry-Pérot form of the sensor interferometers is achieved using matched-wavelength Bragg grating pairs and the fiber between them as individual sensors. The arrangement offers the simplicity of many sensors located serially on a single fiber. The sensor array is interrogated by pulsed lasers. The sensor provides two nonoverlapped pulse returns. By passing them through a compensating interferometer with a path mismatch equal to the two-way sensor length, the pulses are realigned and interference occurs. To implement this sensor configuration on a small diameter mandrel it is necessary to wind Bragg gratings at a small diameter. Arrays of gratings were purchased that had been manufactured in the 50- μm fiber by writing the gratings through the coating, without removing the coating.⁷ This method provides gratings that have nearly the same strength as the original fiber. Pairs of gratings were used that reflect at matched wavelength. The gratings had 1% reflection and a 1-nm bandwidth, with 40 m of sensor fiber between them. The gratings were purchased with zero ten-

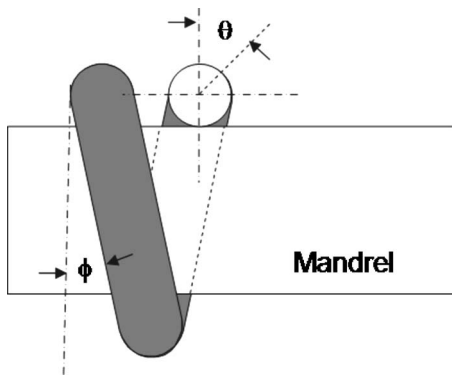


FIG. 1. Mandrel wrapping geometry showing key angles in Eq. (1).

sion reflection wavelength lower than the interrogation laser wavelength and the gratings were tensioned when wound to achieve a reflection wavelength matching the laser. There was some concern that bend-induced birefringence in the fiber would create polarization dependent shifting of the reflection bandshape, but this was not observed and grating winding and spectral adjustment was sufficient to achieve alignment to the laser.

C. Hydrophone responsivity calculation

Simple analytic calculations of the responsivity of a closed hollow cylinder are insufficient because they do not account for the optical fiber, which considerably stiffens the mandrel structure. For this reason, finite element analysis (FEA) was used to evaluate the responsivity and survival pressure of the hydrophones. The FEA model was three dimensional with solid elements representing the mandrel and Acrylate fiber coating, while beam elements represented the fiber glass. Different versions with one, two, and three fiber layers were analyzed. Figure 3 shows the model for the one layer version. All versions modeled the fiber winding as close packed. Two end conditions are studied; one with the ends free to move axially and the other with fixed ends which simulates an infinitely stiff cage attached to the mandrel ends. In both end conditions the nodes at each end of the mandrel cannot move in the radial direction. A steady state harmonic frequency sweep was performed with a $1\text{-}\mu\text{Pa}$ pressure load and the resulting responsivity is shown in Fig. 4 for the three fiber layer case. This exhibits a low frequency responsivity of -166.9 or -168.8 dB//rad/ $\mu\text{Pa m}$ for either fixed or free end conditions. The same was done for the one fiber layer case, yielding low frequency responsivity of -149.5 and -151.7 dB//rad/ $\mu\text{Pa m}$ for fixed or free end

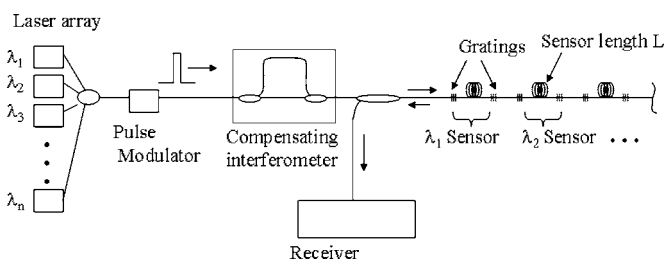


FIG. 2. Differential or two-part interferometer based on TDM/WDM Bragg grating based interferometric sensors.

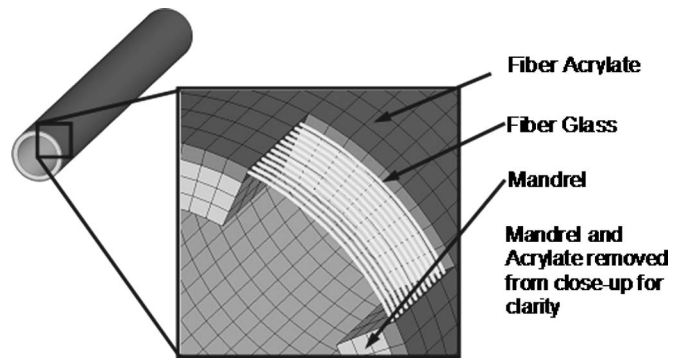


FIG. 3. Diagram of finite element analysis model showing mandrel, fiber coating, and fiber glass elements.

conditions. The first hoop resonance is barely visible near 7050 Hz due to the low resolution of the sweep and the lack of damping in the model. To determine if there were any resonances of the structure not excited by the pressure load an eigenvalue extraction analysis was performed and the resonant frequencies of the mandrel are listed in Tables I and II.

D. Hydrophone construction

The hydrophone mandrels were formed from polycarbonate extruded tubing with an outer diameter of 3.8 mm and a wall thickness of 0.29 mm. The bending strain in the fiber at such a small diameter is not suitable for systems that require extended 20-year life. However, it was deemed sufficient for a short term demonstration, with greater reliability being possible in the future if hermetic coatings on the fiber compatible with grating writing can be developed. If these are not available, smaller diameter fiber or larger mandrels could be used to achieve necessary lifetime. The mandrels were cut to length and then sealed at the ends. The hydrophone design was for an extended hydrophone group where the sensing fiber between the gratings is distributed across four mandrels. The total hydrophone group length was 190.5 mm. A diagram of the hydrophone group with dimensions is shown in Fig. 5. Flexible interlinks⁸ were used to connect the mandrels and provide a path to wind the fiber from mandrel to mandrel. The fiber was wet wound onto each mandrel in three layers with an ultraviolet-cured adhesive. Perforated metal cage halves were installed over the mandrels to protect the fiber. The completed hydrophone is shown in Fig. 6.

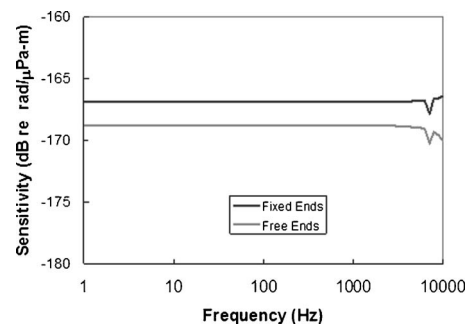


FIG. 4. Responsivity vs frequency for FEA model with three fiber layers.

TABLE I. Resonant frequencies vs mode and end conditions for a mandrel with one fiber layer.

Mode	Resonant frequency (kHz)	
	Free ends	Fixed ends
Bend 1	3.808	3.808
Axial	8.861	...
Bend 2	9.072	9.072
Hoop	9.645	9.644
Bend 3	15.488	15.488

III. HYDROPHONE TESTING

Hydrophone operational testing and calibration was carried out in a closed, water filled pressure vessel. A comparison calibration method was used with a pressure compensated projector delivering a swept sine signal. The hydrophone under test and a single reference hydrophone were rigged adjacent to each other at the bottom of the tank. The calibration tank operates in a standing wave mode.

The hydrophone was interrogated by an interferometric system such as is shown in Fig. 2, but with only one laser and one sensor. Pulses of 100 ns were formed by a lithium niobate modulator. A heterodyne interrogation method was used, whereby a high frequency carrier was introduced by frequency shifting the light in one arm of the Mach-Zehnder compensator using an acousto-optic modulator.

Figure 7 shows the calibration for the miniature optical hydrophone over the range from 50 to 450 Hz. The frequency range was limited by the test tank characteristics and not the fiber optic hydrophone. It can be seen that the hydrophone exhibits a typically flat response across the band, with a responsivity or scale factor of $-138 \text{ dB//rad/}\mu\text{Pa}$ and a responsivity per unit physical length of fiber of $-170 \text{ dB//rad/}\mu\text{Pa m}$. This compares to the modeled responsivity of $-168.8 \text{ dB//rad/}\mu\text{Pa m}$ for the free end conditions or $-166.9 \text{ dB//rad/}\mu\text{Pa m}$ for the fixed end conditions. The actual end conditions are in between the two extremes due to the specifics of the cage rigidity and its' attachment to the mandrels. The slight difference in responsivity between the data and model may be due to some ellipticity and wall thickness variation in the mandrels. Given typical interferometric system phase detection thresholds of $1-10 \mu\text{rad/Hz}^{1/2}$, this represents a sensitivity of $18-38 \text{ dB//}\mu\text{Pa/Hz}^{1/2}$. This responsivity is significantly

TABLE II. Resonant frequencies vs mode and end conditions for a mandrel with three fiber layers.

Mode	Resonant frequency (kHz)	
	Free ends	Fixed ends
Bend 1	2.998	2.998
Hoop 1	7.050	7.050
Bend 2	7.141	7.141
Bend 3	12.184	12.184
Hoop 2	13.984	13.984

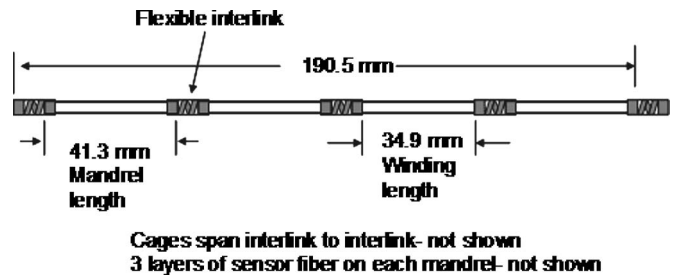


FIG. 5. Diagram of miniature hydrophone group with dimensions.

less than can be achieved with longer, single layer hydrophones. This reduction in responsivity can be attributed to the reduction in mandrel compliance due to the stiffening of the mandrel from multiple layers of glass fiber. Once the mandrel compliance is dominated by the glass fiber rather than by the polycarbonate, there is little advantage from using longer sensor fiber. This demonstrates a fundamental tradeoff with this technology. For higher frequency arrays, shorter groups are required. This leads to a lack of total mandrel length for storing sensing fiber which impacts the achievable responsivity. This limitation is not present for long, low-frequency hydrophone groups where there is sufficient total mandrel length to wind the sensor fiber without impacting the mandrel compliance. Other factors affecting the sensitivity are the amount of sensor fiber and the survival depth requirement on the sensor. The former can vary from system to system based on optical architecture considerations. The later depends on the requirement of the application. Sensor performance may be improved in the future by either a smaller diameter fiber or a smaller diameter fiber coating. A smaller outer coating allows tighter winding and more sensor fiber in the available total mandrel space. A smaller diameter fiber has the same effect, but also reduces the responsivity reduction encountered using multiple fiber layers.

IV. CONCLUSIONS

A first demonstration of a miniature fiber optic hydrophone has been presented that achieves hydrophone diameter of a third of previous designs. Calibration measurements show flat spectral response typical of fiber optic hydrophones. The responsivity or scale factor of the hydrophone is shown to depend on the design frequency (and hence, length) of the extended hydrophone group.

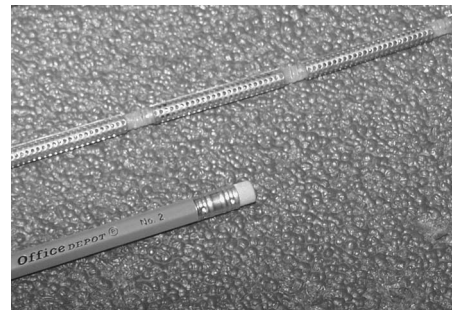


FIG. 6. Photo of miniature hydrophone group showing flexible interlinks and perforated metal cages.

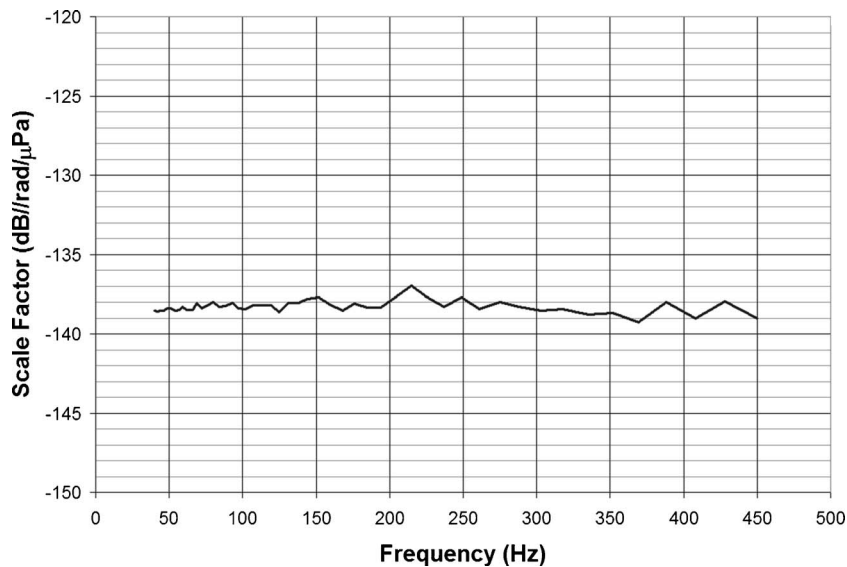


FIG. 7. Calibration results giving scale factor or responsivity of the hydrophone group.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research, ONR321MS, Dr. James McEachern.

¹A. Dandridge, "Fiber-optic sensors for acoustic array applications," *J. Acoust. Soc. Am.* **102**, 3125 (1997).

²G. A. Cranch, P. J. Nash, and C. K. Kirkendall, "Large-scale remotely interrogated arrays of fiber-optic interferometric sensors for underwater acoustic applications," *IEEE Sens. J.* **3**, 19–30 (2003).

³P. J. Nash and J. Keen, "Design and construction of practical optical fiber hydrophones," *Proc. Inst. Acoust.* **12**, 201–212 (1990).

⁴R. Olshansky and R. D. Maurer, "Tensile strength and fatigue of optical fibers," *J. Appl. Phys.* **47**, 4497 (1976).

⁵D. Kalish and B. K. Tariyal, "Probability of static fatigue failure in optical fibers," *Appl. Phys. Lett.* **28**, 721 (1976).

⁶S. A. Al-Chalabi, B. Culshaw, and D. Davies, "Partially coherent sources in interferometric sensors," *Proc. First Inter. Conf. on Optical Fiber Sensors*, London (1983), pp. 132–135.

⁷D. S. Starodubov, V. Grubsky, and J. Feinberg, "Efficient Bragg grating fabrication in a fibre through its polymer jacket using near-UV light," *Electron. Lett.* **33**, 1331–1333 (1997).

⁸G. H. Ames, "Interlink for spatially extended hydrophones," US Patent No. 6,118,733 (12 September 2000).

Aberrations in materials with random inhomogeneities

J. A. Hernández,^{a)} M. Clark, S. D. Sharples, and M. G. Somekh
School of Elec. and Elec. Engineering, The University of Nottingham, NG7 2RD, United Kingdom

V. H. Lopez
Instituto de Investigaciones Metalúrgicas, UMSNH, Morelia, Mich., México

(Received 20 August 2005; revised 13 December 2006; accepted 13 December 2006)

Materials that consist of a random microstructure can affect ultrasonic measurements—reducing signal strength, increasing noise, and reducing measurement accuracy—through scattering and aberration of the acoustic field. To account for these adverse effects a phase screen model, alongside the stochastic wave equation, has been developed. This approach allows the field and study aberrations to be modeled from a statistical point of view. Experimental evidence of aberration and statistical properties of the measured acoustic field are shown. A measured correlation function of the acoustic field is interlinked to mean crystallite size by using a theoretical coherence function that can be mainly described by the correlation length and wave velocity variation of microstructure. The estimation of the mean crystallite size using this technique would provide some insight into material characterization. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431582]

PACS number(s): 43.35.Cg, 43.20.Fn, 43.28.Lv [RLW]

Pages: 1396–1405

I. INTRODUCTION

It is well established that the properties of materials are dictated, to a great extent, by their crystallite structure. Crystal refining is a popular practice to enhance or manipulate the properties of materials. For example, in aluminum, a fine equiaxed microstructure leads to benefits such as less shrinkage, reduced porosity, uniform distribution of second phases and obviously improved mechanical properties which are uniform throughout the material. Moreover, it facilitates subsequent processing. To characterize the crystallite structure nondestructive methods are valuable tools to assess the quality of materials and predict mechanical behavior. In this paper the effects of wave-front distortions on ultrasonic surface acoustic wave (SAW) propagation produced by crystallites of polycrystalline materials is studied and used as a means of gathering information about the microstructure itself. The technique is at an early stage and does not provide a means of estimating crystallite size distribution but it does give a measure of mean crystallite size since this can be estimated based on the coherence of the acoustic field. These effects, generally speaking, will be termed ultrasonic aberrations.

There exist in the literature several techniques for measuring the crystallite size. The microstructure itself is assumed to have certain statistical properties, from which it is theoretically possible to calculate the acoustic first moment and in consequence an attenuation coefficient which is proportional to mean crystallite size.^{1,2} The mean crystallite size is directly related to a correlation length that defines a random medium that idealizes real microstructure. The energy correlation of the acoustic field is dependent on the correlation length of the medium. So it is indirectly possible to obtain mean crystallite size. A good deal of theoretical and

experimental work related to second-order moments and crystallite-noise scattering of acoustic field has been reported.^{3–5}

The paper is organized as follows. Section II looks at some aspects of ultrasonic propagation in materials. The theory is based on the stochastic wave equation for random media alongside a phase screen model. This approach allows us to approximate the field in a random medium, therefore propagation of the energy correlation function developed in Sec. III, which is later used for obtaining microstructure characteristics. A set of experiments was performed on three different samples of aluminum, each with a different crystallite size. This experimental work is presented in Sec. IV which describes the experimental procedure (including instrumentation) for collection of the ultrasonic information necessary for statistical analysis of the measured aberrations. In Sec. V the analysis is explained using an estimated transverse energy correlation function of the field. We have also included in this section simulation of acoustic propagation in a random medium to further support the analysis. This measured correlation function is compared in the last section to the modeled function outlined at the beginning of the paper, so that we can check its consistency. It is believed that the technique will be useful for material characterization as it is possible to relate the aberration strength to the mean crystallite size of the material.

II. THEORY

There exist several theoretical approximations of acoustic wave propagation, for both SAW or bulk waves in polycrystalline materials. Some authors have tried to reduce the problem to a scalar description,⁶ using integral techniques or even more complicated methods where the first moment is calculated for certain types of linearly elastic solids.⁷ In this paper the theoretical description of ultrasound in an inhomogeneous medium is based on the stochastic wave equation in

^{a)}Electronic mail: eexjah@nottingham.ac.uk

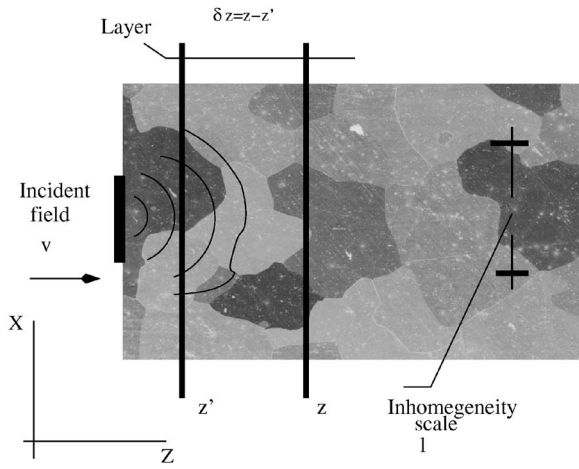


FIG. 1. Schematic representation of ultrasonic propagation in a random medium using phase screen theory.

two dimensions, Eq. (2.1), and a phase screen model widely used in optics and atmospheric calculations.^{8,9} Surface acoustic waves in homogeneous materials can be described fairly well by Eq. (2.1).^{10,11} It will be seen that this description can be approximately extended to a heterogeneous medium. For that purpose, the ultrasound is assumed to be monochromatic, that is $U(\mathbf{r}) = u(\mathbf{r})e^{-i\omega t}$ which then leads to Helmholtz's equation for u :

$$\begin{aligned} [\Delta + \bar{k}^2]u(\mathbf{r}) &\approx -\bar{k}^2\mu(\mathbf{r})u(\mathbf{r}), \\ u(x,0) &= v(x), \end{aligned} \quad (2.1)$$

where $k^2(\mathbf{r}) = \bar{k}^2 \left(\frac{\bar{c}}{c(\mathbf{r})}\right)^2 = \bar{k}^2(1 + \mu(\mathbf{r}))$, μ is a zero mean Gaussian process characterizing the inhomogeneity of the medium, and $\mathbf{r} = (x, z)$. The mean velocity \bar{c} here refers to the Rayleigh wave velocity in aluminum. The variance $\sigma^2 = \langle (k - \bar{k})^2 \rangle / \bar{k}^2$ measures the random fluctuation with respect to the mean wave number \bar{k} . The function $v(x)$ is the initial normal displacement, at the plane $z=0$ so Eq. (2.1) would give the normal displacement.

The calculations of first and second moments of Eq. (2.1) has been the subject of several papers. The second-order moments are related to the energy correlation as they coincide under stationary conditions. In this section an approximate representation for the acoustic field is given with the aim of calculating the energy correlation function; this function is similar to the one given in Refs. 12–14, for example. The propagation of v is performed by dividing the space into layers of thickness δz , as shown in Fig. 1, where only half-space is depicted. The acoustic field is then approximated within each layer using equation Eq. (2.1). The thickness δz is taken to be of the order of the correlation length of $k(\mathbf{r})$, although the suitability of this choice is not mathematically established in this article.

In order to quantify aberrations or wavefront distortions of the acoustic field, forward scattering only will be considered. This can be done by making the substitution $u = e^{ikz}e^{\Phi}$, where Φ is an unknown function to be determined. This leads to parabolic or paraxial approximations of Eq. (2.1) for Φ , which accounts only for forward scattering.^{15,16} The func-

tion Φ accounts for amplitude and phase variations of the field within a random medium. However, in this article, an alternative method is used to account for amplitude and phase variations of the field u based on a phase screen model. In the general case, Φ is a more complicated function and related to ϕ in Eq. (2.2) but is not discussed in this paper.

The use of a phase screen model to approximate solutions to Eq. (2.1) has long been established.^{17–19} The model is stated here with no theoretical analysis but is used to approximate the correlation function of the field. The phase screen model is directly related to solution of the parabolic version of Eq. (2.1) using continual path integrals;⁸ this is perhaps the formal justification of this model.

The phase screen model states that if the inhomogeneities do not significantly bend rays with respect to the axis of propagation z , the field within a layer can be approximated by the field in free space and a random screen. This approximation is only valid for weak inhomogeneities or for short path lengths. As a consequence of this approximation the amount of phase change ϕ that an incident field v to a layer of thickness δz would experience is given^{8,19,20} by

$$\phi(x) = \frac{\bar{k}}{2} \int_0^{\delta z} \mu(x, z') dz'. \quad (2.2)$$

Thus, the integral in Eq. (2.2) is adding all contributions in phase change that the field would experience along propagation within a layer of thickness δz . This is an approximation since in reality points belonging to wavefront will not follow straight lines, so the integral in Eq. (2.2) will have to be changed to a continual integral.⁸

Let us assume that the screen, which mathematically takes the form of a complex number as $e^{i\phi}$, is located at distance z , which will be the entrance of one layer, then the field behind the screen is given by

$$u(x, z) = v(x)e^{i\phi(x)}. \quad (2.3)$$

The next step is to propagate u in free space to the exit of the layer, which is from z to $z + \delta z$ and is given by Eq. (2.1) with $\mu=0$ and initial condition $u(x, z)$. It is well known that the solution to that problem is given by a plane wave expansion,²¹ which is presented below. It has been assumed, in writing Eq. (2.3) that entrance of the layer is located at an arbitrary point $z \neq 0$, thus the limits of integration in Eq. (2.2) should be from z to $z + \delta z$. For simplicity let us assume that the screen is located at $z=0$ then the initial field will be $u(x, 0) = v$.

Equation (2.3) indicates that the forward field, at the exit of the layer, is equivalent to the background field phase shifted by ϕ . If the screen is in the middle of the layer, v is propagated to $\frac{\delta z}{2}$ in free space, phase shifted by $e^{i\phi}$, Eq. (2.3) and then free space propagation is again applied up to δz . The propagation and the phase shifting is performed in the spatial frequency domain. So if $\hat{v}(p)$ denotes the Fourier transform of $v(x)$ with respect to x , the solution to Eq. (2.1) with $\mu=0$ is given by²¹

$$u_{\text{free}}\left(x, \frac{\delta z}{2}\right) = \int_{-\infty}^{\infty} \hat{v}(p) h\left(p, \frac{\delta z}{2}\right) e^{ipx} dp, \quad (2.4)$$

where $h(p, z) = \exp\left[iz\bar{k}\left(1 - \frac{p^2}{2k^2}\right)\right]$ and p is a variable denoting spatial frequency. We are temporally denoting the field u in free space by u_{free} to differentiate it from the one in Eq. (2.5) in a random layer. Taking the Fourier transform of Eq. (2.3) with respect to the variable x and using Eq. (2.4), and reverting back again to the spatial domain by performing the inverse transform gives the representation of u in a random layer as follows:

$$u = \int_{-\infty}^{\infty} \left[\hat{v}(p) h\left(p, \frac{\delta z}{2}\right) \otimes \hat{s}(p) \right] h\left(p, \frac{\delta z}{2}\right) e^{ipx} dp. \quad (2.5)$$

Here \otimes is the convolution operator of two functions and \hat{v} , \hat{s} denote the Fourier transform of v , $s = e^{i\phi}$, respectively. Notice, that u in (2.4) and (2.5) is written in the paraxial version by making the approximation $\sqrt{k^2 - p^2} \approx \bar{k}\left(1 - \frac{p^2}{2k^2}\right)$, which comes from an exact solution to Helmholtz equation in free space.²¹ This approximation is used with the calculations on the correlation function since integrals involving quadratic terms can be performed under suitable conditions. A special case of Eq. (2.5) is $\phi=0$, which results in Eq. (2.4) as expected. Equation (2.5) represents the ultrasound propagating through an inhomogeneous layer which has been approximated by distorting the phase of its elementary components by ϕ .

To extend the propagation to larger propagation distances than δz , Eq. (2.5) is applied recursively to an arbitrary number of layers. So, for instance, since δz is the thickness, u_n indicates the field at $z = n\delta z$ away from source, where $n = 1, \dots, N$, and N is the number of screens within 0 to $z = N\delta z$. Replacing u_n for v and using p_{n+1} instead of p as the dummy variable in expression (2.5) one can see that the field at the exit of the n th layer in the spatial frequency domain is given by

$$\begin{aligned} \hat{u}_{n+1} &= [\hat{u}_n(p_{n+1}) h_{p_{n+1}} \otimes \hat{s}(p_{n+1})] h_{p_{n+1}} \\ &= h_{p_{n+1}} \int_{-\infty}^{\infty} [\hat{u}_n(p_n) h_{p_n} \hat{s}(p_{n+1} - p_n)] dp_n, \end{aligned} \quad (2.6)$$

where $h_{p_n} = h\left(p_n, \frac{\delta}{2}\right)$ and the convolution operator has been replaced by its definition. Here, also $u_0 = v$ as a special case.

Back substitution of the recursive definition for u_n in expression (2.6) shows that the total field $u = u_{n+1}$ at an arbitrary point in space $\mathbf{r} = (x, z)$ can then be written as

$$\begin{aligned} u(x, z) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \hat{v}(p_0) \prod_{j=0}^{n-1} h_{p_j}^2 \hat{s}(p_{j+1} - p_j) \\ &\quad \times h_{p_n} e^{ip_n x} dp_0 \cdots dp_n. \end{aligned} \quad (2.7)$$

Equation (2.7) represents the ensemble of the acoustic field in a random medium and it will serve as a basis to calculate the energy correlation function. It is a multiple integral and there are as many integrals as there are screens within the slab, however, it is computationally efficient as these can be implemented using the fast Fourier transform (FFT) algo-

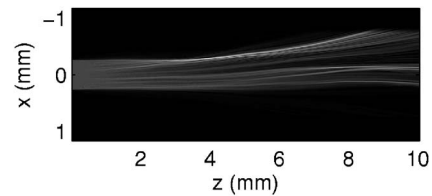


FIG. 2. Amplitude distribution according to Eq. (2.7). The simulation is based on real parameters and a numerical procedure. This can be compared with the experimental amplitude distribution shown in Fig. 6.

rihm. Equation (2.7) represents a SAW traveling in the forward direction in the half-space ($x, 0 \leq z$). The above procedure can be applied to propagate the initial displacement v in the other direction by using the symmetry of the Green's function for the Helmholtz equation. The acoustic field is statistically symmetric with respect to axis z , so only half of it will be considered. We have only measured half of the acoustic field in the experimental work, although it is possible to image SAW in both directions.

The amplitude of a realization numerically implemented from Eq. (2.7) is shown in Fig. 2. It is a truncated acoustic plane wave propagating in simulated inhomogeneous medium characterized by $\sigma=0.02$ according to Eqs. (2.1) and (2.7). As the wave travels from left to right (z direction) the phase is being altered by screens placed to simulate the aberration caused by real microstructure. The overall phase is distorted as well as the amplitude distribution breaking up as it can be seen in Fig. 2. This spread is a typical situation of acoustic surface waves in real experiments. This simulated acoustic field, however, is not expected to follow precisely all wave paths that one would expect to follow in real experiments, such as that shown in Fig. 6. Their similarities are assessed by looking at their statistical properties.

III. PROPAGATION OF ENERGY CORRELATION

The source v can be nonstationary or a wide sense stationary random function. The energy correlation function of a random function is used to measure the strength of aberration and it has the following definition:

$$\Gamma_v(\tau) = \int_{-\infty}^{\infty} v(x)v(x+\tau)dx, \quad (3.1)$$

where $\tau = x - x'$ and x, x' are two arbitrary points in the transverse axis. It is well known that under stationary conditions the average Γ_v is infinite. In this case it is meaningless to consider Eq. (3.1); instead, the power correlation function, $\lim_{X \rightarrow \infty} \frac{1}{X} \int_{-X}^X v(x)v(x+\tau)dx$ has to be considered. Statistical properties of the medium are crucial in this discussion. Real materials can have very complicated microstructure, so an approximate description is potentially susceptible to large errors. Crystals can in general be considered as randomly distributed spatially with preferred or random orientation, and macroscopically the material can be isotropic or anisotropic. "Randomly distributed" or just "randomly" is being used here as a generic word; so the spatial arrangement of crystallites may follow any probability distribution. Here, for theoretical simplifications

the sample is considered as being composed of randomly oriented scatterers which are either isotropic or weakly anisotropic. This is a restrictive approximation but it appears to be justified as it explains many of the observed phenomena. Another important point is the crystallite shape, which can be described based on the scale length l . This length explicitly describes the form of correlation $\Gamma_k = \langle k(\mathbf{r})k(\mathbf{r}') \rangle$, which also fully describes μ . A single model is being used which characterizes μ statistically in terms of σ and l . This is a fair representation of metals with equiaxed crystallites whose spatial distribution can be described by an isotropic random process. Complex structures such as an inhomogeneous crystallite size distribution—elongated crystallites—will require a more sophisticated model. Microstructures with elongated crystallites in a preferred direction can experimentally be investigated by propagating ultrasound in multiple directions. The above assumptions can be summarized by saying that μ can be taken to be an isotropic process as a good approximation to describe microstructure for the polycrystal used in the experimental work. Under this condition the correlation function for μ can take any form as long as it is a function through its difference, that is $\Gamma_\mu = \sigma^2 f(|\mathbf{r}-\mathbf{r}'|)$, f is a suitable function, and $|\mathbf{r}-\mathbf{r}'|^2 = (x-x')^2 + (z-z')^2$. In the literature^{1,15} the function $\Gamma_\mu = \sigma^2 \exp[-\frac{|\mathbf{r}-\mathbf{r}'|^2}{l^2}]$ is extensively used and has been shown to be accurate in some applications.¹⁵ Regardless of the geometry for the boundary problem Eq. (2.1), isotropy in both directions is simply establishing that wave velocity variations between grains can be modeled based on a stationary process with exponential correlation in both directions. This is assuming that velocity variations can be measured in a sample big enough or infinite in both directions.

Let us calculate the mean of the squared difference of the phase, that is $\langle [\phi(x) - \phi(x')]^2 \rangle$ at two arbitrary points x and x' using relation Eq. (2.2). In doing so by using the exponential form for Γ_μ , the correlation function $C_\phi = \frac{1}{2} \langle [\phi(x) - \phi(x')]^2 \rangle$ follows as

$$\begin{aligned} C_\phi &= \frac{\bar{k}^2}{4} \int_0^{\delta z} \int_0^{\delta z} \langle [\mu(x, z) - \mu(x', z')]^2 \rangle dz dz' \\ &= \frac{\bar{k}^2 \sigma^2}{4} [1 - e^{-\tau^2/l^2}] \int_0^{\delta z} \int_0^{\delta z} e^{-(z-z')^2/l^2} dz dz' \\ &= \frac{\sqrt{\pi} \bar{k}^2}{4} l \sigma^2 \delta z (1 - e^{-\tau^2/l^2}) \operatorname{erf}\left(\frac{2\delta z}{l}\right), \end{aligned} \quad (3.2)$$

where erf denotes the error function and $\tau = x - x'$. erf can be ignored as a good approximation providing $\delta z \approx l$ or $\delta z \gg l$ which gives $\operatorname{erf}(2\delta z/l) \approx 1$. Under these circumstances the calculation of C_ϕ is equivalent to assume from the very beginning that μ is delta correlated in the z direction or that the correlation of μ takes the form $\Gamma_\mu = \sigma^2 \exp[-\tau^2/l^2] \delta(z-z')$. Using this correlation function, the calculation of C_ϕ leads exactly to Eq. (3.2) with $\operatorname{erf}(\frac{2\delta z}{l})$ replaced by 1, which will be used to calculate the correlation function of the field.

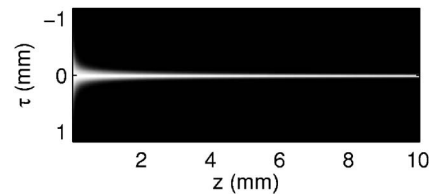


FIG. 3. Image of the energy correlation propagating in a random medium— z direction—with parameters $\sigma=0.02$ and $l=150 \mu\text{m}$, Eq. (3.3).

The energy correlation function of the field is transversally calculated for each propagation distance. Thus, by denoting the correlation function of the field by Γ_u , which is the ensemble average $\langle u(x, z)u^*(x', z) \rangle$ of the field for each propagation distance z , the explicit expression of C_ϕ will allow us to explicitly write Γ_u .

In general, u_n may be statistically related to ϕ for a single layer, because as v propagates from layer to layer u_n depends on ϕ . It will be shown that it is possible, at least mathematically, that the energy correlation of the field can be calculated if the medium is statistically independent of the incident field. Hence, based on those approximations, the expression for the energy correlation function at the entrance and exit of a region are simply related by

$$\Gamma_u = \Gamma_v(\tau) e^{-NC_\phi}, \quad (3.3)$$

where C_ϕ is given in Eq. (3.2). An example of the energy correlation over a distance corresponding to several crystallites, as calculated according to Eq. (3.3), is illustrated in Fig. 3. The decay and width as it propagates is determined by σ and l , respectively. The extreme case is for a highly aberrated medium, that is, $\sigma \rightarrow 1$ and small l —small crystallites—then Eq. (3.3) decays rapidly having a narrow tail. The ideal case occurs when $\mu=0$, that is homogeneous medium, so Eq. (3.3) does not change with propagation distance.

To show that the last expression is valid in this approximation, Eq. (2.7) will be considered by transforming it to the spatial domain.²¹ For that purpose, an independent coordinate is attached for each screen. Therefore, let $\mathbf{x} = (x_0, \dots, x_n)$ be that coordinate system. So if $\phi_0(x_0), \dots, \phi_n(x_n)$ denotes the phase variation at each screen, then one can define the correlation function at each screen as $C_\phi(x_s, y_s) = \langle [\phi_s(x_s) - \phi_s(y_s)]^2 \rangle$. Recall that ϕ is isotropic so C_s is a function of the difference $\tau_s = x_s - y_s$. Let us also consider the product of screen defined as $s(\mathbf{x}) = e^{i2s_s \phi_s(x_s)}$. Now the Green's function in the paraxial approximation of Eq. (2.1) is given by $g(x-x', z) = \sqrt{k/2\pi z} e^{i(k/2z)(x-x')^2}$ then the propagator in several variables is simply

$$G(\mathbf{x}, \mathbf{z}) = \prod_{j=1}^n g(x_j^-, z_j^-), \quad (3.4)$$

where $x_j^- = x_j - x_{j-1}$ and $z_j^- = z_j - z_{j-1}$.

Recall that one wants to calculate the transverse correlation of the field, i.e., $\Gamma_u(x, x', z) = \langle u(x, z)u^*(x', z) \rangle$ at distance z from the source, thus one has to take the ensemble average Eq. (2.7) in the spatial domain. In order to do that let us make $H(\mathbf{x}, \mathbf{y}) = G(\mathbf{x}, \mathbf{z})G^*(\mathbf{y}, \mathbf{z})$ and define the ensemble av-

erage of vs as $f(\mathbf{x}, \mathbf{y}) = \langle v(x_0)v(y_0) \rangle \langle s(\mathbf{x})s^*(\mathbf{y}) \rangle$; the ensemble has been split because v and s are statistically independent. Thus the ensemble average $\langle u(x_n)u^*(y_n) \rangle$ using Eq. (2.7) is given by

$$\Gamma_u = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}, \mathbf{y}) H(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (3.5)$$

which can be shown to lead to Eq. (3.3). To be able to integrate Eq. (3.5) one would need to calculate the average f but this is not necessary as long as f is a function of the difference $\mathbf{x} - \mathbf{y}$ only. This is a consequence of ϕ being Gaussian and an isotropic process. Thus, using a standard result in Gaussian multivariate statistics²¹ f takes the form

$$f = \langle v(x_0)v(y_0) \rangle e^{-\sum_{s=1}^N C_\phi(x_s - y_s)}. \quad (3.6)$$

To continue the evaluation of integral Eq. (3.5) more notation is introduced to shorten the length of the equations. Let $\Lambda_s^- = \frac{2(z_{s-1} - z_s)}{\bar{k}}$ and $r_s = (x_s - x_{s-1})^2 - (y_s - y_{s-1})^2$ with obvious definition in vectorial form. Then Eq. (3.5) can be rewritten as

$$\Gamma_u = b \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}, \mathbf{y}) \prod_{s=1}^N \left\{ \frac{\exp\left[-i \frac{r_s}{\Lambda_s^-}\right]}{\Lambda_s^-} \right\} d\mathbf{x} d\mathbf{y}, \quad (3.7)$$

where $b = \left(-\frac{1}{\pi}\right)^N$. A further step in calculating the above integral follows by making the following change of variables: $2\mathbf{x} = \mathbf{p} + \mathbf{q}$, $2\mathbf{y} = \mathbf{q} - \mathbf{p}$, therefore $r_s = (p_s - p_{s-1})(q_s - q_{s-1})$ or $r_s = p_s^- q_s^-$. Thus, Γ_u in the new coordinate system is

$$\Gamma_u = b \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{p}, \mathbf{q}) \prod_{s=1}^N \left\{ \frac{\exp\left[-i \frac{p_s^- q_s^-}{\Lambda_s^-}\right]}{\Lambda_s^-} \right\} d\mathbf{p} d\mathbf{q}. \quad (3.8)$$

Now, using that Eq. (3.6) depends only on the difference of its coordinates, therefore f would be a function of \mathbf{p} only, we can perform integration with respect to \mathbf{q} . Recognizing that the function to be integrated is the Fourier transform of the identity, this results in a product of delta functions. But first, let us express the term appearing inside the exponential function as

$$-i \frac{p_s^- q_s^-}{\Lambda_s^-} = i \sum_s \left[\frac{p_s^-}{\Lambda_s^-} - \frac{p_{s+1}^-}{\Lambda_{s+1}^-} \right] \quad (3.9)$$

with $p_1^- = p_{n+1}^- = 0$, since we have added extra terms for convenience. After inserting Eq. (3.9) in Eq. (3.8) and performing integration with respect to \mathbf{q} , except for the single variable q_0 , we have

$$\Gamma_u = \int_{-\infty}^{\infty} f(\mathbf{p}, q_0) \prod_{s=1}^N \Lambda_s^- \delta\left(p_s^- - \frac{\Lambda_s^- p_{s+1}^-}{\Lambda_{s+1}^-}\right) \prod_{s=2}^N \frac{1}{\Lambda_s^-} d\mathbf{p} dq_0, \quad (3.10)$$

where δ is the delta of Dirac. In the above expression N is an even integer otherwise one would have to multiply the term on the right by $(-1)^N$.

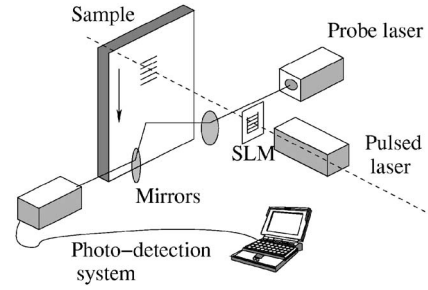


FIG. 4. O-SAM, the optical scanning acoustic microscope. A pattern generated by a spatial light modulator is imaged onto the sample using a pulsed laser. This pattern acts as the source of the surface acoustic waves. The waves are detected by another laser, using the optical beam deflection technique.

Finally integration can be completed by noting that if $\Lambda_s^- = \Lambda_{s+1}^-$ for all s , i.e., all screens are allocated at equal distance in space then we have

$$f(\vec{p}_N, q_0) \prod_{s=2}^N \Lambda_s^- = \int_{-\infty}^{\infty} f(\mathbf{p}, q_0) \prod_{s=1}^N \Lambda_s^- \delta\left(p_s^- - \frac{\Lambda_s^- p_{s+1}^-}{\Lambda_{s+1}^-}\right) d\mathbf{p}. \quad (3.11)$$

Here, $\vec{p}_N = (p_N, \dots, p_N)$. The final expression for Γ_u is obtained by inserting Eqs. (3.6), (3.9), and (3.11) into Eq. (3.10). In doing so

$$\Gamma_u = e^{-NC_\phi(p_N)} \int_{-\infty}^{\infty} \left\langle v\left(\frac{p_N + q_0}{2}\right) v^*\left(\frac{p_N - q_0}{2}\right) \right\rangle dq_0 = e^{-NC_\phi(p_N)} \Gamma_v(p_N) \quad (3.12)$$

which is equivalent to Eq. (3.3) with $p_N = \tau$. The above calculations show that the energy function at distance L is equivalent to the product of individual energy functions at the exit of each layer. Letting N tend to ∞ in Eq. (3.3), Γ_u approximates to a continuous solution of the second-order moment of the Helmholtz's equation. An approximate solution for the second-order moment of Eq. (2.1) is given in Ref. 12 and closely coincides with Γ_u . An expression for the coherence function of the backscattering field is also given in that paper.

IV. MEASUREMENTS

A. The O-SAM instrument

Over the past few years an optical scanning acoustic microscope (O-SAM) has been developed.²² This highly flexible instrument can be fully automated and is capable of performing multiple acoustic measurements over the surface of a sample. A complete set of software and electronics has been developed for gathering information at high speed. Typically, an amplitude and phase c-scan over an area of $1.5 \text{ cm} \times 1.5 \text{ cm}$ with a resolution of $10 \mu\text{m}$ takes a few minutes. Advantage has been taken of these capabilities and the O-SAM was used to build up an ensemble of the acoustic field over the surface of an aberrating material. The main components of O-SAM are shown in Fig. 4. It uses a Q-switched mode-locked Nd-YAG laser for SAW generation, by using a spatial light modulator (SLM) to image any

desirable pattern—typically a set of arcs or straight lines—onto the surface of the material under investigation. This image, illuminated by the pulsed laser, acts as the source of the surface waves. The fundamental frequency at which the O-SAM generates ultrasound is 82 MHz, but multiples of that frequency can also be generated. A second continuous wave laser was used to detect the propagating surface waves using an optical beam deflection technique. Both the detection system and the sample are mounted on computer-controlled automated stages, and so the O-SAM is capable of rapidly imaging the propagating wavefront, at any position on the sample. A more complete overview and technical details are given in Ref. 22.

B. Sample preparation

Three different aluminum samples were created, each with a different mean crystallite size. The procedure is similar for all of them so only a detailed description for one sample is described here. An Al (99.9%) charge of 500 g in mass, contained in a clay bonded SiC crucible, was heated to 730 °C in a muffle furnace. After melting of the Al charge and in order to obtain a lightly refined Al ingot, 0.2 wt. % of an Al-Ti-B commercial crystallite refiner was added and dissolved into the melt. Prior to removal of the oxide skin from the surface of the molten metal, the melt was cast into a rectangular steel mold in which it was allowed to solidify naturally. The Al ingot was released from the mold and sectioned with a band saw. Due to the geometry of the steel mold, a coarse columnar crystallite structure is expected in the top part of the Al ingot. For this reason, that section was removed and four useful blocks were obtained. Owing to the symmetry of the ingot, only three blocks were used; one for the counter part for metallographic characterization and the third was subjected to macro-etching to reveal the overall crystallite structure. Samples for metallography were taken from one block and were mounted, ground and polished down to 1 μm following standard procedures. The same preparation was given to the counter face of the other block. To reveal the crystallite structure, the Al block was repeatedly immersed into a solution (38% H₂O, 45% HCl, 15% HNO₃, and 2% HF) and washed until a good contrast was achieved. Also, the Al-polished samples were anodized in a 2% solution of HBF₄ in water for 1 min at 25 V.²³ After washing and drying, the samples were viewed and imaged in an optical microscope, equipped with a digital camera, under cross-polarized light.

Finally the crystallite size distribution of each of the three samples was determined by applying open source software.²⁴ Using the software, the perimeter was measured for each region contained within each micrograph and stored in a file for mean estimation. The mean caliper diameter, as defined in Ref. 25, was obtained by dividing the mean diameter of each region by π . The results were approximately as follows: 1345 μm , 785 μm , and 134 μm for A, B, and C, respectively. Also Fig. 5 shows the standard deviation and number of regions considered, σ_i , N_i where $i=A,B,C$, respectively for each block characterized.

Sample A has a more complicated microstructure com-

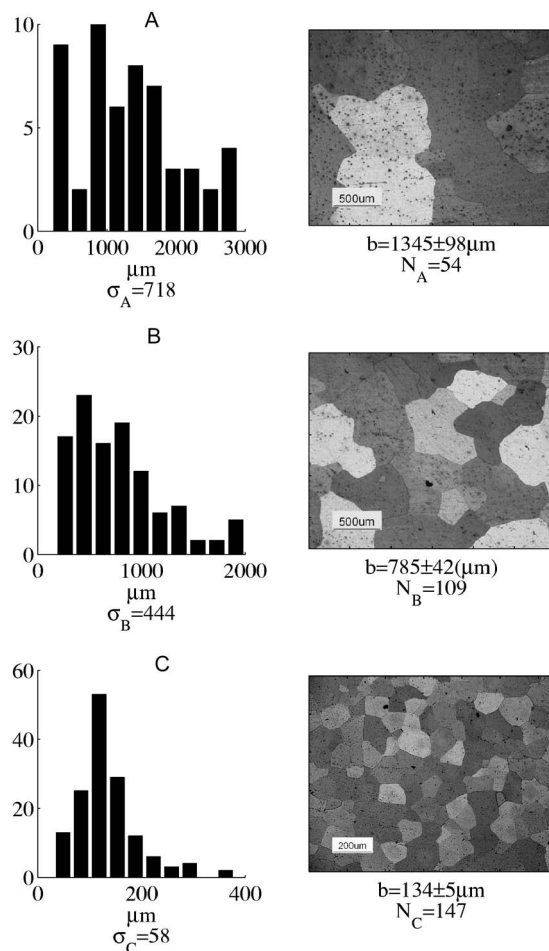


FIG. 5. The first column is the graph for the crystallite size distribution for each micrograph to the right. The second column shows micrographs of aluminum with variations on microstructure.

pared to samples labeled B and C. The latter two have homogeneous distribution of convex crystallites, as can be seen in the micrographs in Fig. 5. On the other hand A has large crystallites of a more complicated form. This feature made samples B and C easy for characterization and ultrasonic analysis, while A was more difficult.

C. Aberrations in aluminum

Aberrations were investigated in blocks of aluminum of 6 cm \times 4 cm \times 1 cm, which gives sufficient room for multiple measurements since the scanning area is typically 3 mm \times 10 mm. The SLM was programmed to project a series of straight lines onto the surface of the sample, each line separated from its nearest neighbor by a distance equal to the wavelength of the surface waves at 82 MHz, which is the fundamental frequency of the excitation laser. This pattern is used to propagate a plane wave. The propagation resembles a diffraction pattern through a slit since the SLM has finite aperture, so diffraction occurs near the edges. Figure 6 is a typical image of a plane wave propagating from left to right in an inhomogeneous medium. The wavefront breaks up due to aberration caused by the crystallite structure, leading to variations in the amplitude of the wavefront. The statistical analysis of the speckle pattern obtained is expected to pro-

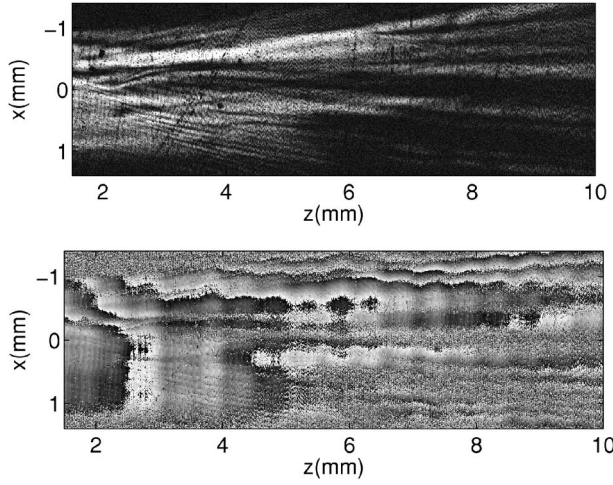


FIG. 6. Amplitude (top) and phase (bottom) of 82 MHz surface acoustic waves, propagating on aluminum. The wavefront breaks up as the ultrasonic wave travels through the crystallites.

vide information about the microstructure. In order to allow the ultrasound to interact with different sets of crystallites the ultrasonic source was positioned at different locations, scanning an area of $3 \text{ mm} \times 10 \text{ mm}$ each time. In this way, approximately 100 pictures similar to those in Fig. 6 were obtained; this was then used to produce an ultrasonic ensemble. All of them are different in detail but are assumed to arise from the same statistical population. Figure 7 is a schematic representation of the procedure used to make multiple measurements. It is important to highlight that the procedure is fully automated so it was only necessary to create a script in order to perform all the measurements.

V. STATISTICAL ANALYSIS

A. Procedure

The discussion is based on statistical concepts such as the second-order moment or energy correlation function for finite sequences. Notation will be introduced to explain some of the concepts and to be able to compare them with the theory earlier developed in Sec. III. Measurements are being denoted by U_{kl}^n as in Fig. 7, so U_{kl}^n represents any of the acoustic ensemble measured on block A, B, and C, where $k=1, \dots, K$, $l=1, \dots, L$ and K, L are determined by the resolution of the measured acoustic field in x, z direction, respectively. The superindex $n=1, \dots, N$ is to indicate the number of field measured on each sample. By definition the cross

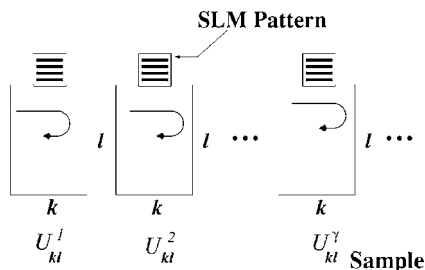


FIG. 7. Schematic representation of scanning area and source locations to build up an ultrasonic ensemble.

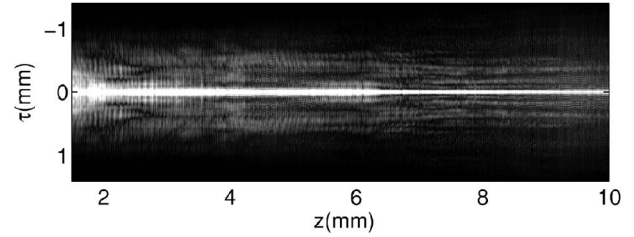


FIG. 8. Measured correlation function according to Eq. (5.2). This is a single instance of Γ_e . Averaging over many instances gives Γ_e which resembles the function in Fig. 3.

correlation is $\langle U_{kl}^n U_{k'l}^{n*} \rangle$ where $\langle \rangle$ denotes the ensemble average. The estimation of the ensemble average of $Z_{kk'l}^n = U_{kl}^n U_{k'l}^{n*}$ is rather complicated since there is little statistical information about U_{kl}^n . Instead, two different averages will be performed. By making $m=k-k'$, $Z_{kk'l}^n$ can be rewritten as $Z_{k,k+m,l}^n = U_k^n U_{k+m,l}^{n*}$. Since there is a waveform for each m , the average over k is performed as well as the ensemble average, leading to

$$Z = \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K Z_{k,k+m,l}^n \quad (5.1)$$

Z is a complex function so its modulus will be considered, and it will be termed energy correlation or the autocorrelation function; hence,

$$\Gamma_e = \|Z\|. \quad (5.2)$$

The subindex e stands for experimental and is used to differentiate it from the theoretical one. The measured autocorrelation function Γ_e , implemented numerically, is shown in Fig. 8. The importance of this function will become apparent when it is compared to the theoretical Γ_u in Eq. (3.3), since it is related to microstructure. Figure 8 shows that the autocorrelation propagates as an ultrasonic disturbance. The overall decay of this function is strongly related to aberrations due to the interaction between the crystallite structure and the ultrasound. There are two parameters which are free in Eq. (3.3), being σ and l which represents the average crystallite size. An estimation of σ, l has been obtained, for cases A, B, and C by minimizing the following function along propagation distance:

$$\chi^2(\sigma, l) = \sum_{i=1}^K \sum_{j=1}^L [\Gamma_u(\tau_i, z_j; \sigma, l) - \Gamma_e]^2 \quad (5.3)$$

This method of estimating the parameter σ, l proved to be useful only for samples labeled B and C as it will be shown in the last section.

B. Corroboration of model and statistical analysis technique

In Sec. III, Eq. (3.3) linked the theoretical statistical properties of the material microstructure with the autocorrelation function obtained from an ensemble of measured acoustic fields. In order to corroborate the approach of the analysis of this measured data, the phase screen approximation model described in Sec. II was used to simulate a set of

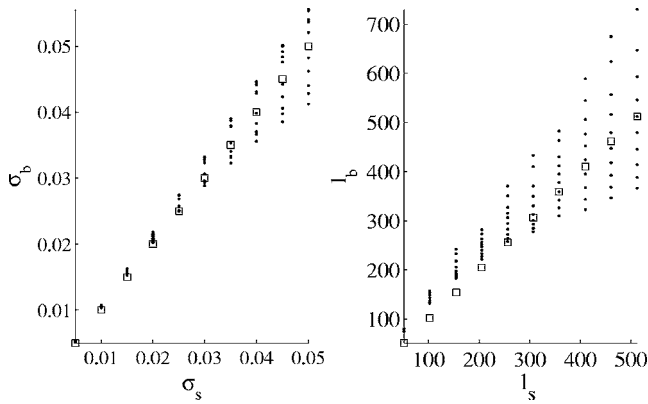


FIG. 9. Subscript s stands for simulation and subscript b for fitted parameters according to Eq. (3.3).

ultrasonic fields propagating through a simulated aberrating medium of known statistical properties. Each of the fields propagated through different simulated crystallite structures, and their corresponding propagating correlation functions were combined into an ensemble average as described by Eq. (5.2). The statistical analysis described in Sec. V A was performed, and the results for standard deviation (σ) and mean crystallite size (l) were compared to the values used to generate the ultrasonic fields. The simulations were repeated for different values of variance and mean crystallite size, and the results are illustrated in Fig. 9. We are using the symbols l_s , σ_s for correlation length and standard deviation used in the simulations, respectively, whereas l_b , σ_b will stand for the

best values obtained by minimizing χ^2 , Eq. (5.3), for each simulation. In Fig. 9, in the left graph, σ_s is used for the abscissas as well as the ordinate, that is (σ_s, σ_s) . In the same box, (σ_s, σ_b) is also being plotted for comparison. So if σ_s and σ_b are the same then the graph will be a straight line. The same principle applies for the graph on the right where we are plotting (l_s, l_s) as well as (l_s, l_b) . The small squares in the left graph in Fig. 9 correspond to values of σ_s whereas the squares in the right to l_s . The black dots correspond to σ_b and l_b in the same order as before for each simulation.

One hundred different mediums were simulated by feeding the algorithm with ten values for σ_s and ten for l_s , varying σ_s from 0.01 to 0.05, and l_s from 51 to 512. For each pair (σ_s, l_s) , one hundred fields were generated to be able to have a good estimation of average Eq. (5.1). The agreement between the values used to simulate the random microstructure, and the values obtained from statistical analysis of the ensemble autocorrelation functions is good, particularly for the standard deviation. The spread on the estimated crystallite size according to Eq. (5.3) needs more careful analysis since function Eq. (3.3) becomes complicated in that region because σ and l are tightly related, so the algorithm has some difficulties in identifying the right values, particularly so for the correlation length.

It should be remembered that the method used to generate the simulated random crystallite structure and the analysis of the autocorrelation functions are not directly math-

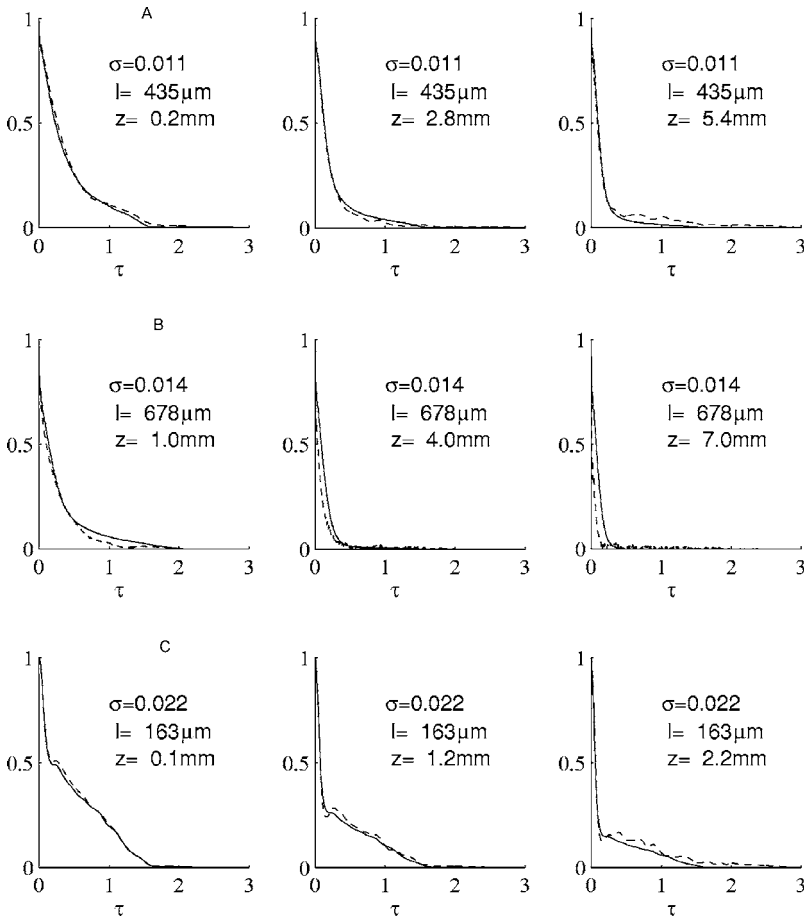


FIG. 10. Comparison of normalized energy correlation functions Γ_u and Γ_e for samples in Fig. 5. The continuous line (—) is derived from Eq. (3.3) at different propagation distances, with values of σ and l shown in each graph. The dashed line (---) represents the correlation function of the experimental acoustic field at the same propagation distances. The parameters σ and l are within a 20% accuracy.

ematically related. The simulation has been used in a previous paper to study the effects of aberration on wave velocity measurements.²⁶

C. Analysis of experimental data

The main result in this section is the comparison between theory and the measured autocorrelation function. The scan area for each picture was taken to be 3 mm wide whereas the propagation distance was chosen until the ultrasonic field became diffuse. In the first case, sample C, the maximum propagation distance was 2.5 mm for instance, using steps of 5 μm in the x axis and 100 μm in the direction of propagation (the z axis). The experimental data acquired by the O-SAM instrument was processed in the same way as the modeled data used at the beginning of this section, and comparisons are made between the measured (Γ_e) and predicted (Γ_u) energy correlation functions at various propagation distances.

Accumulated noise at the central peak of the energy function gives a very sharp peak affecting the overall decay of the energy function, and therefore the estimated values of σ , l . The data were filtered assuming a linear model of the form $y_1 = y_2 + e$ where e is a white noise statistically uncorrelated to y_1 , and y_2 is data free of noise. When filtered with an optimum filter the residual is delta correlated, which corresponds to noise. This can be removed with retention of the desired signal. This very simple model considerably reduces the central peak due to noise, giving the results shown in Fig. 10.

Figure 10 shows the comparison of the measured (Γ_e) and predicted (Γ_u) energy correlation functions—where Γ_u is shown in solid lines—for samples A, B, and C. In each case, it is shown at three different propagation distances in order to illustrate the decay of the correlation function with distance. The dashed lines in Fig. 10 represent the measured energy correlation function on the samples at the same propagation distances, derived from the acoustic ensemble in samples A, B, and C. There is good agreement for samples B and C. The standard deviation which measures the velocity variations from crystallite to crystallite used in Fig. 10, for comparison, is approximately $\sigma \approx 0.015$ which is a value that one would expect for aluminum.¹

The nonlinear fitting Eq. (5.1), numerically solved, gives the following estimation for σ and l . [σ l] = [0.011 435 μm], [σ l] = [0.014 678 μm], and [σ l] = [0.022 163 μm] for samples A, B, and C, respectively. The estimation of σ is reasonable in all cases, however, the estimated correlation length for sample A is significantly different from the values obtained visually, which are 1345 μm , 785 μm , and 134 μm as shown in Fig. 5.

Possible reasons for this are as follows. First, due to mechanical limitations in the O-SAM instrument, the acoustic field on sample A could not be mapped in its entirety. This effectively truncated the available dataset from which an estimation could be made. Second, we note that the measured mean crystallite size (1345 μm) is approaching the width of the acoustic source (≈ 2 mm). This is significant, because Γ_e is influenced more by the acoustic aperture in this case than

by the correlation length. Finally, as noted in Sec. IV B, the large crystallites in sample A have a complicated form, in that many of the crystallites are nonconvex.

VI. CONCLUSIONS

A theory has been developed to relate statistical parameters of material microstructure to statistics of ultrasonic aberrations. From this relationship it is possible to investigate—indirectly—mean crystallite size and variance of the acoustic wave velocity under certain conditions.

Therefore, the technique as a whole may be used as a tool for material characterization. This technique is particularly suitable for materials that are weakly anisotropic and the crystallite structure is not very complicated. Also, it works well if the width of the source is bigger than the average crystallite size. A phase screen model, alongside the stochastic wave equation, has been developed to simulate ultrasound propagating through random media. This model has been used to corroborate the technique of statistical analysis of the propagating energy correlation function, and provides a useful test bed for developing the theory, alongside the experimental work. The experimental results shown are in good agreement with the analytic expression for the correlation function calculated in Sec. III, provided the conditions discussed in the article are met.

ACKNOWLEDGMENTS

I would like to thank The University of Nottingham for the funding provided, The Engineering and Physical Science Research Council (EPSRC), and the UK center for NDE for supporting this work. We would like also to thank the referees for their helpful and constructive comments.

- ¹R. Ranganathan and S. Dattagupta, "Theory of ultrasonic scattering by grains in polycrystals," *Phys. Status Solidi A* **54**, 537 (1979).
- ²J. Saniie and T. Wang, "Statistical evaluation of backscattered ultrasonic grain signal," *J. Acoust. Soc. Am.* **84**, 400–408 (1988).
- ³P. Haldipur, F. J. Margetan, L. Yu, R. B. Thompson, and J. A. Turner, "Ultrasonic attenuation measurements in jet-engine nickel alloys," **20**, 1338–1345 (2001).
- ⁴F. J. Margetan, L. Yu, and R. B. Thompson, "Computation of crystallite-noise scattering coefficients for ultrasonic pitch/catch inspection of metals," *Rev. Prog. Quant. Nondestr. Eval.* **24**, 1300–1315 (2005).
- ⁵R. B. Thompson, L. Yu, and F. J. Margetan, "A formal theory for the spatial correlation of backscattered ultrasonic crystallite noise," *Rev. Prog. Quant. Nondestr. Eval.* **24B**, 1292–1315 (2005).
- ⁶J. McCoy, "A parabolic theory for stress wave propagation through inhomogeneous linearly elastic solids," *J. Appl. Mech.* **44**, 462–468 (1977).
- ⁷M. J. Beran and J. McCoy, "Mean field variations in a statistical sample of heterogeneous linearly elastic solid," **6**, 1035–1054 (1970).
- ⁸V. U. Zavorotnyi, V. I. Klyatskin, and V. I. Tatarskii, "Strong fluctuations of the intensity of electromagnetic waves in randomly inhomogeneous media," *Sov. Phys. JETP* **46**, 252 (1977).
- ⁹J. W. Goodman, *Statistical Optics* (Wiley, New York, 2000).
- ¹⁰K. F. Graff, *Wave Motion in Elastic Waves* (Ely House, London, 1975).
- ¹¹L. Rose, "Point-source representation for laser-generated ultrasound," *J. Acoust. Soc. Am.* **75**, 723–732 (1984).
- ¹²D. A. de Wolf, "Backscatter corrections to the parabolic wave equations," *J. Opt. Soc. Am.* **6**, 174–179 (1989).
- ¹³V. E. Ostashev and V. I. Tatarskii, "Representation of the Helmholtz equation solution in the form of a series based on backscattering multiplicity," *Waves Random Media* **5**, 125–135 (1995).
- ¹⁴C. L. Rino, "On propagation in continuous random media," *Waves Random Media* **2**, 59–72 (1991).
- ¹⁵A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Aca-

- demic, New York, 1970).
- ¹⁶K. Sobczyk, *Stochastic Wave Propagation* (PWN-Polish Scientific Publishers, Warsaw, 1984).
- ¹⁷M. D. Feit and J. J. A. Fleck, "Light propagation in graded-index optical fibers," *Appl. Opt.* **17**, 3990–3997 (1978).
- ¹⁸J. M. Martin and S. M. Flatté, "Intensity images and statistics from numerical simulation of wave propagation in 3-d random media," *Appl. Opt.* **27**, 2111–2125 (1988).
- ¹⁹D. J. Thomson, "A wide-angle split-step algorithm for the parabolic equation," *J. Acoust. Soc. Am.* **74**, 1848–1854 (1983).
- ²⁰J. A. Neubert, "Asymptotic solution of the stochastic Helmholtz equation for turbulent water," *J. Acoust. Soc. Am.* **48**, 1203–1211 (1970).
- ²¹L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).
- ²²S. D. Sharples, M. Clark, and M. G. Somekh, "All-optical adaptive scanning acoustic microscope," *Ultrasonics* **41**, 295–299 (2003).
- ²³V. H. Lopez, S. Truelove, and A. Kennedy, "Fabrication of Al-TiC master composites and their dispersion in Al, Cu and Mg melts," *Mater. Sci. Technol.* **19**, 925–930 (2003).
- ²⁴ImageJ: <http://rsb.info.nih.gov/ij/>. Last viewed 2/7/07.
- ²⁵J. Ohser and F. M., *Statistical Analysis of Microstructure in Materials Science* (Wiley, New York, 2000).
- ²⁶M. Clark, S. D. Sharples, and M. G. Somekh, "Optimization using measured Green's function for improving spatial coherence in acoustic measurements," *Ultrasonics* **42**, 205–212 (2004).

Reproducible cavitation activity in water-particle suspensions

Bram M. Borkent, Manish Arora, and Claus-Dieter Ohl^{a)}

Faculty of Science and Technology, Physics of Fluids, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 1 October 2006; revised 23 December 2006; accepted 2 January 2007)

The study of cavitation inception in liquids rarely yields reproducible data, unless special control is taken on the cleanliness of the experimental environment. In this paper, an experimental technique is demonstrated which allows repeatable measurements of cavitation activity in liquid-particle suspensions. In addition, the method is noninvasive: cavitation bubbles are generated using a shock-wave generator, and they are photographed using a digital camera. The cavitation activity is obtained after suitable image processing steps. From these measurements, the importance of the particle's surface structure and its chemical composition is revealed, with polystyrene and polyamide particles generating the highest yields. Further findings are that cavitation nuclei become depleted with an increasing number of experiments, and the existence of nuclei with varying negative pressure thresholds. Finally, a decrease of the cavitation yield is achieved by prepressurization of the suspension—indicating that the cavitation nuclei are gaseous. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436646]

PACS number(s): 43.35.Ei [AJS]

Pages: 1406–1412

I. INTRODUCTION

The discrepancy between the theoretical tensile strength of pure water and the much lower values reported in various cavitation experiments has been attributed to the presence of cavitation nuclei. Harvey *et al.*¹ suggested that stable, minute gas nuclei are present in cavities on the solid surfaces of suspended particles, and that such cavities could act as nucleation sources. Greenspan and Tschiegg² reported that removing particles larger than 0.2 μm in diameter increased the tensile strength of water to about 200 bar. The hypothesis of Harvey *et al.* developed over several years into the so-called “crevice model”^{3–7} which predicts the acoustic cavitation threshold required to nucleate a vapor cavity from a mote, and the threshold's dependency on bulk liquid properties and crevice attributes such as size, wettability, and geometry.

The experiments^{2,3,8} revealed that cavitation inception depends not only on various liquid properties but also on acoustic parameters such as pulse duration, frequency and the geometric form of the sound field, and (if present) the number of cavitation nuclei and their respective nuclei properties. The latter observation stimulated investigations into the role played by controlled amounts of impurities, such as suspended particles.^{9–14} As a result of that work it was recognized that the method used to determine whether a cavitation event has happened can play a crucial role in determining cavitation thresholds. In the majority of experiments an active detector is used^{9–11,13,14} which consists of a focused transducer operated in pulse-echo mode, typically at 30 MHz (see, for example, Madanshetty *et al.*^{13,14} for details). However, this invasive method affects the cavitation process,¹⁴ for example the acoustic field can convect the particles into the cavitation zone through acoustic radiation forces, and high acoustic frequencies might set particles into oscillatory

motion. The latter could stimulate the formation and subsequent merging of surface bubbles into microscale cavities, which has been termed acoustic coaxing.^{10,14} This hypothesis is also put forward in a more recent study in which a Keller-vortex flow system is used to induce cavitation on microparticles.¹²

Cavitation experiments, when performed under uncontrolled lab conditions, are usually difficult to reproduce, see, for example, the comments of Strasberg in 1959.³ His data showed a “surprising” mean deviation of 10%–15% for 10–20 measurements with unchanged experimental conditions. Madanshetty *et al.*¹⁴ and Deng *et al.*¹¹ suggest that cavitation thresholds are strongly influenced by the cleanliness of the system. They both stress the importance of monitoring and accounting for fine details in the liquid preparation and handling. Thus great experimental care has to be taken to set up the nuclei environment in order to obtain repeatable cavitation measurements. Still, cavitation threshold measurements typically have a standard deviation of about 15%, which is regarded as good; see, for example, Atchley *et al.*¹⁵

In the present work we emphasize the reproducibility of the experiment. This addresses the fact that the results have been reproduced in our lab a few months later with freshly prepared particle-water suspensions. Thus, we expect that this technique is a suitable method to compare cavitation activity—within the later specified error margins—also in different groups.

Given that most of the problems faced in cavitation experiments are related to the invasiveness of the cavitation detector and the cleanliness of the nuclei environment, we designed a method which overcomes both. In this study, we make use of a single acoustic cycle from a shock wave generator to nucleate cavitation bubbles in an ultraclean environment. Since the shock wave passage takes only a few microseconds, the interaction of the acoustic signal with the

^{a)}Electronic mail: c.d.ohl@tnw.utwente.nl

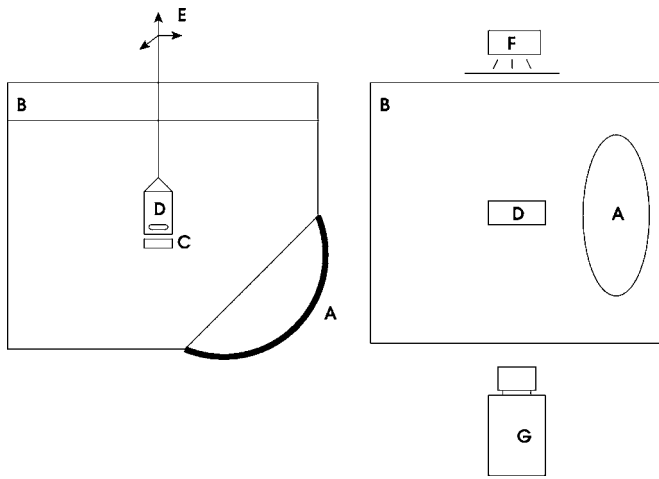


FIG. 1. Side and top view (respectively, left and right) of the experimental setup. A: shock-wave generator, B: water tank, C: magnetic stirrer, D: flask with particle suspension, E: translation stage, F: flash light, and G: camera.

cavitation bubbles is reduced to a minimum which distinguishes this approach from the ones presented in Refs. 2, 3, 8–11, and 13. In addition, we detect cavitation bubbles optically, thus quantitative and noninvasive. With this method we showed previously that cavitation bubbles nucleate on the microparticles itself and, as a direct consequence, the particles are accelerated away with velocities up to tens of meters per second.¹⁶ In contrast, we report here on the depletion of cavitation nuclei due to both the nucleation event itself and prepressurization of the particles. Finally, we report how new sets of cavitation nuclei can be activated by increasing the tensile stress in the liquid in a stepwise manner.

II. MATERIALS AND METHODS

A. Experimental setup

The noninvasive technique to detect cavitation has evolved from a prior experiment discussed in Arora *et al.*¹⁶ Cavitation is induced by a focused shock wave source and is detected with a camera, see Fig. 1. The shock wave source is attached at an angle of 45° in one of the walls of a large container filled with approximately 50 l of partially degassed demineralized water (~ 3 mg/l O_2 concentration at a temperature of approximately $20^\circ C$).

The probe suspension is contained within a polystyrene sterile flask (75 ml, EasyFlask, Nunclon). This is positioned with its center aligned with the acoustic focus of the shock wave source using an *xyz*-translation stage. A watertight magnetic stirrer (Telemodul 90407, Variomag) located below the flask is mixing the suspension in the flask homogeneously with a glass-coated magnet. Before submerging the flask in the basin its content is sealed with a foil (Parafilm, American National Can, Chicago). Here, care is taken that no air bubbles become entrapped while closing the seal.

Shock waves are generated with a focused piezoelectric source which is a slightly modified extracorporeal lithotripter (Piezolith 3000, Wolf GmbH). The strength of the applied tensile stress is set through the discharge voltage ranging from 0 to 9 kV on the piezoelectric transducers. A typical pressure profile (5 kV discharge voltage) *inside* the polysty-

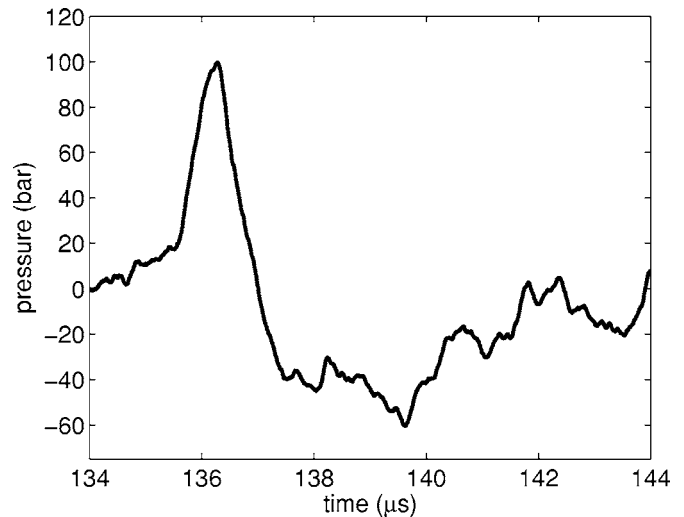


FIG. 2. Pressure as a function of time as recorded inside the flask (without particles) with the fiber optic probe hydrophone FOPH-500 for a charging voltage of 5 kV. The time $t=0$ corresponds to the start of the lithotripter pulse.

rene flask is shown in Fig. 2. The pressure pulse has been recorded with a calibrated glass fiber hydrophone (FOPH-500, RP Acoustics) located inside the sealed flask and close to the acoustic focus point, which is marked by two crossing laser diodes. The measurement shows that the travel time for the pressure wave from the front shell to the acoustic focus is approximately $136 \mu s$. Once the acoustic wave arrives, a steep overpressure-peak (maximum of 100 bar, half-width of $1 \mu s$) is followed by a more elongated tensile wave (minimum of -60 bar, $5 \mu s$ duration). During the latter period the threshold pressure is exceeded and cavitation bubbles are created (Fig. 3). If sufficient cavitation nuclei are present, an approximately cigar-like cavitation cluster is found.¹⁷

Bearing in mind that each expanding and collapsing bubble is acting as an acoustic source, the shape of the pressure wave following the positive peak in Fig. 2 may vary strongly with cavitation activity.¹⁸

The bubble activity is illuminated with a flash lamp (Strobolume 1540, General Radio) positioned behind a diffuser and photographed with a single frame from a digital charged-coupled device camera (Pulnix TM-6710, Alzenau,

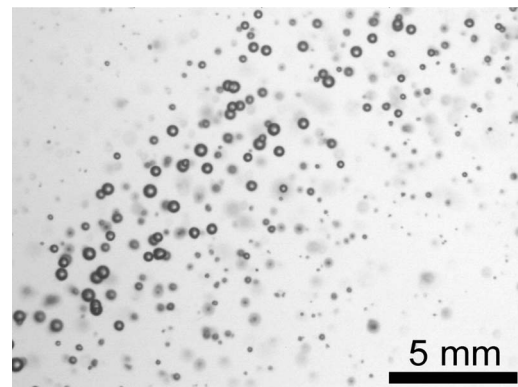


FIG. 3. Typical bubble cloud which appears approximately $14 \mu s$ after passage of the tensile wave (Fig. 2) inside of the flask. Settings: 10^4 particles/75 ml and 5 kV discharge voltage.

Germany). The camera is equipped with a macro lens (Makro-Planar T*2.8/60, Carl Zeiss) and placed around 30 cm from the acoustic focus, which is just outside the water tank. The devices are triggered with a home-made delay generator. The diffuse backillumination provides a good contrast for detecting the bubbles; they are imaged as dark dishes (Fig. 3). The spatial resolution of the images obtained is $32.1 \pm 0.1 \mu\text{m}$ per pixel, resulting in an image area of $15.4 \times 20.6 \text{ mm}^2$. In this image a region of interest (ROI) is selected parallel to the direction of the shock wave propagation being 45° to the horizontal. Bubbles in this ROI are segmented, counted, and analyzed with standard image processing tools (MATLAB, The Mathworks, Natick, MA). To limit the analysis to bubbles within a certain depth range a threshold technique is applied on the bubble border. Bubbles imaged in focus possess a strong gray scale gradient at their edge. Bubbles positioned away from the object plane are imaged with a lower magnitude of the edge gradient. Applying a threshold filter on the averaged gradient of the bubble boundary allows the detection algorithm to limit the counting of bubbles within a fixed depth. This depth has been determined by moving an artificial bubble object back and forth in the imaged volume. It was obtained that the bubble counting algorithm limits the detection depth to 8.0 mm.

B. Liquid handling

Stock solutions of the particles were made in test tubes and their number densities were determined with a Neubauer cell-counting chamber under the microscope. The corresponding volume containing 10^4 particles was pipetted from the test tube with a micropipette into the sterile 75 ml flask. Then, the flasks are filled up to their rim with Milli-Q water and sealed carefully. After placing the flask in the holder at the acoustic focus a shock wave is generated. The bubble activity is recorded approximately $14 \mu\text{s}$ after the passage of the tensile wave through the focus. To check for the water quality of the suspending liquid control experiments were conducted. Only if the flasks containing pure Milli-Q water showed less than five cavitation bubbles in the field of view the water was considered appropriately clean.

In the case of prepressurization (Sec. III C) the suspension was first introduced into a pressure chamber. This consists of a stainless steel cylinder having a maximum volume of 40 ml. The liquid in the cylinder is pressurized with a gas tight piston connected to the labs high pressure air supply. First, the liquid handling system is gently flushed with the suspension. Thereby, it was verified that most bubbles in this system are removed. Then, 40 ml of the suspension are sucked into the cylinder and pressurized with 3 or 5 bar absolute pressure for 15 min. In the control experiment all liquid handling steps were conducted in the same way with the exception of pressurizing the chamber. After 15 min the suspension was pushed smoothly out of the chamber into the flask and an additional volume of 35 ml of partly degassed Milli-Q water with no particles is gently added to fill up the flask, which is then sealed and put into the water tank.

C. Estimation of the cavitation yield

The cavitation yield, α , is defined as the percentage of particles in the acoustic volume which develop into cavitation bubbles, i.e.,

$$\alpha = \frac{N_{\text{bub}}}{N_{\text{particles}}} \times 100 \% . \quad (1)$$

Here, N_{bub} is the number of imaged bubbles and $N_{\text{particles}}$ is the number of particles. The acoustic volume is the volume in the flask where the pressure becomes low enough such that cavitation bubbles become visible. Although the particles are distributed homogeneously in the flask, it is important to stress that the negative pressure amplitude varies strongly in space. Thus, not all of the particles in the acoustic volume can develop into cavitation bubbles. To estimate this acoustically active volume we sum up many images with bubble activity. This averaged image reveals that bubbles appear in an approximately cylindrical region with its central axis aligned with the acoustic pathway: The cylinder length is $14.2 \pm 0.1 \text{ mm}$ and its diameter is $7.3 \pm 0.1 \text{ mm}$ leading to an optically registered acoustic volume as $0.59 \pm 0.02 \text{ cm}^3$. Clearly, these values vary with the applied acoustic pressure but here we limit to the case of 5 kV discharge voltage (corresponding to -60 bar peak tensile stress, see Fig. 2). With the number density of 133 particles/ml given and the assumption of a homogeneous distribution of particles, approximately 78 ± 3 particles are present in the optically registered acoustic volume. For instance, in the first shot of Fig. 6 66 bubbles were counted resulting into a cavitation yield of $\alpha = 85 \pm 3\%$. Thus, most of the particles serve as a cavitation nucleus.

III. RESULTS

The key request of the experiment was that a quantitative measure of cavitation activity of a controlled amount of particles could be obtained *in a reproducible way*. Reproducibility in cavitation experiments is a notorious problem since many factors may stimulate or impose nucleation, especially contamination. In the end, it was found that reliable quantitative data could be obtained by counting the number of bubbles produced by a controlled number of particles in a fixed volume at given shock wave settings while carefully handling the liquids and ultrapure water conditions.

A. Cavitation activity of microparticles

First we investigated different types of particles from which we select one type for further quantitative studies. Figure 4 depicts the number of cavitation events per particle for the ones listed in Table I. Here 10^4 particles per flask, thus 1.3×10^2 per ml, have been prepared and the shock wave generator has been operated from 1 to 5 kV discharge voltage with steps of 1 kV. (The pressure pulse corresponding to 5 kV discharge voltage is depicted in Fig. 2 and has a peak negative pressure of -60 bar .) The total number of bubbles, nucleated in these five subsequent shots, is depicted in Fig. 4. Clearly, polyamide and polystyrene enhance cavi-

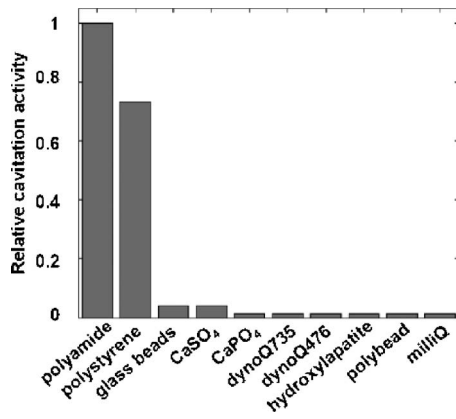


FIG. 4. Relative cavitation activity for different particles suspended in Milli-Q water. The control with the plain suspension liquid gives a count of 0.01.

tation significantly, while the other studied particles (either hydrophilic or covered with surfactants) show only cavitation activity similar to the plain liquid (Milli-Q water). Polyamide and polystyrene particles differ from the other particles as both can be categorized as hydrophobic and rough-structured on the surface. Interestingly, polyamide particles show more cavitation activity than polystyrene. This is surprising since polystyrene (consisting of methyl-groups and phenyl-rings) can be assessed as more hydrophobic than polyamide (which contains many polar amide-groups) and second because the polystyrene particles have around 40 times as much surface area compared with the relatively small polyamide particle, where cavitation nuclei could be entrapped (provided that necessary conditions are satisfied). On the other hand, scanning electron microscopy reveals that polyamide particles have a much rougher surface structure,

at least on the microscale as compared to polystyrene (Fig. 5). Thus, the higher surface roughness of polyamide may facilitate the entrapment and stabilization of gas pockets on the surface. The further studies are conducted with the polyamide particles.

B. Depletion of cavitation activity

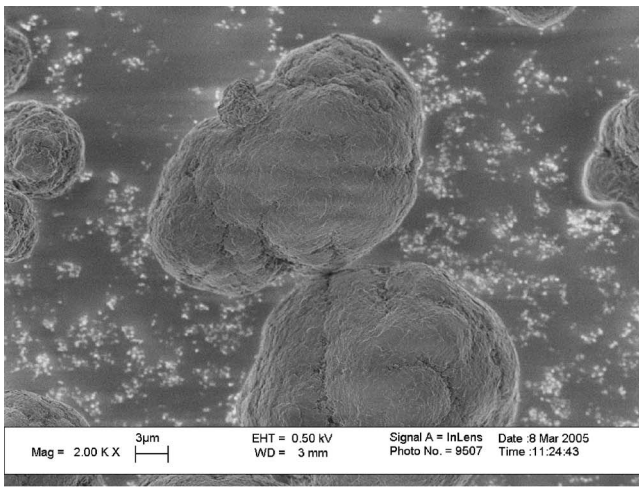
1. Effect of repetitive tensile waves

In the next experiment the effect of multiple pressure waves on the cavitation activity was investigated. 10^4 polyamide particles were suspended in a flask, and 40 subsequent waves at 5 kV discharge voltage (-60 bar peak negative pressure) were applied. The experiment was repeated three times and the number of cavitation bubbles measured. From the average number of bubbles, the mean percentage of nucleated particles per shot was calculated as the cavitation yield, α , and is shown in Fig. 6.

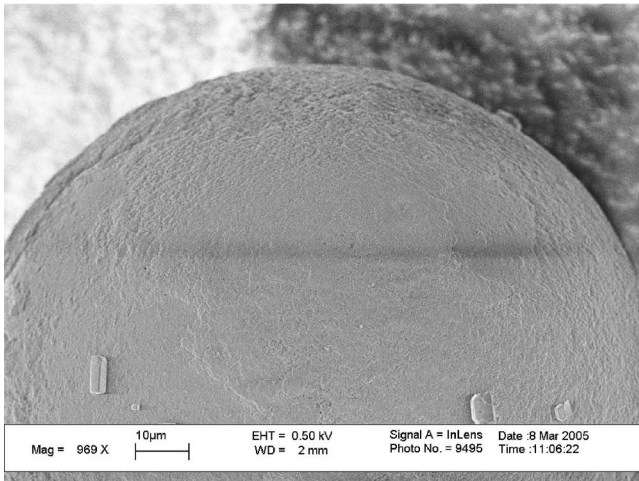
We find a decreasing number of cavitation events with a low variability between experiments, thus high reproducibility. Interestingly, it is found that the number of particles which can be nucleated decreases per shot. This can be explained that each particle has a limited number of nuclei which finally become used up. Additionally, the waiting time between two shots was varied from 10 to 60 s for the circles and the squares in Fig. 6, respectively. Cavitation yield is unaffected by the time interval. The dissolution time for the gaseous remains from cavitation bubbles can be estimated for bubbles reaching a maximum diameter of $400 \mu\text{m}$ during $40 \mu\text{s}$ in water saturated with air at room temperature. A bubble will thereby collect approximately 1.1×10^{12} g of air,¹⁷ which is equivalent to a bubble of $6 \mu\text{m}$ equilibrium radius. These sized bubbles again dissolve within 1.2 s, see,

TABLE I. Characterization of the particles investigated in this study on cavitation activity. An asterisk (*) behind the mean diameter indicates the manufacturers specification. In the other cases the particle sizes were determined with a microscope in the lab.

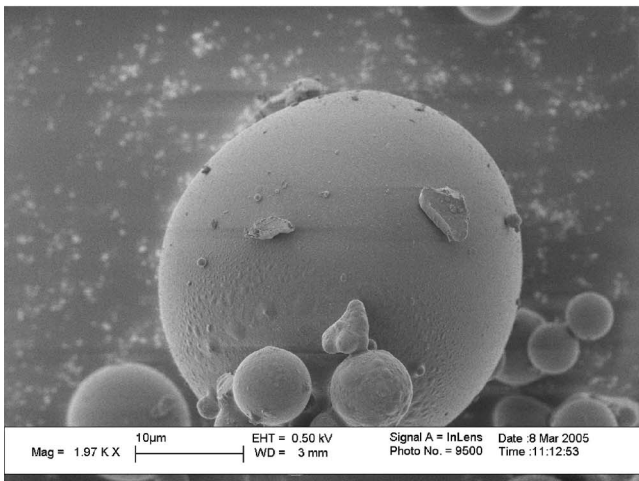
Particle name	Mean diameter (μm)	Features	Manufacturer (product number)
Hollow glass spheres (glass oxide)	11*	Hydrophilic, smooth surface	Potters Industries (110P8)
Polyamide seeding particles	20*	Hydrophobic, rough surface structure	Dantec Dynamics GmbH (80A4011)
Polystyrene-2% divinylbenzene	80–150*	Hydrophobic, rough surface structure	Serva GmbH
Dynosphere Q735, from Ref. 12	40	Hydrophilic, with hydroxyl groups	Dyno Particles AS, Norway (EXP-SS-42.3-RSH)
Dynosphere Q476, from Ref. 12	3.0*	Hydrophilic, smooth surface	Dyno Particles AS, Norway (EXP-SA-3.2-RNI)
Polybead polystyrene microspheres	1.00*	Filled with water, surfactant on surface	Polysciences
Hydroxylapatite ($\text{Ca}_{10}(\text{OH})_2(\text{PO}_4)_6$)	10–100	Biocompatible, slow-dissolving salt	Sigma Aldrich (289396)
β -tri-Calcium phosphate hydroxide ($\text{Ca}_3(\text{PO}_4)_3$)	5–10	Slow-dissolving salt	Fluka (21218)
Calcium sulphate Dihydrate ($\text{CaSO}_4 \cdot \text{H}_2\text{O}$)	10–100	Slow-dissolving salt	Sigma Aldrich (255548)



(a)



(b)



(c)

FIG. 5. SEM pictures of (a) polyamide, (b) polystyrene, and (c) glass spheres: Further specifications are in Table I.

for example, Refs. 19 and 20. Thus, a waiting time between two experiments of 10 s assures that gaseous remains from prior cavitation activity have dissolved.

2. Effect of tensile stress level

If one thinks of cavitation nuclei as becoming “used up” after nucleation, it is interesting to investigate what would

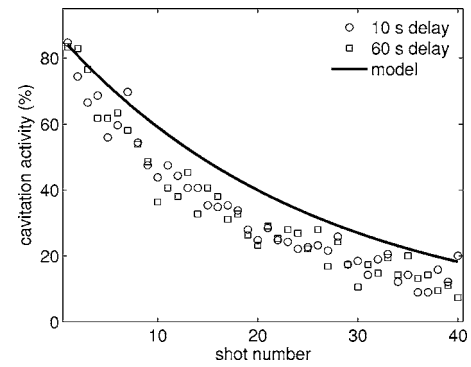


FIG. 6. Cavitation activity given as the percentage of particles that lead to a cavitation event. Each data point is the mean of three experiments. The waiting time of 10 s (circles) and 60 s (squares) has no effect on the cavitation yield. The line shows the values as calculated by the model, Eq. (3), with no adjustable parameter.

happen when the discharge voltage (and thus the tensile stress) is increased stepwise. Are we able to activate a new population of nuclei by applying higher tensile stresses? To answer this question, an experiment similar to the previous one has been performed, but now beginning with a discharge voltage of 4 kV and after 20 shots followed by 20 shots of 5 kV. This was repeated three times and the result is plotted in Fig. 7.

During the first 20 shots the nuclei which can be activated with a 4 kV discharge voltage are growing into cavitation bubbles. The trend of decrease in the cavitation yield resembles the one as depicted in Fig. 6. However, if after 20 shots the acoustic tensile stress is increased, more bubbles become visible, hence, a new set of nuclei is triggered. With the higher tensile stress the number of bubbles again starts to decrease for successive shots.

C. Prepressurization

The crevice model⁷ predicts that a pressurization of the suspension will force the meniscus of the gaseous nuclei to a more convex shape, until it reaches the advancing contact angle. An increased or prolonged pressurization will move the meniscus of the interface toward the apex of the assumed conical crevice: The bubble shrinks and liquid moves into the crevice. Here, the question is addressed if preexisting nuclei on the particles can be affected by pressurization of

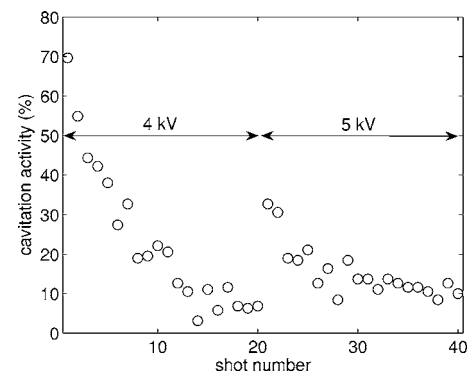


FIG. 7. Cavitation activity for two different tensile stresses, 4 and 5 kV experiments, respectively. Each data point is the mean of three experiments.

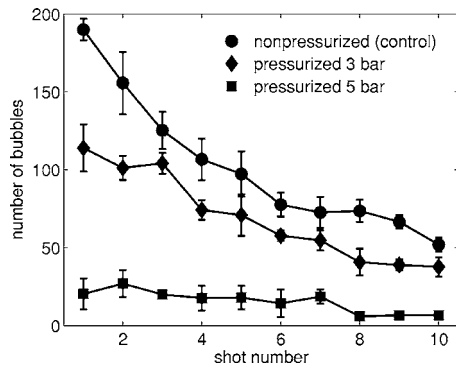


FIG. 8. Temporal evolution of cavitation after preexperimental pressurization. A decrease of cavitation of around 30% has been measured when shooting with 3 kV shots after pressurization with 3 bars (diamonds), compared to the nonpressurized scenario (circles). Pressurization with 5 bars (squares) leads to a cavitation reduction of more than 80%. Every data point is the mean of three experiments (bars depict standard deviation).

the suspension prior to the nucleation experiment. The effect of prepressurization of controlled water-particle-suspensions has not been investigated before, yet it was proposed by Deng *et al.*¹¹ Interestingly, Strasberg found that it is more difficult to obtain cavitation in tap water which had been subjected temporarily to a high pressure.³

The liquid handling during the pressurization step is described in Sec. II B. Two different preexperimental pressurization levels are compared with the case with no pressurization but the same fluid handling steps in Fig. 8. Again, each experiment is repeated three times and the mean number of cavitation bubbles is plotted. The control case, circles in Fig. 8, shows the typical decrease of cavitation activity with shot number. A prepressurization of the suspension with 3 bar absolute pressure for 15 min leads to a significant decrease of cavitation (diamonds in Fig. 8). Cavitation activity is largely diminished with a preexperimental treatment of the suspension with 5 bar overpressure, see squares in Fig. 8. Here, only about 20% of the bubbles are counted as compared to the control.

IV. DISCUSSION

A. Effect of exposure time

Interestingly, the particles (dynoQ735 and dynoQ476, see Fig. 4) which are from the same batch as the ones tested in Marschall *et al.*¹² did not show pronounced cavitation activity in response to the 60 bar peak negative pressure wave (Fig. 2). This is in bold contrast to Keller's vortex-flow nozzle experiment as described in Ref. 12. In that work a negative pressure of 0.87 bar was sufficient to induce cavitation with particles from the same batch. The obvious difference between the two experimental conditions is the exposure duration for the particles to the tensile stress. In the shock wave experiments, particles are exposed only a few microseconds, whereas in the flow cavitation experiments the tensile stress lasts two orders of magnitude longer.²¹ This difference suggests that not only the tensile strength but also the duration of negative pressure exposure is critical for the nucleation process. Our hypothesis is that a longer exposure to tensile stress allows a dynamic rearrangement and possi-

bly a coalescence of gaseous nuclei on the particle surface. These now larger nuclei possess a reduced cavitation threshold and thus explode at smaller absolute pressures.

B. Model

Assuming that each nucleus in the acoustic volume can grow into a cavitation bubble the observed number of bubbles, N_{bub} , should be equal to the number of nuclei present in the observed acoustic volume V_{ob} .

The number of nuclei in the observed acoustic volume is determined from

$$N_{\text{bub}} = \frac{V_{\text{ob}}}{V_{\text{flask}}} N_{\text{nuclei}}, \quad (2)$$

where $V_{\text{flask}} = 75$ ml is the total flask volume and N_{nuclei} the total number of nuclei available in the flask.

The observed acoustic volume is approximated from the experimental geometry with a cylinder, $V_{\text{ob}} = \pi r^2 l$ having a radius $r = 3.6$ mm and a length $l = 14.2$ mm. Initially (exposure number $n = 1$), the number of nuclei in the flask, N_{nuclei} , is related to the number of particles $N_{\text{particles}}$ through the cavitation yield, thus $N_{\text{nuclei}} = N_{\text{particles}} \alpha$, where α is the cavitation yield. We assume that at later stages ($n > 1$) nuclei become used up due to cavitation, thus N_{nuclei} decreases with every shot by a factor of $V_{\text{tav}}/V_{\text{flask}}$. Here, V_{tav} is the total acoustic volume $\pi r^2 l_{\text{flask}}$ where $l_{\text{flask}} = 71$ mm is the length of the acoustic volume in the entire flask. Therefore, the decrease of cavitation nuclei as a function of the exposure number n is given by

$$N_{\text{nuclei}}^{n+1} = \left(1 - \frac{V_{\text{tav}}}{V_{\text{flask}}}\right) N_{\text{nuclei}}^n. \quad (3)$$

Now the number of nucleated bubbles N_{bub} can be calculated from Eq. (2). Here, all model parameters are determined from the experiment and no fitting parameters come into play. The model Eq. (2) gives an exponential decrease in cavitation activity which resembles the trend in the experimental curve reasonably well. It is possible to obtain a very good fit to the measured decay and the model by changing the experimental parameters by no more than by 10%. Yet, we see the good agreement with the measured data only as an indication that indeed a particle is acting as one-time triggerable source of cavitation. This finding underlines the importance of the history of the suspension in cavitation studies, e.g., how many times it has been exposed to a rarefaction wave; and the golden rule of Robert Apfel.²² "Know thy liquid!" applies.

V. CONCLUSION

A novel experimental approach to study cavitation inception within a well controlled environment has been presented. A flask filled with a suspension of clean water and an adjustable number of particles is exposed to a single shock wave-rarefaction wave cycle. At a peak negative pressure amplitude of 60 bar, hydrophobic and corrugated polymer particles facilitate cavitation inception, while smooth and hydrophilic particles do not enhance cavitation activity. In the case of polyamide, around 80% of the particles nucleate after

the passage of the first shock wave. For successive shock waves, the number of cavitation bubbles decreases exponentially, which is shown by running a simple cavitation nucleation model. This exponential decrease suggests that particles have a limited number of nucleation sites which are consumed during the experiment. Therefore the level of cavitation activity reflects the history of the liquid. It was also found that more bubbles are observed with increasing negative pressure. This suggests that the nucleation sites on the particles have varying cavitation thresholds. Finally, it was observed that the cavitation yield decreases when the suspension is prepressurized suggesting that the cavitation nucleus is of gaseous origin.

ACKNOWLEDGMENTS

We are grateful to Professor Knud Aage Mørch and Professor Detlef Lohse for fruitful discussions and Gert-Wim Bruggert for technical support. The work is partly financed by “Stichting Technische Wetenschappen (STW),” “Stichting voor Fundamenteel Onderzoek der Materie (FOM),” and the “Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).”

¹E. N. Harvey, D. K. Barnes, W. D. McElroy, A. H. Whiteley, D. C. Pease, and K. W. Cooper, “Bubble formation in animals,” *J. Cell. Comp. Physiol.* **24**, 1–22 (1944).

²M. Greenspan and C. E. Tschiegg, “Radiation-induced acoustic cavitation: Apparatus and some results,” *J. Res. Natl. Bur. Stand., Sect. C* **71**, 299–312 (1967).

³M. Strasberg, “Onset of ultrasonic cavitation in tap water,” *J. Acoust. Soc. Am.* **31**, 163–176 (1959).

⁴R. E. Apfel, “The role of impurities in cavitation-threshold determination,” *J. Acoust. Soc. Am.* **48**, 1179–1186 (1970).

⁵L. A. Crum, “Tensile strength of water,” *Nature (London)* **278**, 148–149 (1979).

⁶K. A. Mørch, “Cavitation nuclei and bubble formation—A dynamic

liquid-solid interface problem,” *J. Fluoresc.* **122**, 494–498 (2000).

⁷A. A. Atchley and A. Prosperetti, “The crevice model of bubble nucleation,” *J. Acoust. Soc. Am.* **86**, 1065–1084 (1989).

⁸W. J. Galloway, “An experimental study of acoustically induced cavitation in liquids,” *J. Acoust. Soc. Am.* **26**, 849–857 (1954).

⁹R. A. Roy, S. I. Madanshetty, and R. E. Apfel, “An acoustic backscattering technique for the detection of transient cavitation produced by microsecond pulses of ultrasound,” *J. Acoust. Soc. Am.* **87**, 2451–2458 (1989).

¹⁰S. I. Madanshetty, “A conceptual model for acoustic microcavitation,” *J. Acoust. Soc. Am.* **98**, 2681–2689 (1995).

¹¹C. X. Deng, Q. Xu, R. E. Apfel, and C. K. Holland, “Inertial cavitation produced by pulsed ultrasound in controlled host media,” *J. Acoust. Soc. Am.* **100**, 1199–1208 (1996).

¹²H. B. Marschall, K. A. Mørch, A. P. Keller, and M. Kjeldsen, “Cavitation inception by almost spherical solid particles in water,” *Phys. Fluids* **15**, 545–553 (2003).

¹³S. I. Madanshetty and R. E. Apfel, “Acoustic microcavitation: Enhancement and applications,” *J. Acoust. Soc. Am.* **90**, 1508–1514 (1991).

¹⁴S. I. Madanshetty, R. Roy, and R. E. Apfel, “Acoustic microcavitation—Its active and passive acoustic detection,” *J. Acoust. Soc. Am.* **90**, 1515–1526 (1991).

¹⁵A. A. Atchley, L. A. Frizzell, R. E. Apfel, C. K. Holland, S. Madanshetty, and R. A. Roy, “Thresholds for cavitation produced in water by pulsed ultrasound,” *Ultrasonics* **26**, 280–285 (1988).

¹⁶M. Arora, C. D. Ohl, and K. A. Mørch, “Cavitation inception on micro-particles: A self-propelled particle accelerator,” *Phys. Rev. Lett.* **92**, 174501–1–4 (2004).

¹⁷M. Arora, L. Junge, and C. D. Ohl, “Cavitation cluster dynamics in shock-wave lithotripsy. 1. Free field,” *Ultrasound Med. Biol.* **31**, 827–839 (2005).

¹⁸Y. A. Pishchalnikov, O. A. Sapozhnikov, and M. R. Bailey, “Cavitation selectively reduces the negative-pressure phase of lithotripter shock pulses,” *ARLO* **6**, 280–286 (2005).

¹⁹O. A. Sapozhnikov, V. A. Khokhlova, M. R. Bailey, J. C. Williams, J. A. McAteer, R. O. Cleveland, and L. A. Crum, “Effect of overpressure and pulse repetition frequency on cavitation in shock wave lithotripsy,” *J. Acoust. Soc. Am.* **112**, 1183–1195 (2002).

²⁰P. S. Epstein and M. S. Plesset, “On the stability of gas bubbles in liquid-gas solutions,” *J. Chem. Phys.* **18**, 1505–1509 (1950).

²¹K. A. Mørch, private communication.

²²R. E. Apfel, “Acoustic cavitation inception,” *Ultrasonics* **22**, 168–173 (1984).

Parallel capillary-tube-based extension of thermoacoustic theory for random porous media

Heui-Seol Roh

Davidson Laboratory, Stevens Institute of Technology, Hoboken, New Jersey 07030

Richard Raspet^{a)} and Henry E. Bass

National Center for Physical Acoustics and Department of Physics and Astronomy,
University of Mississippi, University, Mississippi 38677

(Received 12 July 2006; revised 19 December 2006; accepted 30 December 2006)

Thermoacoustic theory is extended to stacks made of random bulk media. Characteristics of the porous stack such as the tortuosity and dynamic shape factors are introduced into the thermoacoustic wave equation in the low reduced frequency approximation. Basic thermoacoustic equations for a bulk porous medium are formulated analogously to the equations for a single pore. Use of different dynamic shape factors for the viscous and thermal effects is adopted and scaling using the dynamic shape factors and tortuosity is demonstrated. Comparisons of the calculated and experimentally derived thermoacoustic properties of reticulated vitreous carbon and aluminum foam show good agreement. A consistent mathematical model of sound propagation in a random porous medium with an imposed temperature is developed. This treatment leads to an expression for the coefficient of the temperature gradient in terms of scaled cylindrical thermoviscous functions. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436632]

PACS number(s): 43.35.Ud, 43.28.Kt, 43.20.Mv [MFH]

Pages: 1413–1422

I. INTRODUCTION

Symko *et al.*^{1,2} have built small high frequency thermoacoustic refrigerators and prime movers. The small pore sizes involved makes it difficult to construct stacks with parallel pores. Instead, random porous media such as metal, glass, or cotton wools are used. We thus investigate the extension of thermoacoustics to this case.

A linear wave equation which serves as the theoretical basis for the development of thermoacoustic devices was first established by Roti³ and extended by Swift.⁴ It was expanded by Arnott *et al.*,⁵ who introduced a variety of stack geometries. The thermoacoustic wave equation commonly used is rigorously applicable to a stack with a single pore but is then applied to stacks with parallel pores. Devices with stacks filled with random porous media show promise of improved operation. The potential for improved performance provides motivation for the development of a theory that can account for a stack composed of bulk porous media.

The development of the theory for thermoacoustic devices resembles that for sound propagation in porous media.^{6–14} The theory of sound propagation in a random porous medium is often derived assuming the medium is composed of capillary tubes. Capillary-tube-based theories employ the concepts^{7,8,10,15} of dynamic shape factor and tortuosity. Resultant predictions of absorption, wave velocity, etc., agree reasonably well with measurements using the shape factor as an adjustable variable.

In this paper, a porous media approach is used to extend thermoacoustic theory to describe a device with a stack consisting of bulk porous media. Scaled thermoviscous func-

tions for such a stack are introduced into the thermoacoustic wave equation. The wave number appearing in the thermoviscous functions includes the dynamic shape factors as scaling parameters. The tortuosity enters as a scaling parameter in the propagation constant. By using the low reduced frequency approximation,^{9,16} basic equations for a bulk porous medium are formulated in a manner similar to that employed for a single pore in the derivation of the thermoacoustic wave equation. To demonstrate the utility of the thermoacoustic wave equation, the calculated thermoviscous functions are compared with those deduced from published experimental data. Wilen and Petculescu^{17,18} measure the viscous function from the experimental specific acoustic impedance and the thermal function from the experimental compressibility. The extension of porous media scaling is used to predict the temperature gradient coefficient in terms of the scaled cylindrical functions. The implications for thermoacoustics are then discussed.

II. THERMOACOUSTIC THEORY FOR A RANDOM MEDIUM

Extension of the thermoacoustic wave equation to a bulk porous medium is developed in this section. The low reduced frequency approximation^{9,16} is used with a temperature gradient along the axis of sound propagation. This approach is consistent with the separate treatment of the viscous and thermal effects for sound propagation in a porous medium.^{7,10} A bulk medium stack is represented by a stack with identical, arbitrary cross-section pores inclined with respect to the sound propagation direction. Dynamic shape factors and tortuosity are used to describe the effects of the random porous medium on the acoustic properties of the stack. Basic assumptions for the extension of thermoacoustic

^{a)}Electronic mail: raspet@olemiss.edu

theory to a bulk random medium are as follows: The derivation of the wave equation is restricted to the linear regime. The stack materials are assumed rigid and stationary. Wavelengths are much bigger than penetration lengths. The pore walls have high heat capacity and thermal conductivity compared with those of the working gas. The acoustic pressure is constant over the cross section of the stack. The viscous and thermal effects for a random porous medium may be described by the same mathematical function with different scaling factors.

A. Basic equations for a single straight tube

Basic equations of motion for a fluid in a circular straight tube with a temperature gradient can be obtained in the low reduced frequency approximation by following Rasper *et al.*¹⁶ First-order acoustic normalized variables are

$$p(z,t) = (\rho_0 c^2 / \gamma) [1 + p^*(z) \exp^{-i\omega t}], \quad (1)$$

$$\mathbf{v}(r,z,t) = c [v^*(r,z) \hat{e}_r + u^*(r,z) \hat{e}_z] \exp^{-i\omega t}, \quad (2)$$

$$T(r,z,t) = T_0 [1 + T^*(r,z) \exp^{-i\omega t}], \quad (3)$$

and

$$\rho(r,z,t) = \rho_0 [1 + \rho^*(r,z) \exp^{-i\omega t}], \quad (4)$$

where the time dependence $e^{-i\omega t}$ is used. Symbols with a superscript (*) and a subscript 0 denote small, dimensionless fluctuation quantities and ambient quantities, respectively.

Dimensionless parameters or coordinates^{9,16} are defined as follows: The reduced frequency is defined by $\psi = \omega R / c$, where R is the tube radius and c is the sound velocity; the normalized axial coordinate is defined by $\xi = \omega z / c$; the normalized radial coordinate by $\eta = r / R$; and the normalized axial variable by $Z = z / L$ where L is the stack length.

The low reduced frequency approximation assumes that the internal tube radius is smaller than the wavelength, the transverse velocities are smaller than the axial velocity, and the pore radius is smaller than the stack length: $\psi \ll 1$, $v / u \ll 1$, and $R / L \ll 1$. Equations are linearized by assuming that steady flows are negligible and retaining terms with dT_0 / dz and $d\rho_0 / dz$.

The axial component of the Navier-Stokes equation is given by

$$iu^* = \frac{1}{\gamma} \frac{dp^*}{d\xi} - \frac{1}{\lambda_\mu^2} \left(\frac{\partial^2 u^*}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial u^*}{\partial \eta} \right) \quad (5)$$

and the radial component is

$$0 = \frac{1}{\gamma} \frac{dp^*}{d\eta}. \quad (6)$$

The viscous wave number λ_μ is defined as

$$\lambda_\mu = R \sqrt{\frac{\rho_0 \omega}{\mu}} = \frac{\sqrt{2} R}{\delta_\mu}, \quad (7)$$

where R is the tube radius, ρ_0 the gas density, ω the angular frequency of the sound, and μ the gas viscosity. λ_μ is a measure of the radius of the tube in terms of the viscous

penetration depth of the gas, δ_μ . The equation of continuity becomes

$$i\psi \rho^* = \psi \frac{\partial u^*}{\partial \xi} + \frac{\partial v^*}{\partial \eta} + \frac{v^*}{\eta} - \frac{R u^*}{L T_0} \frac{dT_0}{dZ}. \quad (8)$$

The equation of state is written as

$$p^* - \rho^* - T^* = 0. \quad (9)$$

The equation of heat transfer has the form of

$$T^* - \frac{\gamma - 1}{\gamma} p^* - \frac{i}{\lambda_\kappa^2} \left(\frac{\partial^2 T^*}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial T^*}{\partial \eta} \right) = \frac{1}{\psi L T_0} \frac{dT_0}{dZ}. \quad (10)$$

The thermal wave number λ_κ is defined by

$$\lambda_\kappa = R \sqrt{\frac{\rho_0 \omega c_p}{\kappa}} = \frac{\sqrt{2} R}{\delta_\kappa}, \quad (11)$$

where R and ρ_0 are as defined earlier, c_p is the heat capacity of the gas, and κ the thermal conductivity of the gas. δ_κ is the thermal penetration depth. λ_κ is related to λ_μ by the Prandtl number:

$$\lambda_\kappa^2 = N_{Pr} \lambda_\mu^2. \quad (12)$$

The boundary conditions at the tube wall $\eta = 1$ are as follows: The temperature oscillation vanishes; the heat capacity of the walls is high so the temperature of the tube wall does not fluctuate; there is no axial slip for the particle velocity at the tube wall; and the fluid cannot penetrate the tube wall.

B. Thermoacoustic theory for a bulk porous medium

Arnott *et al.*⁵ have shown that the thermoacoustic equations for a long straight parallel wall tube of arbitrary cross section corresponds to the form of the circular pore equations with the radial Laplacian replaced by the appropriate transverse Laplacian for the particular geometry

$$\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \rightarrow \nabla_T^2. \quad (13)$$

The solutions to Eqs. (5) and (10) can be written in terms of the solution to the differential equation

$$\frac{1}{\lambda^2} \nabla_T^2 F(x, y, \lambda) + iF(x, y, \lambda) = 1. \quad (14)$$

$F(x, y, \lambda)$ determines the behavior of the oscillating axial velocity and temperature as determined by viscosity, thermal conductivity, frequency, and pore shape for the given geometry. The acoustical behavior of a bulk medium composed of parallel pores is determined by the porosity Ω and by the area average of $F(x, y, \lambda)$; $F(\lambda)$. $F(\lambda)$ is called the thermo-viscous function:

$$F(\lambda) = \frac{1}{A} \int F(x, y, \lambda) dA. \quad (15)$$

For random porous media, the exact form and boundary conditions for the transverse Laplacian ∇_T^2 of the longitudinal particle velocity and temperature in thermoacoustics are not known. Good agreement between measurements and cylin-

drical pore theory have been obtained by adjusting the parameters λ_κ and λ_μ in the cylindrical thermoviscous function by a scaling parameter n :

$$F_A(\lambda) \equiv F_{\text{cyl}}(\lambda/n), \quad (16)$$

where $F_A(\lambda)$ is the exact undetermined thermoviscous function and $F_{\text{cyl}}(\lambda/n)$ is the thermoviscous function for the cylindrical pore and λ is the wave number calculated from the estimated characteristic radius R . For a parallel wall pore the characteristic radius is defined as twice the area divided by the perimeter. For random media, the characteristic radii are estimated from the pore linear density.

This follows porous media theory in which the thermoviscous functions for closed pores^{5,7,10} are fitted well using a scaled wave number. Cylindrical tube theory is thus employed as a basis for further derivations. Research in porous media for sound propagation often involves small pores where the viscous effect dominates. For that case, a single scaling factor suffices.

Champoux and Stinson¹⁹ have developed a model porous media using segments of cylindrical pores with different radii. For this canonical medium, two scaling factors are required. Consideration of a porous medium composed of cylinders perpendicular to the flow axis illustrates that different thermoviscous functions and different boundary conditions must be used to solve for $F(x, y, \lambda_\mu/n_\mu)$ and $F(x, y, \lambda_\kappa/n_\kappa)$ where n_μ is the viscous dynamic shape factor and n_κ is the thermal dynamic shape factor. For standing wave thermoacoustics,

$$F(\lambda_\mu/n_\mu) = 1 - (2/\lambda_{\mu n})\sqrt{i}(J_1(\sqrt{i}\lambda_{\mu n})/J_0(\sqrt{i}\lambda_{\mu n})) \quad (17)$$

and

$$F(\lambda_\kappa/n_\kappa) = 1 - (2/\lambda_{\kappa n})\sqrt{i}(J_1(\sqrt{i}\lambda_{\kappa n})/J_0(\sqrt{i}\lambda_{\kappa n})) \quad (18)$$

must be scaled differently. Wilen and Petculescu^{17,18} have shown that measurements of the thermal and viscous functions for two porous media can be related by simple scaling and that the thermal functions for reticulated vitreous carbon foams (RVC) can be fitted to the circular pore thermal functions with a reasonable fit.

The viscous and thermal dynamic shape factors, n_μ and n_κ , are included in the dimensionless scaling parameters

$$\lambda_{\mu n} = \lambda_\mu/n_\mu \quad (19)$$

and

$$\lambda_{\kappa n} = \lambda_\kappa/n_\kappa. \quad (20)$$

To apply this parallel pore approach to a random porous medium, the dynamic shape factors are used to account for pore shapes, the effect of changes in pore cross sections along their lengths, and uncertainty in how the characteristic radius should be determined for random porous media.

To develop predictions of the coefficient of the temperature gradient term in thermoacoustics we assume the Laplacian ∇_T^2 for local interaction can be applied to a random porous medium. The unknown transverse Laplacians $\nabla_{T\mu}^2$ and $\nabla_{T\kappa}^2$ for the viscous effect and thermal effect are replaced with the scaled transverse Laplacians in the cylindrical coor-

inate. The following replacements may be made to the basic equations for a single straight pore to get the equations for the equivalent bulk porous medium:

$$\begin{aligned} & \frac{1}{\lambda_\mu^2} \left(\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \right) u^*(\eta, \xi) \\ & \rightarrow \frac{1}{\lambda_{\mu n}^2} \nabla_{T\mu}^2 u^*(x, y, z) \rightarrow \frac{1}{\lambda_{\mu n}^2} \left(\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \right) u^*(\eta, \xi) \end{aligned} \quad (21)$$

and

$$\begin{aligned} & \frac{i}{\lambda_\kappa^2} \left(\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \right) T^*(\eta, \xi) \\ & \rightarrow \frac{i}{\lambda_{\kappa n}^2} \nabla_{T\kappa}^2 T^*(x, y, z) \rightarrow \frac{i}{\lambda_{\kappa n}^2} \left(\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \right) T^*(\eta, \xi), \end{aligned} \quad (22)$$

where $\lambda_{in} = \lambda_i/n_i$ with $i = \mu, \kappa$.

Again following porous media theory, the tortuosity q is introduced to express additional acoustic path length through the random porous medium. For a sample of thickness l with actual path length l' , $q = l'/l$. For a straight pore at angle θ to the normal $q = 1/\cos \theta$.^{7,8,10,15}

The goal is to develop equations which model sound propagation through the pores as plane wave propagation through a homogeneous medium with complex properties. u^* is now the particle velocity along the axis. The velocity within the pore (the velocity in the single pore differential equations) is given by qu^* ,

$$u^* \rightarrow qu^*. \quad (23)$$

The tortuosity is also inserted in the axial coordinate z in this bulk porous medium extension:

$$z \rightarrow qz. \quad (24)$$

The normalized axial coordinate and the axial variable are accordingly changed:

$$\xi \rightarrow \frac{q\omega z}{c} = q\xi \quad (25)$$

and

$$Z \rightarrow \frac{qz}{L} = qZ. \quad (26)$$

The tortuosity accounts for two effects. The pores themselves are longer so that viscous effects are larger in proportion to the length and the pressure gradient is also reduced by a factor proportional to the actual length.

Applying replacements from Eqs. (21) to (26), basic equations for a single circular straight pore from Eqs. (5), (9), and (10) become basic equations for a bulk porous medium as follows. The axial Navier-Stokes equation becomes

$$iu^* = \frac{1}{q^2 \gamma} \frac{dp^*}{d\xi} - \frac{1}{\lambda_{\mu n}^2} \left(\frac{\partial^2 u^*}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial u^*}{\partial \eta} \right). \quad (27)$$

The equation of state remains the same:

$$p^* - \rho^* - T^* = 0. \quad (28)$$

The equation of heat transfer is given by

$$T^* - \frac{\gamma-1}{\gamma} p^* - \frac{i}{\lambda_{kn}^2} \left(\frac{\partial^2 T^*}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial T^*}{\partial \eta} \right) = \frac{1}{\psi L T_0} \frac{dT_0}{dZ}. \quad (29)$$

At the tube wall $\eta=1$, there is no axial slip for the particle velocity: $u^*=0$. The temperature fluctuation vanishes at the wall: $T^*=0$.

The nondimensional particle velocity $u^*(\eta, \xi)$ is found¹⁶ from Eq. (27) by applying the boundary condition $u^*=0$ at the tube wall $\eta=1$:

$$u^*(\eta, \xi) = -\frac{1}{q^2} \frac{i}{\gamma} \frac{dp^*(\xi)}{d\xi} F(\eta, \lambda_{\mu n}), \quad (30)$$

where

$$F(\eta, \lambda_{\mu n}) = 1 - \frac{J_0(\eta, \sqrt{i\lambda_{\mu n}})}{J_0(\sqrt{i\lambda_{\mu n}})}. \quad (31)$$

The area-averaged particle velocity $u^*(\xi)$ from Eq. (30) is obtained by using $(1/A) \int F(\eta, \lambda_{\mu n}) dA = (1/\pi) \int_0^1 F(\eta, \lambda_{\mu n}) 2\pi\eta d\eta = F(\lambda_{\mu n})$:

$$u^*(\xi) = -\frac{1}{q^2} \frac{i}{\gamma} \frac{dp^*(\xi)}{d\xi} F(\lambda_{\mu n}), \quad (32)$$

where the viscous function $F(\lambda_{\mu n})$ is in the cylindrical coordinate,

$$F(\lambda_{\mu n}) = 2 \int_0^1 F(\eta, \lambda_{\mu n}) \eta d\eta = 1 - \frac{2}{\lambda_{\mu n} \sqrt{i}} \frac{J_1(\sqrt{i\lambda_{\mu n}})}{J_0(\sqrt{i\lambda_{\mu n}})}. \quad (33)$$

Substituting Eq. (30) into Eq. (29) yields

$$\begin{aligned} T^* - \frac{\gamma-1}{\gamma} p^* - \frac{i}{\lambda_{kn}^2} \left(\frac{\partial^2 T^*}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial T^*}{\partial \eta} \right) \\ = -\frac{1}{q^2} \frac{R}{\gamma \psi L T_0} \frac{dT_0}{dZ} \frac{dp^*(\xi)}{d\xi} F(\eta, \lambda_{\mu n}). \end{aligned} \quad (34)$$

The temperature fluctuation T^* in the absence of the temperature gradient dT_0/dz is the solution of

$$T^* - \frac{i}{\lambda_{kn}^2} \left(\frac{\partial^2 T^*}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial T^*}{\partial \eta} \right) = \frac{\gamma-1}{\gamma} p^*, \quad (35)$$

which has the same form as Eq. (27). $T^*(\eta, \xi)$ is found by applying the boundary condition $T^*(\eta, \xi)=0$ at the tube wall:

$$T_r^*(\eta, \xi) = \frac{\gamma-1}{\gamma} F(\eta, \lambda_{kn}) p^*(\xi). \quad (36)$$

Substituting Eq. (36) into Eq. (35), $F(\eta, \lambda_{kn})$ satisfies

$$\left(\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \right) F(\eta, \lambda_{kn}) = i\lambda_{kn}^2 (1 - F(\eta, \lambda_{kn})). \quad (37)$$

The boundary condition $F(\eta, \lambda_{kn})=0$ at the tube wall is required due to the boundary condition $T^*(\eta, \xi)=0$ at the tube wall.

Substituting Eq. (30) into Eq. (27), $F(\eta, \lambda_{\mu n})$ satisfies

$$\left(\frac{\partial^2}{\partial \eta^2} + \frac{1}{\eta} \frac{\partial}{\partial \eta} \right) F(\eta, \lambda_{\mu n}) = i\lambda_{\mu n}^2 (1 - F(\eta, \lambda_{\mu n})). \quad (38)$$

The boundary condition $F(\eta, \lambda_{\mu n})=0$ at the tube wall is a result of the boundary condition $u^*(\eta, \xi)=0$ at the tube wall.

The form of Eq. (34) suggests a trial solution of the form

$$T^*(\eta, \lambda_{\mu n}, \lambda_{kn}) = AF(\eta, \lambda_{kn}) + BF(\eta, \lambda_{\mu n}). \quad (39)$$

Combining Eq. (39) into Eq. (34), using Eqs. (38) and (37), and applying the boundary condition $T^*(1, \xi)=0$, the coefficients A and B are determined by setting the terms including $F(\eta, \lambda_{kn})$ or $F(\eta, \lambda_{\mu n})$ to zero. The temperature fluctuation T^* is given by

$$\begin{aligned} T^*(\eta, \xi) = \frac{\gamma-1}{\gamma} F(\eta, \lambda_{kn}) p^*(\xi) - \frac{1}{q^2} \frac{1}{\psi} \frac{R}{\gamma L T_0} \\ \times \frac{F(\eta, \lambda_{kn}) - (\lambda_{kn}/\lambda_{\mu n})^2 F(\eta, \lambda_{\mu n})}{1 - (\lambda_{kn}/\lambda_{\mu n})^2} \frac{dT_0}{dZ} \frac{dp^*}{d\xi}. \end{aligned} \quad (40)$$

The area-averaged temperature fluctuation can be written from Eq. (40) as

$$\begin{aligned} T^*(\xi) = \frac{\gamma-1}{\gamma} F(\lambda_{kn}) p^*(\xi) - \frac{1}{q^2} \frac{1}{\psi} \frac{R}{\gamma L T_0} \\ \times \frac{F(\lambda_{kn}) - (\lambda_{kn}/\lambda_{\mu n})^2 F(\lambda_{\mu n})}{1 - (\lambda_{kn}/\lambda_{\mu n})^2} \frac{dT_0}{dZ} \frac{dp^*}{d\xi}, \end{aligned} \quad (41)$$

where the thermal function $F(\lambda_{kn})$ is represented in cylindrical coordinates by

$$F(\lambda_{kn}) = 2 \int_0^1 F(\eta, \lambda_{kn}) \eta d\eta = 1 - \frac{2}{\lambda_{kn} \sqrt{i}} \frac{J_1(\sqrt{i\lambda_{kn}})}{J_0(\sqrt{i\lambda_{kn}})}. \quad (42)$$

From the equation of state (28), the averaged density fluctuation becomes

$$\rho^*(\xi) = p^*(\xi) - T^*(\xi). \quad (43)$$

Substituting Eq. (41) into Eq. (43), the averaged density fluctuation is found to be

$$\begin{aligned} \rho^*(\xi) = \left[1 - \frac{\gamma-1}{\gamma} F(\lambda_{kn}) \right] p^*(\xi) + \frac{1}{q^2} \frac{1}{\psi} \frac{R}{\gamma L T_0} \\ \times \frac{F(\lambda_{kn}) - (\lambda_{kn}/\lambda_{\mu n})^2 F(\lambda_{\mu n})}{1 - (\lambda_{kn}/\lambda_{\mu n})^2} \frac{dT_0}{dZ} \frac{dp^*}{d\xi}. \end{aligned} \quad (44)$$

At this point, the continuity equation for a bulk porous medium with the averaged variables is in first order

$$-i\omega \rho^*(z) + \frac{d[\rho_0(z)u^*(z)]}{dz} = 0. \quad (45)$$

Using Eq. (43), Eq. (45) leads to

$$-i\omega(p^*(z) - T^*(z)) + \rho_0(z) \frac{du^*(z)}{dz} - \beta \rho_0(z) u^*(z) \frac{dT_0}{dz} = 0 \quad (46)$$

or

$$-ic(p^*(\xi) - T^*(\xi)) + \rho_0(\xi) \frac{du^*(\xi)}{d\xi} - \beta \rho_0(\xi) u^*(\xi) \frac{dT_0}{d\xi} = 0. \quad (47)$$

The expression⁵

$$d\rho_0(\xi)/dz = -\beta\rho_0(z)dT_0/dz \quad (48)$$

and the transformation

$$dz = (c/\omega)d\xi \quad (49)$$

are used in Eqs. (46) and (47). Equation (46) is one of the new approximation method equations which is necessary for thermoacoustic predictions. The other is Eq. (32). These two equations play important roles in calculating thermoacoustic properties.²⁰

Substituting Eqs. (32) and (40) into Eq. (47), Eq. (47) becomes

$$\begin{aligned} \frac{i\psi}{2\gamma} \frac{d^2 p^*}{d\xi^2} F(\lambda_{\mu n}) + \frac{i q^2 \psi p^*}{2} \left(1 - \frac{\gamma-1}{\gamma} F(\lambda_{\kappa n})\right) \\ + \frac{i}{2\gamma L T_0} \left(\frac{F(\lambda_{\kappa n}) - F(\lambda_{\mu n})}{1 - N'_{Pr}}\right) \frac{dT_0}{dZ} \frac{dp^*}{d\xi} = 0, \end{aligned} \quad (50)$$

where the modified Prandtl number N'_{Pr} is

$$N'_{Pr} = \left(\frac{\lambda_{\kappa n}}{\lambda_{\mu n}}\right)^2 = N_{Pr} \left(\frac{n_{\mu}}{n_{\kappa}}\right)^2. \quad (51)$$

Converting Eq. (50) to dimensional form in terms of the complex pressure amplitude, $p_{ac} = p_0 p^*$, a second-order differential wave equation for a bulk fibrous medium is obtained:

$$\begin{aligned} \frac{\rho_0}{F(\lambda_{\mu n})} \frac{d}{dz} \left(\frac{F(\lambda_{\mu n})}{\rho_0} \frac{dp_{ac}(z)}{dz} \right) + 2\alpha_b(\lambda_{\mu n}, \lambda_{\kappa n}) \frac{dp_{ac}(z)}{dz} \\ + k_b^2(\lambda_{\mu n}, \lambda_{\kappa n}) p_{ac}(z) = 0. \end{aligned} \quad (52)$$

The coefficient α_b with the temperature gradient becomes

$$\alpha_b(\lambda_{\mu n}, \lambda_{\kappa n}) = \frac{\beta}{2} \frac{dT_0}{dz} \frac{F(\lambda_{\kappa n})/F(\lambda_{\mu n}) - 1}{(1 - N'_{Pr})}, \quad (53)$$

where $\beta = 1/T_0$ for an ideal gas. The propagation constant squared k_b^2 is

$$k_b^2(\lambda_{\mu n}, \lambda_{\kappa n}) = \frac{q^2 \omega^2}{c^2} \frac{1}{F(\lambda_{\mu n})} [1 + (\gamma-1)(1 - F(\lambda_{\kappa n}))]. \quad (54)$$

Equation (52) is the thermoacoustic wave equation for pressure fluctuation p_{ac} in the presence of a mean temperature gradient dT_0/dz in the stack. The propagation constant squared $k_b^2 = q^2 k_s^2$ for a bulk porous medium has the tortuosity factor squared q^2 to modify the propagation constant squared k_s^2 for a single pore.

Equation (54) recovers the porous media result for the propagation constant in terms of the scaled cylindrical thermoviscous functions. Equation (53) is a prediction of the temperature gradient coefficient in terms of the scaled functions. That term determines the acoustic gain when a temperature gradient is imposed on the stack. This prediction is the principal result of this paper.

Figure 1 displays photographs of two random porous media discussed in this paper and a conceptual drawing of the model porous media used for calculations.

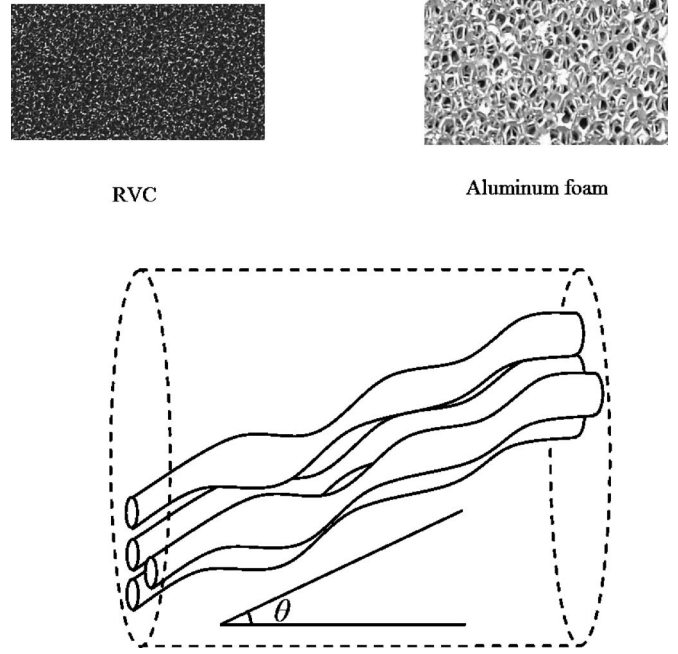


FIG. 1. Images of RVC and aluminum foam, as well as a conceptual diagram of the model porous media with cylindrical pores of varying radius inclined at an angle θ to the normal.

III. COMPARISON OF THEORETICALLY AND EXPERIMENTALLY DERIVED THERMOVISCIOUS FUNCTIONS

In this section, the experimentally derived thermoviscous functions are compared to the scaled cylindrical thermoviscous functions to demonstrate how well they agree for RVC and aluminum foam. The theoretically derived thermoviscous functions (33) and (42),

$$F(\lambda_{\mu n}) = 1 - \frac{2}{\lambda_{\mu n} \sqrt{i}} \frac{J_1(\sqrt{i} \lambda_{\mu n})}{J_0(\sqrt{i} \lambda_{\mu n})} \quad (55)$$

and

$$F(\lambda_{\kappa n}) = 1 - \frac{2}{\lambda_{\kappa n} \sqrt{i}} \frac{J_1(\sqrt{i} \lambda_{\kappa n})}{J_0(\sqrt{i} \lambda_{\kappa n})}, \quad (56)$$

are expressed as a function of the wave numbers (7) and (11), $\lambda_{\mu} = R\sqrt{(\rho_0 \omega / \mu)}$ and $\lambda_{\kappa} = R\sqrt{(\rho_0 \omega c_p / K)}$, which are functions of measurable quantities.

The experimental thermoviscous functions of RVC and aluminum foam are taken from Wilen and Petculescu.^{17,18} Figure 2 shows the viscous and thermal functions as a function of wave number. According to Ref. 18, the RVC stack has 20 pores per inch (ppi), characteristic radius $R = 0.00127$ m, porosity $\Omega = 0.97$, and length $L = 0.0895$ m. The Duocel aluminum foam stack has 40 ppi, characteristic radius $R = 0.000635$ m, porosity $\Omega = 0.94$, and length $L = 0.0889$ m. We will comment on the choice of the characteristic radius later.

The viscous function for a bulk medium with arbitrary pore shape (33) is calculated as a function of scaled viscous wave number (19) and the thermal function (42) is calculated

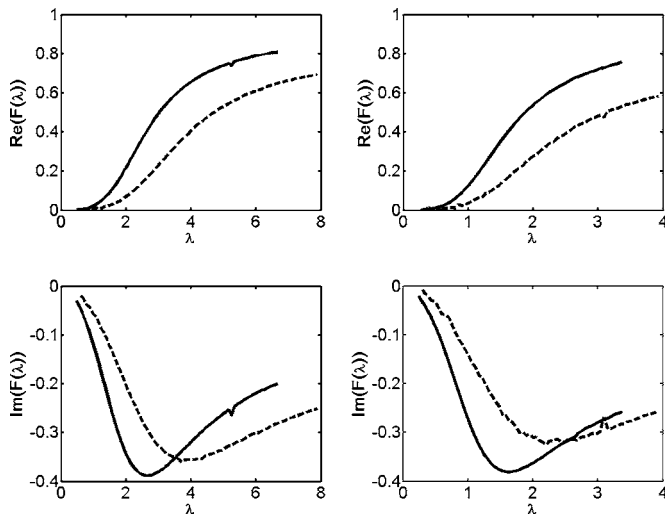


FIG. 2. Experimentally derived, thermoviscous functions for RVC (left) and aluminum foam (right). The dotted lines denote the effective viscous function and the solid lines the thermal function, taken from Petculescu and Wilen. λ for the viscous function is calculated from Eq. (7) and λ for the thermal function is calculated from Eq. (11).

as a function of scaled thermal wave number (20). Within this section, the effective viscous function is that defined by Petculescu and Wilen:¹⁸

$$F^e(\lambda_{\mu n}) = \frac{i\omega\rho_0 L}{Z_{\text{mea}} A_{\text{por}}}, \quad (57)$$

where the pore area is $A_{\text{por}}=A\Omega$, the cross-section area is A , stack length is L , and Ω is the porosity, the fraction of gas filled area of the porous medium. The impedance $Z_{\text{mea}}=p/(A_{\text{por}}u)$ was measured.

In this paper we have chosen to define the thermoviscous functions in terms of the pore space variables. The thermoviscous functions should approach unity as λ becomes large; indicating that the thermal or viscous waves are affecting a vanishingly small fraction of the pore space. To evaluate a pore space viscous function we replace L with Lq and u by uq in Eq. (57):

$$F(\lambda_{\mu n}) = \frac{i\omega\rho_0 Lq^2}{Z_{\text{mea}} A_{\text{por}}} = q^2 F^e(\lambda_{\mu n}). \quad (58)$$

The viscous function (58) rather than the effective viscous function (57) is used for comparison purposes. Note that the unscaled thermoviscous wave numbers (7) and (11) rather than the scaled wave numbers (19) and (20) are used as the horizontal axes in Fig. 2.

Comparisons of the calculated and experimentally derived viscous and thermal functions for the RVC stack are shown in Fig. 3 as a function of wave number λ calculated from Eq. (7) for the viscous function and Eq. (11) for the thermal function. The viscous dynamic shape factor $n_{\mu}=1.490$, thermal dynamic shape factor $n_{\kappa}=1.013$, and tortuosity squared $q^2=1.08$ are used as input for theoretical calculations of the thermoviscous functions of the RVC stack. The scaling of the viscous function to fit the thermal function is possible in both calculated and experimentally derived data. The calculated curve for the thermal function lies on

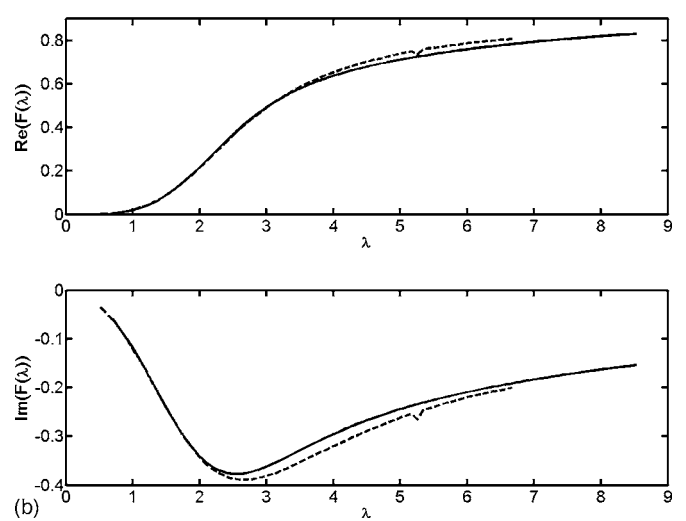
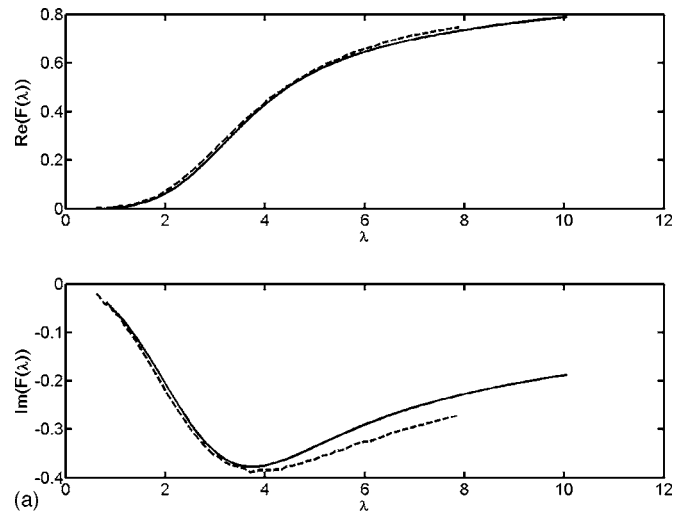


FIG. 3. Comparisons of the calculated and experimentally derived, viscous function (a) and thermal function (b) for RVC. The dotted lines indicate the experimental data and the solid lines the theoretical data. λ for the viscous function is calculated from Eq. (7) and λ for the thermal function is calculated from Eq. (11).

the top of the calculated curve for the viscous function. The scaling factor used in the horizontal axis is $n_{\mu}/n_{\kappa}=1.47$ and the scaling factor used in the vertical axis is $q^2=1.08$. These theoretical scaling factors are the same as the scaling factors experimentally deduced by Petculescu and Wilen.¹⁸

Figure 4 shows comparisons of computed and experimentally derived viscous and thermal functions for the aluminum foam stack. Viscous dynamic shape factor $n_{\mu}=0.900$, thermal dynamic shape factor $n_{\kappa}=0.638$, and tortuosity squared $q^2=1.2$ are used as input for calculations of the thermoviscous functions of the aluminum foam stack. Scaling of the viscous function to fit the thermal function is also possible in both calculated and experimentally derived quantities for the aluminum foam stack. The two computed curves for thermoviscous functions also match. Scaling factors in the horizontal axis and vertical axis are $n_{\mu}/n_{\kappa}=1.41$ and $q^2=1.20$, respectively. These theoretical scaling factors are the same as experimentally deduced by Petculescu and Wilen¹⁸ as is expected.

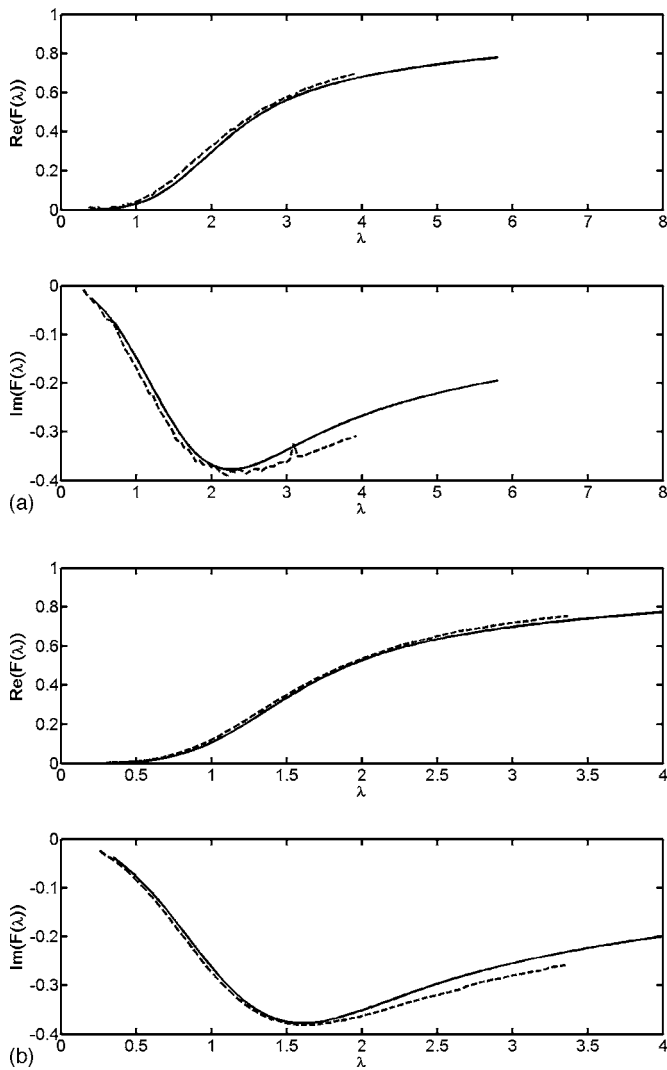


FIG. 4. Comparisons of the calculated and experimentally derived, viscous function (a) and thermal function (b) for aluminum foam. The dotted lines indicate the experimental data and the solid lines the theoretical data. λ for the viscous function is calculated from Eq. (7) and λ for the thermal function is calculated from Eq. (11).

Comparisons of computed and experimentally derived thermoviscous functions as a function of the wave number show good agreement. The two dynamic shape factors predicted for the viscous and thermal effects play the role of scaling the dimensionless wave numbers. The viscous and thermal dynamic shape factors are determined so as to fit the experimental data for the viscous and thermal effects. Champoux and Stinson¹⁹ studied a constructed porous media of straight cylindrical pore with alternative section of large and small radius. In that study, they noted that the effective characteristic radius for the viscous function was determined by the small radius of the pore and the thermal function depends strongly on the larger radius. We find the same type of result in RVC and aluminum foam. The viscous dynamic shape factor is larger than the thermal factor, implying the effective characteristic radius is smaller for the viscous effects than for the thermal effects.

The dynamic shape factors for the RVC stack are bigger than one but those for the aluminum foam stack are smaller than one when the characteristic radii are taken as the pore

spacing specified by the manufacturer. However, these values can be reevaluated based on a consistent definition of the characteristic radius for a random porous medium. The definition of characteristic radius should be consistent even though the effective radius R depends on geometry. For a cylindrical pore the effective radius is defined by $R=a$ where a is the radius of a cylindrical pore. In Ref. 18, many of the random porous media are specified in terms of the linear pore density, the pores per inch.¹⁸ To achieve a unified definition consistent with regular shape pores, the characteristic radius of a random porous medium is redefined to be half the inverse of the linear pore density. The characteristic radius $R=1/(20 \text{ ppi})=0.00127 \text{ m}$ used in Ref. 18 becomes $R=0.000635 \text{ m}$ for the 20 ppi RVC. Considering these definitions, a characteristic radius $R=0.000635 \text{ m}$ is obtained for the 20 ppi RVC stack. The viscous dynamic shape factor $n_\mu=0.745$ and the thermal dynamic shape factor $n_\kappa=0.507$ are reduced from the initial values $n_\mu=1.490$ and $n_\kappa=1.013$. Similarly, a characteristic radius $R=0.000508 \text{ m}$ is obtained for the 40 ppi aluminum foam stack since a spacing of 0.001016 m is given by the manufacturer for the specific porosity of 0.94. The viscous dynamic shape factor $n_\mu=0.789$ and the thermal dynamic shape factor $n_\kappa=0.559$ are reevaluated values instead of the initial ones $n_\mu=0.900$ and $n_\kappa=0.638$ for $R=0.000635 \text{ m}$.

The argument of the scaled thermoviscous functions is inversely proportional to the dynamic shape factors. Decreasing the dynamic shape factor from one means that the effective characteristic radius of the pore increases or the penetration depth decreases. Decreasing the dynamic shape factor also means that the optimum operating frequency decreases. The dynamic shape factors shift both the thermoviscous function and the temperature gradient coefficient. If the dynamic shape factor is less than one, the thermoviscous function shifts to a smaller frequency range than the viscous function with a dynamic shape factor one and vice versa. For a cylindrical pore, the viscous function has a larger argument for the same frequency than the thermal function since the Prandtl number of 0.72 is smaller than one. However, for both RVC and aluminum foam, the viscous function has a smaller argument than the thermal function since the modified Prandtl number $N'_{pr}=N_{pr}(n_\mu^2/n_\kappa^2)$ is bigger than one.

The tortuosity effect in the effective viscous function (57) is illustrated by the data. Figure 2 shows lower absolute amplitude for the effective viscous function than for the thermal function for both stacks at high wave numbers. The tortuosity plays the role of scaling the amplitude of the effective viscous function along the vertical axis. The vertical scaling factor ratio, q^2 , of the thermal function to the effective viscous function explains the fact that the measured data show higher absolute amplitude for the thermal function than for the viscous function for both stacks at high wave numbers after scaling. The higher the tortuosity, the bigger the amplitude gap between the effective viscous function and thermal function. This is explicitly illustrated since the vertical scaling factor 1.20 for the aluminum foam stack has a bigger amplitude gap than that of the RVC stack 1.08 in these comparison data. Note that the tortuosities matching theory to measurement do not differ greatly from one for parallel pores

in both RVC and aluminum foam. This is likely related to the fact that both stacks have large porosity values. The scaling data for the viscous function to the thermal function for random porous media such as RVC and aluminum foam provide good agreement to theoretical results and are in reasonable agreement with experimentally derived data. Reference 18 notes that the measurements do not extend to high enough frequencies to identify the scale factor unequivocally with the tortuosity. To get the viscous function $F(\lambda_{\mu n})$ defined by Eq. (33), it is emphasized that the tortuosity squared, q^2 , should be multiplied by measured values of the effective viscous function $F^e(\lambda_{\mu n})$ defined by Eq. (57). This verifies the validity of the initial assumption that the viscous function and thermal functions may be represented by the cylindrical thermoviscous functions with different scaling factors.

IV. TEMPERATURE GRADIENTS IN RANDOM MEDIA STACKS

In Sec. III, measured and scaled cylindrical thermoviscous functions for random porous media were computed and shown to be in good agreement for a zero temperature gradient. Based upon success at the zero temperature gradient, the capillary-tube-based extension of thermoacoustic theory will next be applied at a nonzero temperature gradient. The temperature gradient coefficient (53) unique to the thermoacoustic wave equation will be used to predict the performance for RVC and aluminum foam stack versus parallel pore stacks.

The temperature gradient coefficient is a function of the wave number in the presence of a temperature gradient. It would be nice to compare predicted values with experimental data reflecting the effect of a temperature gradient. In fact, data are not yet available. We define the normalized temperature gradient coefficient α in terms of the thermoviscous functions as

$$\alpha(\lambda_{\mu n}, \lambda_{\kappa n}) = \frac{F(\lambda_{\kappa n})/F(\lambda_{\mu n}) - 1}{(1 - N'_{Pr})}. \quad (59)$$

Figure 5 shows the normalized temperature gradient coefficient as a function of wave number for RVC and aluminum foam. The normalized temperature gradient coefficient depends upon several parameters. Among them, three (four) quantities are distinguished: frequency (or wave number), Prandtl number, and dynamic shape factor(s). The frequency is nominally fixed by resonator size. The Prandtl number depends on the working fluid type. The dynamic shape factors depend on the type of stack. The three variables are not completely independent. For example, the operating frequency can be adjusted by modifying the working fluid type that may also modify the dynamic shape factor. Similarly, the effective Prandtl number depends upon the working fluid type and the stack geometry/materials.

The modified Prandtl number N'_{Pr} defined in Eq. (51) for a random porous medium plays an essential role in optimization. The modified Prandtl number is bigger than the original Prandtl number N_{Pr} if the ratio n_{μ}^2/n_{κ}^2 is larger than one and vice versa. If $N'_{Pr} < 1$, the viscous function is shifted to a lower frequency range than the thermal function for both a

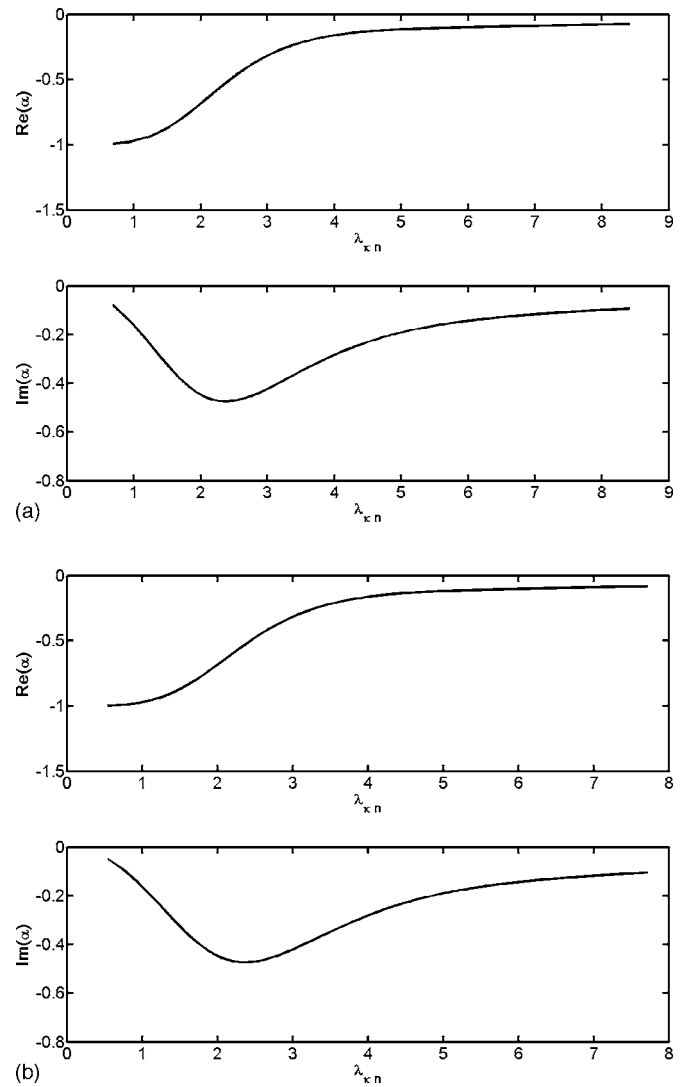


FIG. 5. Normalized temperature gradient coefficients are shown as a function of wave number for RVC (a) and aluminum foam (b). The data denote the predicted coefficients based on the theoretical thermoviscous functions.

random porous medium and cylindrical pores. If $N'_{Pr} > 1$, the thermal function is shifted to a lower frequency range than the viscous function for a random porous medium while the viscous function lies in the lower frequency range than the thermal function for a cylindrical pore. If $N'_{Pr} = 1$, the viscous function and thermal function are the same as a function of frequency. It is noted that the temperature gradient coefficient does not change sign even though N'_{Pr} goes through one.

The thermoacoustic conversion of heat to acoustic work or acoustic work to heat transport in a standing wave device is principally determined by the magnitude of the imaginary part of the temperature gradient coefficient.¹⁶ In Fig. 6, the normalized temperature gradient coefficient calculated for RVC is compared to the coefficient calculated for a circular pore stack and a parallel pore stack. The same characteristic radii, $R=0.000635$ m are used for all the stacks. The solid line represents the temperature gradient coefficient for the thermal dynamic shape factors $n_{\kappa}=0.507$ for RVC. The dotted line indicates the coefficient for a parallel plate stack, and the broken and dotted line denotes the coefficient for a

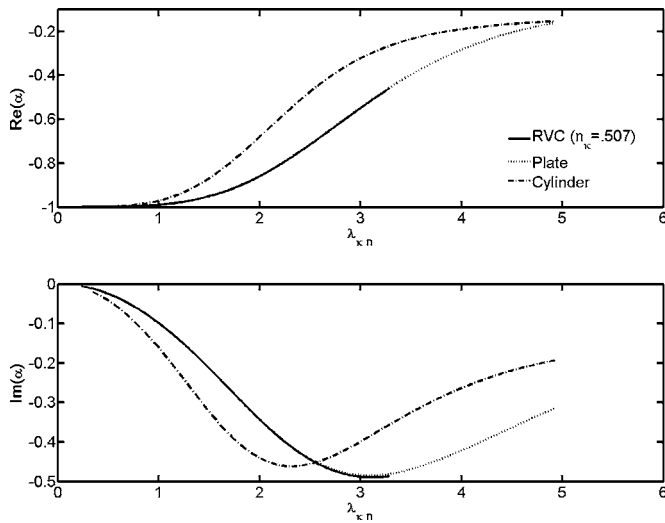


FIG. 6. Theoretical, normalized temperature gradient coefficient as a function of λ_{kn} for RVC, parallel plates, and cylindrical pores. Solid line: $n_k = 0.507$, $n_\mu = 0.745$ for RVC. Dotted line: a parallel plate stack. Broken and dotted line: a straight cylindrical pore stack. The coefficient for RVC with $n_k = 0.507$ overlays the coefficient for parallel plate pore.

straight cylindrical pore stack. The coefficient for RVC with $n_k = 0.507$ overlays the coefficient for parallel plate pore. Its imaginary part is thus as large as for a parallel plate stack. This suggests that RVC can be competitive with a parallel plate stack in thermoacoustic devices. As the thermal dynamic shape factor decreases or the viscous dynamic shape factor increases, the imaginary part of the coefficient increases in its absolute value. Increasing the ratio of the viscous dynamic shape factor to the thermal dynamic shape factor (n_μ/n_k) leads to increasing the imaginary part of the temperature gradient coefficient. This behavior is observed if $|1 - N'_{Pr}| < |1 - N_{Pr}|$. The increased imaginary magnitude is related to an increased performance of thermoacoustic devices. Increasing the thermal term compared to the viscous term in a random porous medium can be useful when choosing a random porous medium for thermoacoustic devices since the thermal term has a positive effect but the viscous term has a negative effect on the performance.^{4,21}

The predicted temperature gradient coefficients for porous media are comparable to parallel plate coefficients and larger than ones for the parallel cylindrical pores. This predicts that porous media should perform at least as well as parallel pore devices in thermoacoustic prime movers where stacks are very short and viscous losses are minimized. The increased viscous losses indicated by the relative scaling factors may reduce the performance of refrigerators since the stacks are longer in refrigerators than in prime movers.

V. CONCLUSIONS

A thermoacoustic theory for a bulk random medium is developed from that for a single pore, based on parallel, capillary-tube-based theories. To extend thermoacoustic theory to a bulk porous medium, the tortuosity, the viscous dynamic shape factor, and the thermal dynamic shape factor are introduced and the low reduced frequency approximation is employed. Basic equations such as the Navier-Stokes

equation, the equation of state, and the equation of heat transfer for a bulk porous medium are formulated for the derivation of the thermoacoustic wave equation as analogously as possible to those for a single pore. This theory develops the explicit expression for the thermoacoustic wave equation that contains the thermoacoustic functions. In the thermoviscous functions, the viscous and thermal dynamic shape factors play the scaling role for the dimensionless wave number. The scaled fits for RVC and aluminum foam are compared with experimentally derived data for the thermoviscous functions. The scaled and measured data show reasonably good agreement and verify the validity of the theoretical model for porous media. A prediction of the temperature gradient coefficient is developed in terms of the scaled thermoviscous functions. This prediction shows that random porous media stacks should perform as well as parallel pore stacks in high frequency prime movers. A more complete analysis is necessary for small refrigerators since enhanced viscosity may affect the performance.

ACKNOWLEDGMENTS

We would like to thank L. A. Wilen for providing measurement data of the thermoviscous functions for RVC and aluminum foam. This work was supported by the Department of Defense, U. S. Army Space and Missile Defense Command.

- ¹O. G. Symko, E. Abdel-Rahman, Y. S. Kwon, M. Emmi, and R. Behunin, "Design and development of high-frequency thermoacoustic engines for thermal management in microelectronics," *Microelectron. J.* **35**, 185–191 (2004).
- ²H. El-Gendy, Y. S. Kwon, and O. G. Symko, "High frequency operation of thermoacoustic coolers and prime movers," *J. Acoust. Soc. Am.* **117**, 2533 (2005).
- ³N. Rott, "Thermoacoustics," *Adv. Appl. Mech.* **20**, 135–175 (1980).
- ⁴G. W. Swift, "Thermoacoustic engines," *J. Acoust. Soc. Am.* **84**, 1145–1180 (1988).
- ⁵W. P. Arnott, H. E. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections," *J. Acoust. Soc. Am.* **90**, 3228–3237 (1991).
- ⁶J. W. Strutt (Lord Rayleigh), *Theory of Sound* (Dover, New York, 1945).
- ⁷C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
- ⁸M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid," *J. Acoust. Soc. Am.* **28**, 168–191 (1956).
- ⁹H. Tijdeman, "On the propagation of sound waves in cylindrical tubes," *J. Sound Vib.* **39**, 1–33 (1975).
- ¹⁰K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular materials," *J. Acoust. Soc. Am.* **73**, 785–799 (1983).
- ¹¹H. S. Roh, W. P. Arnott, J. M. Sabatier, and R. Raspet, "Measurement and calculation of acoustic propagation constant in arrays of small air-filled rectangular tubes," *J. Acoust. Soc. Am.* **89**, 2617–2624 (1991).
- ¹²M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape," *J. Acoust. Soc. Am.* **89**, 550–558 (1991).
- ¹³J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier Applied Science, London, 1993).
- ¹⁴K. Attenborough, "Models for the acoustic properties of air-saturated granular media," *Acta Acust. (Beijing)* **1**, 213–226 (1993).
- ¹⁵P. C. Carman, *Flow of Gases through Porous Media* (Academic, New York, 1956).
- ¹⁶R. Raspet, W. V. Slaton, C. J. Hickey, and R. H. Hiller, "Theory of inert gas-condensing vapor thermoacoustics: Propagation equation," *J. Acoust. Soc. Am.* **112**, 1414–1422 (2002).
- ¹⁷L. A. Wilen, "Dynamic measurements of the thermal dissipation function of reticulated vitreous carbon," *J. Acoust. Soc. Am.* **109**, 179–184 (2001).

- ¹⁸A. Petculescu and L. A. Wilen, "Lumped-element technique for the measurement of complex density," *J. Acoust. Soc. Am.* **110**, 1950–1957 (2001).
- ¹⁹Y. Champoux and M. R. Stinson, "On acoustical models for sound propagation in rigid frame porous materials and the influence of shape factors," *J. Acoust. Soc. Am.* **92**, 1120–1131 (1992).
- ²⁰R. Raspet, J. Brewster, and H. E. Bass, "A new approximation method for thermoacoustic calculations," *J. Acoust. Soc. Am.* **103**, 2395–2402 (1998).
- ²¹M. E. H. Tijani, J. C. H. Zeegers, and A. T. A. M. de Waele, "Prandtl number and thermoacoustic refrigerators," *J. Acoust. Soc. Am.* **112**, 134–143 (2002).

Near-field ultrasound tomography

Serge Mensah^{a)} and Emilie Franceschini

CNRS-Laboratoire de Mécanique et d'Acoustique, 31 chemin Joseph Aiguier, 13402 Marseille, Cedex 20, France

(Received 6 March 2006; revised 22 December 2006; accepted 2 January 2006)

In this paper, a near-field tomographic solution is introduced to solve the imaging problem of fluid objects assumed to be weakly heterogeneous (Born approximation) and excited by spherical waves. The solution to the forward problem is based on the Huygens-Fresnel principle which describes the scattered field as the result of the interference scheme of all the secondary spherical waves. From the derivation of the scattered field, a new Fourier transform that has been called the elliptical Fourier transform is defined: It differs from the standard Fourier transform in that instead of a plane wave decomposition, a harmonic ellipsoidal wave decomposition is obtained. Based on this spectral analysis, a near-field Radon transform is designed that complements the “far-field tools” published in diffraction tomography literature. Then, assuming that the measuring distance is greater than one wavelength, the feasibility of reconstructing either the impedance or the velocity maps of an acoustical (perfect fluid) model is demonstrated. Numerical simulations were performed which confirmed the validity of the theory presented here; a theory which has many potential applications in future wave theory research. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436637]

PACS number(s): 43.35.Wa, 43.80.Ev, 43.80.Vj [TDM]

Pages: 1423–1433

I. INTRODUCTION

The use of wide-band information is becoming a common goal for most research teams working on the ultrasonic imaging of weak-scattering media, since it is known to improve point and contrast resolutions, both of which are important criteria for medical applications. Developing quantitative reconstruction procedures is the ultimate goal, so as to improve the specificity of the system as well as performing automatic tissue characterization. Finally, in order to obtain an efficient reconstruction procedure it is important to take into account the true nature of the wave (cylindrical or spherical waves; transducer directivity), which is generally expressed in terms of near-field conditions. So far, there exist very few approaches which meet all the above-mentioned criteria simultaneously. In his pioneering studies, Norton^{1,2} introduced the near-field concept, which makes it possible to deal with spherical waves transmitted around the whole organ as the punctual broadband transducer moves along a circular path. The resulting projections obtained were found to result from surface integration of the reflectivity over the spherical wave fronts. However, the fundamental hypothesis was based on a circular (or spherical) harmonic decomposition of the object, which makes it necessary to use Hankel transforms. These transforms are numerically very unstable, and further approximations are therefore generally required for practical implementation. The second limitation of this pioneering work is the fact that it involves a single transducer acting as both a transmitter and a receiver in the pure reflection mode. Furthermore, diffraction tomography methods have been implemented either with a number of discrete frequencies^{3–6} or with multiple frequencies in order to enhance the image resolution.⁷ Also, with quantitatively accu-

rate reconstruction techniques⁸ the imaging of object of higher contrast and bigger size than permitted by diffraction tomography (Born approximation) becomes possible. However, these linear and nonlinear solutions of the inverse problem have the great disadvantage of requiring a considerable amount of computational resources, especially in the typical case of quantitative imaging, where many frequencies and/or enhanced Green's operator are used to achieve a sufficiently high imaging quality to compete with the diagnostic systems available at present. In the dual space, several researchers^{9–11} have developed direct time-domain scattering methods that overcome most of the limitations inherent in frequency-domain techniques. These methods were designed to obtain quantitative information by performing bistatic or multistatic measurements and attempting to work in the near-field zone. However, the plane wave assumption on which these methods were based is generally not compatible with the way in which present-day scanners function. Recently, some interesting time-domain reconstructions were carried out in the field of thermoacoustic tomography^{12,13} with a view to reconstructing the relative absorption coefficient of biological tissues based on cylindrical or spherical measurement. The prerequisite that the near-field distance must be much greater than the thermoacoustic signals simplifies the exact formulas, and makes it possible to obtain backprojections and weighted time-domain summations over cylindrical or spherical surfaces. Approaches of this kind implicitly make it necessary to invert a generalized (circular) Radon transform.¹⁴ However, here again, no multistatic configurations have so far been used in studies of this kind, since the ultrasound channels operate in a passive mode in these microwave-induced imaging systems.

Lastly, several researchers were interested at an early stage in quantitative reconstruction, and attempted to use the scattering directivity pattern of any acoustical param-

^{a)}Electronic mail: mensah@lma.cnrs-mrs.fr

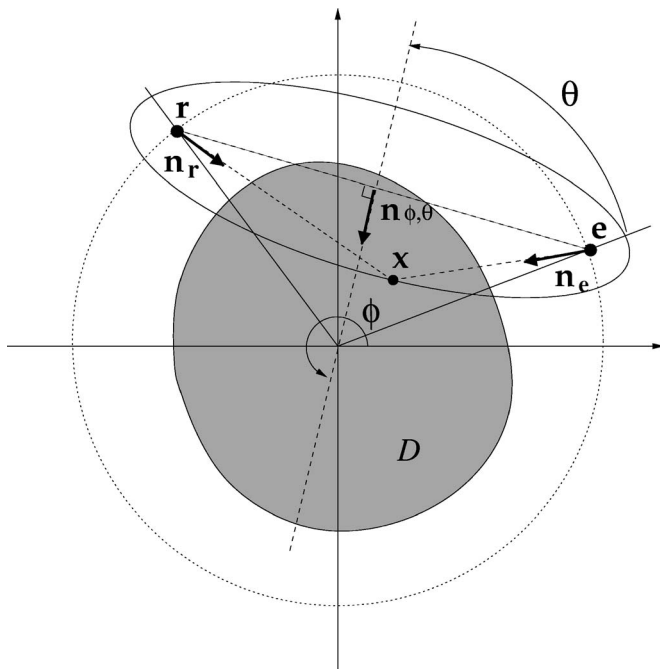


FIG. 1. Scattering acquisition configuration. In active ultrasonic tomography, a transducer \mathbf{e} (which size is smaller than half a wavelength) transmits a spherical wave which is scattered and recorded by a receiver \mathbf{r} placed on the same spherical surface closely surrounding the organ (i.e., the breast).

eters (attenuation and sound speed,¹⁵ density and compressibility,^{16–18} or impedance and sound speed^{19–21}) for this purpose. In addition to these investigations, it is proposed to extend the scope of the research to include near-field conditions (using spherical waves), with broadband signals, the scattering analysis derived using plane wave decomposition methods. Depending on whether the transmission or the reflection mode is involved, this will enable us to separate the respective effects of tissue parameter distributions on the scattering processes. From this analysis, which is valid provided that the observation distance is greater than the wavelength, a two-dimensional (2D) reconstruction scheme was developed, providing an estimate of either impedance or sound speed distributions. The ideal propagative medium is assumed to effect no absorption and to have only weak scattering characteristics. The present reconstruction technique is a near-field extension of the inverse Radon transform: this solution of the linearized scattered inverse problem exploits an optimal fast Fourier transform algorithm.

II. DEFINITION OF THE ELLIPTICAL FOURIER TRANSFORMS

Let us define the compact domain \mathcal{D} of \mathbb{R}^N $N=2,3$, corresponding to the organ to be imaged and the acquisition sphere (or circle, in the 2D case) \mathcal{S} surrounding the compact domain \mathcal{D} . The acquisition sphere radius \mathcal{R} is such that $\forall x \in \mathcal{D}, \|x\| < \mathcal{R}$. A transducer located in \mathbf{e} , $\mathbf{e} \in \mathcal{S}$, transmits a spherical wave with an angular frequency ω . The field scattered by the medium is recorded in \mathbf{r} , $\mathbf{r} \in \mathcal{S}$.

We take $\mathbf{n}_{\phi, \theta}$ to denote the unit vector of the mediatrix in the incident direction specified by the angles (Fig. 1):

$$\phi = \pi + \frac{1}{2}(\text{angle}(\mathbf{e}) + \text{angle}(\mathbf{r})) \quad \text{the incident angle,}$$

$\theta = \frac{1}{2}|\text{angle}(\mathbf{e}) - \text{angle}(\mathbf{r})|$ the configuration angle in the plane (\mathbf{e}, \mathbf{r}) .

The elliptical wave vector $\mathbf{K} = k\mathbf{n}_{\phi, \theta}$ is then defined. The elliptical scalar product defined for a fixed configuration (\mathbf{e}, \mathbf{r}) or equivalently (ϕ, θ) is introduced:

$$\mathbf{K} \odot \mathbf{x} = k\mathbf{n}_{\phi, \theta} \odot \mathbf{x} \triangleq -k(\|\mathbf{x} - \mathbf{e}_{\mathbf{K}}\| + \|\mathbf{x} - \mathbf{r}_{\mathbf{K}}\|), \quad (1)$$

where $k = \omega/c_0$ is the wave number of the illuminating wave having the sound speed c_0 in the host medium.

Note 1: The negative value of the elliptical scalar product results here from the fact that the wave is propagating toward the origin, and is opposed to the outgoing orientation of the \mathbf{x} axis.

Note 2: According to the definition of an ellipse (ellipsoid), the scalar product verifies: $|\mathbf{K} \odot \mathbf{x}| > k\|\mathbf{r}_{\mathbf{K}} - \mathbf{e}_{\mathbf{K}}\|$ or $\theta < \pi$. Thus, pure transmission measurement appears as a limiting case that can be approximated with any arbitrary precision.

Let $f(\mathbf{x})$ be a well-behaved function defined over \mathcal{D} and null elsewhere. The *elliptical Fourier transform* $\mathcal{F}_{\mathcal{E}}$ is defined by

$$\mathcal{F}_{\mathcal{E}}(f(\mathbf{x})) = \check{f}(\mathbf{K}) = \int f(\mathbf{x}) \frac{e^{-i\mathbf{K} \odot \mathbf{x}}}{[\mathbf{x}]_{\mathbf{K}}} d\mathbf{x} = \int f(\mathbf{x}) \varepsilon^{-i\mathbf{K} \odot \mathbf{x}} d\mathbf{x}. \quad (2)$$

In the last integral, the harmonic decomposition function

$$\varepsilon^{-i\mathbf{K} \odot \mathbf{x}} = \frac{e^{-i\mathbf{K} \odot \mathbf{x}}}{[\mathbf{x}]_{\mathbf{K}}} \quad (3)$$

corresponds to the ellipsoidal harmonic function reduced by the distance $[\mathbf{x}]_{\mathbf{K}}$:

$$[\mathbf{x}]_{\mathbf{K}} = 16\pi^2 \|\mathbf{x} - \mathbf{e}_{\mathbf{K}}\| \cdot \|\mathbf{x} - \mathbf{r}_{\mathbf{K}}\|. \quad (4)$$

The *inverse Elliptical Fourier transform* is then defined as follows:

$$\mathcal{F}_{\mathcal{E}}^{-1}(\check{f}(\mathbf{K})) = f(\mathbf{x}) = \frac{1}{(2\pi)^N} \int \check{f}(\mathbf{K}) [\mathbf{x}]_{\mathbf{K}} e^{i\mathbf{K} \odot \mathbf{x}} d\mathbf{K}, \quad (5)$$

and is written

$$f(\mathbf{x}) = \frac{1}{(2\pi)^N} \int \check{f}(\mathbf{K}) \varepsilon^{i\mathbf{K} \odot \mathbf{x}} d\mathbf{K}. \quad (6)$$

Indeed,

$$\begin{aligned} \mathcal{F}_{\mathcal{E}}^{-1}(\check{f}(\mathbf{K})) &= \frac{1}{(2\pi)^N} \int \check{f}(\mathbf{K}) \varepsilon^{i\mathbf{K} \odot \mathbf{x}} d\mathbf{K} \\ &= \frac{1}{(2\pi)^N} \int \left(\int f(\mathbf{x}') \varepsilon^{-i\mathbf{K} \odot \mathbf{x}'} d\mathbf{x}' \right) \varepsilon^{i\mathbf{K} \odot \mathbf{x}} d\mathbf{K} \\ &= \int f(\mathbf{x}') \frac{1}{(2\pi)^N} \int \varepsilon^{i\mathbf{K} \odot (\mathbf{x} \ominus \mathbf{x}')} d\mathbf{K} d\mathbf{x}', \end{aligned} \quad (7)$$

where we note $\mathbf{K} \odot (\mathbf{x}' \ominus \mathbf{x}) = \mathbf{K} \odot \mathbf{x}' - \mathbf{K} \odot \mathbf{x}$, since the elliptical scalar product “ \odot ” is not distributive. The last integral is the Dirac distribution (Appendix A 1):

$$\frac{1}{(2\pi)^N} \int e^{i\mathbf{K}\odot(\mathbf{x}\oplus\mathbf{x}')} d\mathbf{K} = \frac{1}{(2\pi)^N} \int \frac{[\mathbf{x}]_{\mathbf{K}}}{[\mathbf{x}']_{\mathbf{K}}} e^{i\mathbf{K}\odot(\mathbf{x}\oplus\mathbf{x}')} d\mathbf{K} = \delta(\mathbf{x} - \mathbf{x}') \quad (8)$$

then, in the sense of distributions:

$$\mathcal{F}_{\mathcal{E}}^{-1}(\check{f}(\mathbf{K})) = \int f(\mathbf{x}') \delta(\mathbf{x} - \mathbf{x}') d\mathbf{x}' = f(\mathbf{x}). \quad (9)$$

The elliptical Fourier transform is an extension of the classical Fourier transform applied to near-field diffraction measurement.

III. FORWARD PROBLEM

We neglect the absorption-dispersion phenomena in order to concentrate on the diffusion resulting from the inhomogeneities. For practical applications of these results, a time gain compensation, TGC, should possibly be used. The inhomogeneous media are described in terms of their density ρ and velocity c distributions. The acoustic pressure p is governed by the wave equation:

$$-\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} + \rho \nabla \left(\frac{1}{\rho} \nabla p \right) = 0. \quad (10)$$

Let ρ_0, c_0 be the acoustic characteristics of the surrounding medium. The propagation equation can be written as follows:

$$-\frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} + \Delta p = \left(\frac{1}{c^2} - \frac{1}{c_0^2} \right) \frac{\partial^2 p}{\partial t^2} + \frac{\nabla \rho}{\rho} \nabla p. \quad (11)$$

We introduce the parameters $2\alpha = (c^2 - c_0^2)/c^2$ and $\xi = \text{Log}(z/z_0)$ corresponding to the quadratic fluctuations of velocity c and to the logarithmic variations of impedance $z = \rho c$, respectively. In view of the development:

$$\frac{\nabla \rho}{\rho} = \nabla \left(\text{Log} \frac{z}{z_0} \right) + \frac{1}{2} \nabla (\text{Log}(1 - 2\alpha)), \quad (12)$$

and the fact that the biological tissues are weakly inhomogeneous media ($\alpha \sim \xi \sim 10^{-2}$), we obtain, at order one in α , the following impedance-velocity propagation equation:²¹

$$-\frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} + \Delta p = -\frac{2\alpha}{c_0^2} \frac{\partial^2 p}{\partial t^2} - \nabla \alpha \nabla p + \nabla \xi \nabla p. \quad (13)$$

The medium is excited with a spherical wave having angular frequency ω , the emitter being located in \mathbf{e} :

$$P_i(\mathbf{x}, \mathbf{e}, \omega) = \frac{e^{ik\|\mathbf{x}-\mathbf{e}\|}}{4\pi\|\mathbf{x}-\mathbf{e}\|}. \quad (14)$$

The solution of Eq. (13) is $p = P e^{-i\omega t}$. P is the solution of the Helmholtz equation which can be written in its integral form:

$$P(\mathbf{r}, \mathbf{e}, \omega) = P_i(\mathbf{r}, \mathbf{e}, \omega) + \int_{\mathcal{D}} G(\mathbf{r}, \mathbf{x}, \omega) [2k^2 \alpha(\mathbf{x}) P(\mathbf{x}, \mathbf{e}, \omega) - \nabla \alpha(\mathbf{x}) \cdot \nabla P(\mathbf{x}, \mathbf{e}, \omega) + \nabla \xi(\mathbf{x}) \cdot \nabla P(\mathbf{x}, \mathbf{e}, \omega)] d\mathbf{x}, \quad (15)$$

where the integration volume \mathcal{D} corresponds to the compact

support of α and ξ , and G is the free space Green function; the scattered field is measured in \mathbf{r} :

$$G(\mathbf{r}, \mathbf{x}, \omega) = -\frac{e^{ik\|\mathbf{r}-\mathbf{x}\|}}{4\pi\|\mathbf{r}-\mathbf{x}\|}. \quad (16)$$

Throughout this paper, it is assumed that the acoustical parameters are described by ‘‘well-behaved’’ functions defined over \mathcal{D} and null elsewhere. This means in particular that they have a finite number of discontinuities and turning points (maxima and minima), between which they are monotonic and continuous. Eventually, it also means that they may become infinite, provided that $\int f(\mathbf{x}) d\mathbf{x}$ converges absolutely. Furthermore, we assume that the scattering is sufficiently weak to be able to linearize the inverse problem (using the Born approximation) so that the scattered field P_d becomes

$$P_d = \int G\{2k^2 \alpha P_i - \nabla \alpha \nabla P_i + \nabla \xi \nabla P_i\} d\mathbf{x}. \quad (17)$$

Differentiating P_i yields:

$$\nabla P_i(\mathbf{x}, \mathbf{e}, \omega) = \left(ik - \frac{1}{\|\mathbf{x}-\mathbf{e}\|} \right) \nabla (\|\mathbf{x}-\mathbf{e}\|) \frac{e^{ik\|\mathbf{x}-\mathbf{e}\|}}{4\pi\|\mathbf{x}-\mathbf{e}\|}. \quad (18)$$

In Eq. (18), $1/\|\mathbf{x}-\mathbf{e}\|$ is negligible in comparison with ik . Indeed, in soft tissues, $c_0 \approx 1540 \text{ m s}^{-1}$; and in the near-field zone of interest, $\|\mathbf{x}-\mathbf{e}\| \approx 10^{-2} \text{ m}$. The angular frequency components of the transmitted pulse (based on current echographs) are generally larger than $\omega_{\text{ref}} = 2\pi \cdot 2.5 \times 10^6 \text{ rad s}^{-1}$; thus $k\|\mathbf{x}-\mathbf{e}\| \approx 10^2 \text{ rad}$. The observation distance is greater than the wavelength.

We define \mathbf{n}_e (respectively, \mathbf{n}_r) as the unit vector along the emitter (respectively, receiver) -voxel line that we call the $\mathbf{e}-\mathbf{x}$ axis (respectively, the $\mathbf{x}-\mathbf{r}$ axis): $\mathbf{n}_e = \nabla(\|\mathbf{x}-\mathbf{e}\|)$. This gives

$$\nabla P_i(\mathbf{x}, \mathbf{e}, \omega) \simeq ik \mathbf{n}_e \frac{e^{ik\|\mathbf{x}-\mathbf{e}\|}}{4\pi\|\mathbf{x}-\mathbf{e}\|}, \quad (19)$$

and Eq. (17) becomes

$$P_d = - \int \frac{e^{ik(\|\mathbf{r}-\mathbf{x}\| + \|\mathbf{x}-\mathbf{e}\|)}}{16\pi^2 \|\mathbf{r}-\mathbf{x}\| \|\mathbf{x}-\mathbf{e}\|} \times \{2k^2 \alpha - ik \mathbf{n}_e \cdot [\nabla \alpha - \nabla \xi]\} d\mathbf{x}. \quad (20)$$

According to the definition of the elliptical Fourier transform in Sec. II, the scattered field Eq. (20) is

$$P_d = - \int_{\mathcal{D}} \varepsilon^{-i\mathbf{K}\odot\mathbf{x}} \{2k^2 \alpha - ik \mathbf{n}_e \cdot [\nabla \alpha - \nabla \xi]\} d\mathbf{x}, \quad (21)$$

and P_d then reduces to

$$P_d(\mathbf{K}) = - \mathcal{F}_{\mathcal{E}}\{2k^2 \alpha - ik \mathbf{n}_e \cdot [\nabla \alpha - \nabla \xi]\}(\mathbf{K})|_{\mathbf{K}=k\mathbf{n}_{\phi, \theta}}. \quad (22)$$

When taking $\check{\mathcal{K}}_{\mathbf{K}} = \mathcal{F}_{\mathcal{E}}[-ik(1 + \mathbf{n}_{\mathbf{e}_{\mathbf{K}}} \cdot \mathbf{n}_{\mathbf{r}_{\mathbf{K}}})]$, the derivation rule introduced in Appendix A 2 yields

$$P_d(\mathbf{K}) \simeq - \{2k^2 \check{\alpha}(\mathbf{K}) - ik[\check{\mathcal{K}}_{\mathbf{K}} \circledast (\check{\alpha} - \check{\xi})](\mathbf{K})\}, \quad (23)$$

where ‘‘ \circledast ’’ is the elliptical convolution (cf Appendix A 2). We note

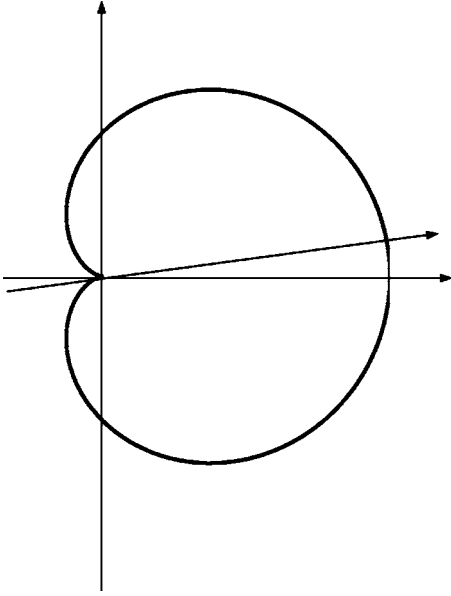


FIG. 2. Directivity diagrams of α : The velocity effects predominate in the transmission mode.

$$P_d(\mathbf{e}, \mathbf{r}, \omega) = -h(\mathbf{e}, \mathbf{r}, \omega), \quad (24)$$

then

$$h(\mathbf{e}, \mathbf{r}, \omega) = k^2 [\mathcal{F}_{\mathcal{E}}(1 - \mathbf{n}_e \cdot \mathbf{n}_r) \otimes \check{\alpha} + \mathcal{F}_{\mathcal{E}}(1 + \mathbf{n}_e \cdot \mathbf{n}_r) \otimes \check{\xi}](\mathbf{K}), \quad (25)$$

where $\check{\alpha}(\mathbf{K})$ and $\check{\xi}(\mathbf{K})$ are the elliptical Fourier transforms of the functions $\alpha(\mathbf{x})$ and $\xi(\mathbf{x})$:

$$\begin{aligned} \check{\alpha}(\mathbf{K}) &= \int \alpha(\mathbf{x}) \varepsilon^{-i\mathbf{K} \odot \mathbf{x}} d\mathbf{x}, \\ \check{\xi}(\mathbf{K}) &= \int \xi(\mathbf{x}) \varepsilon^{-i\mathbf{K} \odot \mathbf{x}} d\mathbf{x}. \end{aligned} \quad (26)$$

h represents the elliptical spatial transfer function of the medium in the case of a specific configuration defined by \mathbf{e} and \mathbf{r} . This configuration involves two parameters offering opposite local directivity diagrams: the velocity predominates in the transmission mode (Fig. 2); whereas the impedance predominates in the reflection mode (Fig. 3). This composite object results in a summation of the elementary contributions integrated over the whole object volume, weighted by their local directivity indexes. When $\mathbf{n}_e = \mathbf{n}_r$, i.e., in the pure reflection mode,

$$h(\mathbf{e}, \mathbf{r}, \omega) = 2k^2 \check{\xi}(\mathbf{K})|_{\mathbf{K}=k\mathbf{n}_{\phi,0}}. \quad (27)$$

When $\mathbf{n}_e \simeq -\mathbf{n}_r$, i.e., in the pure transmission mode,

$$h(\mathbf{e}, \mathbf{r}, \omega) = 2k^2 \check{\alpha}(\mathbf{K})|_{\mathbf{K}=k\mathbf{n}_{\phi,\pi}}. \quad (28)$$

At this stage, we have derived a practical means of separating the respective contributions of the impedance fluctuations and those of the sound speed fluctuations to the near-field data. The simplest way of effecting this separation is to work in either the transmission or the reflection mode.

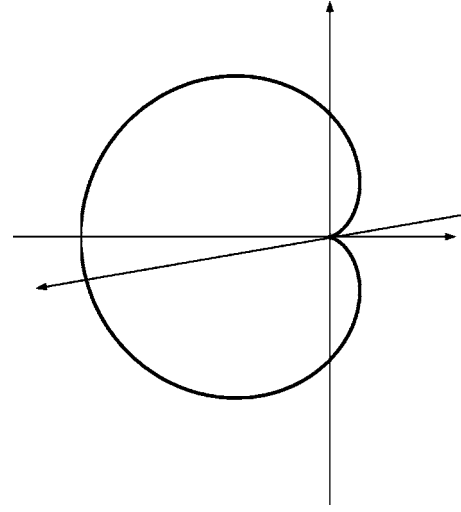


FIG. 3. Directivity diagrams of ξ : The impedance fluctuations account mainly for the scattering in the reflection mode.

IV. INVERSE PROBLEM

By applying the inverse elliptical Fourier transform to the elliptical spectra, we are able to reconstruct either the impedance or the sound speed map. For this purpose, the acquisition procedure used consists of filling in the respective elliptical Fourier domains with the highest sampling densities by performing both a spatial scan and a frequency sweep using broadband signals. However, in practice, since the ultrasonic transducers have a finite bandwidth, one can expect to be able to reconstruct only a bandpass (respectively, low-pass) filtered version of the impedance (respectively, sound speed). Besides, rather than using multidimensional inverse elliptical Fourier transforms to retrieve the object from diffraction measures, it is better to use a reconstruction procedure based on classical one-dimensional (1D) Fourier transforms, with which so many signal processing devices have been optimized. We therefore defined a near-field extension of the Radon transform $\mathcal{R}_{\mathcal{E}}$, thus restricting our derivation to the 2D case.

A. The 2D elliptical Radon transform and the elliptical Fourier projection-slice theorem

Let $f(\mathbf{x})$ be a well-behaved object function defined over the compact domain \mathcal{D} of \mathbb{R}^2 . f represents either the quadratic fluctuation of the velocity α or the logarithmic variation of the impedance ξ . We may define the elliptical Radon transform by

$$\begin{aligned} [\mathcal{R}_{\mathcal{E}}f](s, \phi, \theta) &= \int f(\mathbf{x}) \frac{\delta(s - \|\mathbf{x} - \mathbf{e}\| + \|\mathbf{r} - \mathbf{x}\|)}{[\mathbf{x}]_{\phi, \theta}} d\mathbf{x} \\ &= \int f(\mathbf{x}) d\sigma_{\phi, \theta}(\mathbf{x}), \end{aligned} \quad (29)$$

where $d\sigma_{\phi, \theta}(\mathbf{x})$ is the ellipse $s = \|\mathbf{x} - \mathbf{e}\| + \|\mathbf{r} - \mathbf{x}\|$ of foci $(\mathbf{e}, \mathbf{r}) \in \mathcal{S}^2$ (weighted by $[\mathbf{x}]_{\phi, \theta}$). It is clear, however, that this mapping is overdetermined, since the dimension of triplet (s, ϕ, θ) is $N+1$, while the function f depends on $N=2$ variables only. Thus, from the outset, we fix the configuration

angle $\theta = \theta_0$ at a constant value, we therefore drop this subscript from the following expressions.

The acquisition configuration allows us to obtain one of the 1D projections, $p_\phi(s)$, under the incidence ϕ with $\phi \in [0, \pi]$ that belongs to the elliptical Radon transform of the object function f :

$$[\mathcal{R}_{\mathcal{E}f}](s, \phi) = \{p_\phi(s), s \in \mathbb{R}, 0 \leq \phi \leq \pi\}, \quad (30)$$

$$[\mathcal{R}_{\mathcal{E}f}](s, \phi)|_{\phi \text{ fixed}} = p_\phi(s). \quad (31)$$

The standard Fourier transform F of a projection $p_\phi(s)$, $\hat{p}_\phi(S)$, is written:

$$\hat{p}_\phi(S) = (F[p_\phi])(S) = \int_{-\infty}^{\infty} p_\phi(s) e^{-iSs} ds. \quad (32)$$

Substituting $p_\phi(s)$ by its expression [Eq. (29)], implies that:

$$\begin{aligned} \hat{p}_\phi(S) &= \int \int f(\mathbf{x}) \frac{\delta(s - \mathbf{n}_\phi \odot \mathbf{x})}{[\mathbf{x}]_\phi} e^{-iSs} d\mathbf{x} ds \\ &= \int f(\mathbf{x}) \int \frac{\delta(s - \mathbf{n}_\phi \odot \mathbf{x})}{[\mathbf{x}]_\phi} e^{-iS \mathbf{n}_\phi \odot \mathbf{x}} ds d\mathbf{x}. \end{aligned} \quad (33)$$

We denote $S \mathbf{n}_\phi = \mathbf{S}_\phi$, then

$$\hat{p}_\phi(S) = \int f(\mathbf{x}) \varepsilon^{-iS \mathbf{n}_\phi \odot \mathbf{x}} d\mathbf{x} = \check{f}(S, \phi)|_{\phi \text{ fixed}} = \mathcal{F}_{\mathcal{E}|\phi}[f](S). \quad (34)$$

The standard Fourier transform of a projection is therefore a slice of the 2D elliptical Fourier transform of the object function. This result can be said to be a near-field extension of the Fourier projection-slice theorem.

B. Inversion procedure

The inverse of the elliptical Radon transform is derived in Appendix B. Here, the focus is on the inversion procedure: Let us take a two-dimensional function f , and apply the direct and inverse elliptical Fourier transforms and use polar coordinates to obtain:

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \check{f}(\mathbf{K}) e^{i\mathbf{K} \odot \mathbf{x}} [\mathbf{x}]_{\mathbf{K}} d\mathbf{K} \\ &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{\infty} \mathcal{F}_{\mathcal{E}}[f](S, \phi) e^{iS \mathbf{n}_\phi \odot \mathbf{x}} [\mathbf{x}]_\phi S dS d\phi \\ &= \frac{1}{(2\pi)^2} \int_0^{\pi} \int_{-\infty}^{\infty} \mathcal{F}_{\mathcal{E}|\phi}[f](S) |S| e^{iS \mathbf{n}_\phi \odot \mathbf{x}} [\mathbf{x}]_\phi dS d\phi \\ &= \frac{1}{(2\pi)^2} \int_0^{\pi} \int_{-\infty}^{\infty} \hat{p}_\phi(S) |S| e^{iS \mathbf{n}_\phi \odot \mathbf{x}} [\mathbf{x}]_\phi dS d\phi. \end{aligned} \quad (35)$$

The inversion scheme is as follows:

$$f(\mathbf{x}) = \frac{1}{2} \times \frac{1}{\pi} \int_0^{\pi} \Pi_\phi(\mathbf{x}) d\phi \quad (36)$$

is the sum of the Elliptical backprojections $\Pi_\phi(\mathbf{x})$:

$$\begin{aligned} \Pi_\phi(\mathbf{x}) &= \frac{1}{2\pi} \int \hat{p}_\phi(S) |S| e^{iS \mathbf{n}_\phi \odot \mathbf{x}} [\mathbf{x}]_\phi dS \\ &= \frac{[\mathbf{x}]_\phi}{2\pi} \int \int \hat{p}_\phi(S) |S| e^{iSs} \delta(s - \mathbf{n}_\phi \odot \mathbf{x}) dS ds \\ &= [\mathbf{x}]_\phi p_\phi^F(s = \mathbf{n}_\phi \odot \mathbf{x}) \end{aligned} \quad (37)$$

of the filtered projections p_ϕ^F :

$$\begin{aligned} p_\phi^F(s) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{p}_\phi(S) |S| e^{iSs} dS \\ &= F^{-1}[|S| \hat{p}_\phi(S)] \\ &= F^{-1}[\mathbf{K} |\hat{p}_\phi(\mathbf{K})]. \end{aligned} \quad (38)$$

This inversion algorithm turns out to be quite similar to the classical backprojection of the filtered projection algorithm used so far in computer-assisted tomography; only the nature of the projections differs.

C. Backprojection operator

Consider an arbitrary function $h(s, \phi)$ where $s = \mathbf{n}_\phi \odot \mathbf{x}$. The backprojection operator \mathcal{B} is defined by²²:

$$\mathcal{B}[h(s, \phi)](\mathbf{x}) = \frac{1}{2} \times \frac{1}{\pi} \int_0^{\pi} [\mathbf{x}]_\phi h(\mathbf{n}_\phi \odot \mathbf{x}, \phi) d\phi. \quad (39)$$

It is helpful to introduce the modified projection function f^* as follows:

$$f^*(s, \phi) = F^{-1}[|S| F[\mathcal{R}_{\mathcal{E}f}](S)], \quad (40)$$

the function f is then recovered by backprojection, (cf. Appendix B)

$$f(\mathbf{x}) = \mathcal{B}f^* = \frac{1}{2} \times \frac{1}{\pi} \int_0^{\pi} [\mathbf{x}]_\phi f^*(\mathbf{n}_\phi \odot \mathbf{x}, \phi) d\phi, \quad (41)$$

or

$$f = \mathcal{B} \circ F^{-1} \circ \text{Abs} \circ F \circ \mathcal{R}_{\mathcal{E}} \circ f, \quad (42)$$

where the filter Abs has the response $\text{Abs}(S) = |S|$.

V. NUMERICAL SIMULATIONS

In order to assess the reconstruction procedure presented in Sec. IV, we use a 2D numerical tissue-like phantom giving a scattering response which is computed using a finite element method (FEM). This method, which was developed in Refs. 23 and 24, models the time-domain acoustic wave propagation occurring in fluid media and is based on the discretization of the mixed velocity-pressure formulation of acoustics. One of the advantages of this method is that it requires no physical approximations to be made in the framework of linear acoustics; our method thus automatically accounts for multiple scattering, refraction, and reflection. The space discretization of the problem is based on a mixed finite element method²⁴ and the discretization in time is performed using a second-order-centered finite difference scheme. The simulation grid is surrounded by an absorbing layer (PML) simulating unbounded domains.²⁵

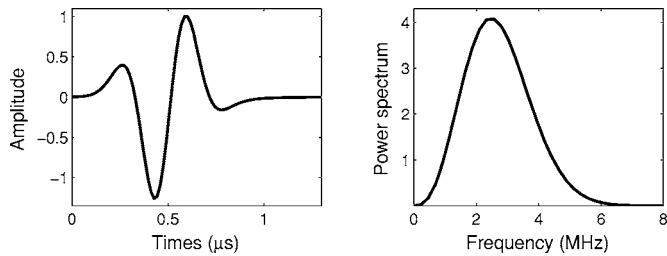


FIG. 4. 2.5 MHz cylindrical wave used in the FEM simulations in the time and frequency domains.

The size of the spatial step in the finite element grid is taken to be one thirtieth of the wavelength. A grid consisting of 1000×1000 pixels ($\Delta x = 0.016$ mm, 1.6 cm \times 1.6 cm) is used here. The ring transducer array is composed of 360 equally spaced transducers (point-like transmitters and receivers, central frequency 2.5 MHz, $\lambda = 0.6$ mm) and has a radius of $R = 7.36$ mm. The diameter of the ring transducer is fixed in order to obtain an acceptable computation time (simulation of the forward problem). This results in an element spacing (pitch) of 0.21λ . In ultrasound research, recent ring transducer systems already use 0.38λ pitch arrays in order to refine the transmitted wave front and to reduce the image reconstruction noise.²⁶ In the simulation, each transducer transmits a short pulse, while the remaining elements act as receivers. The temporal and spectral plots of the transmitted pulses are shown in Fig. 4. The cylindrical academic phantom simulates holes evenly spaced along a spiral immersed in water (Fig. 5). The radii of the holes are $r_1 = r_{1'} = 1.32$ mm, $r_2 = r_{2'} = r_1/2 = 0.66$ mm, $r_3 = r_{3'} = r_1/4 = 0.33$ mm, and $r_4 = r_{4'} = r_1/8 = 0.165$ mm. Each member of a pair of holes having the same radius has either a fixed impedance contrast (and no sound speed contrast, Fig. 6) or a sound speed contrast (and no impedance contrast, Fig. 7). The acoustic properties of the phantom are described in Table I.

Figure 8 shows the impedance tomogram reconstructed with the elliptical backprojection (EBP) procedure when the data were acquired in the pure reflection mode. One can note the image's high-quality contrast and its high resolution: This

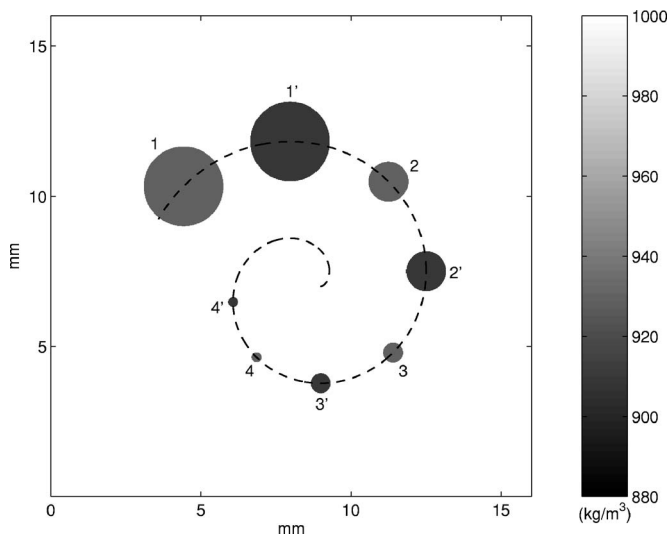


FIG. 5. Density map.

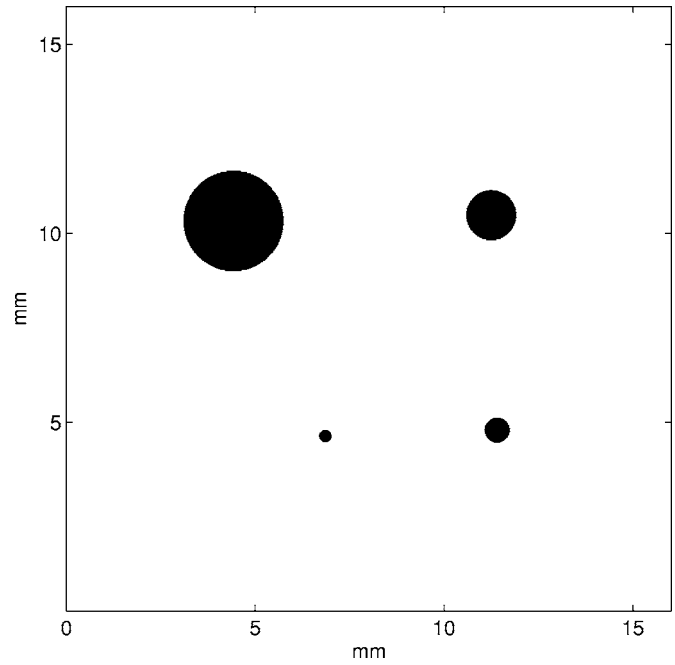


FIG. 6. The numerical phantom of Fig. 5 shows distinct contrast supports (location of the holes) of the acoustical parameters: the impedance map (surrounding medium 1.5 MRa; holes 1.395 MRa).

procedure makes it possible to detect objects of one half the wavelength. Figure 9 illustrates the phantom sound speed reconstructed via the EBP technique in the transmission mode. The time of flight (TOF) between the transmitter and the receiver was estimated using the first zero crossing method. When compared with the impedance reconstructed image, both the contrast and the resolution are poorer because of the low-pass filtering of the projections resulting from the TOF measurements. In the transmission mode, the

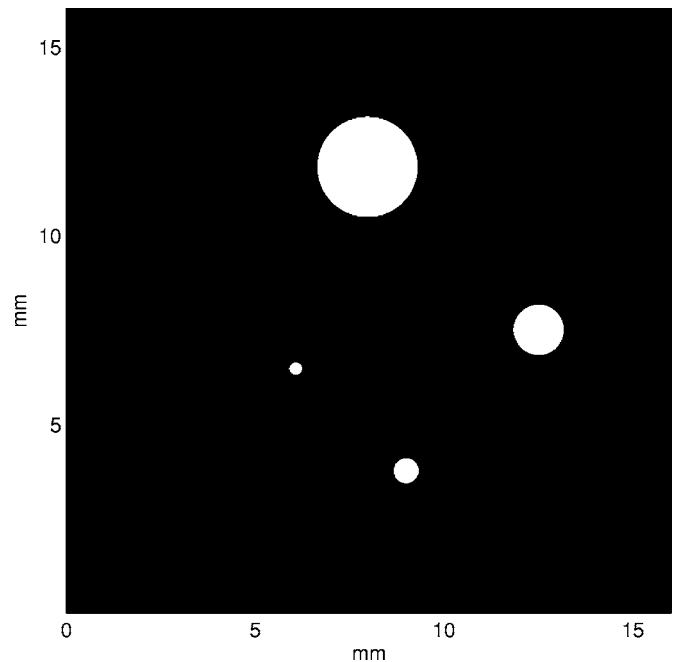


FIG. 7. The numerical phantom of Fig. 5 shows distinct contrast supports (location of the holes) of the acoustical parameters: the sound speed map (surrounding medium 1500 m/s; holes 1650 m/s).

TABLE I. Acoustical properties

Medium	Impedance (MRa)	Sound speed (m/s)	Density (kg/m ³)
Water	1.5	1500	1000
Holes 1, 2, 3,4	1.395	1500	930
Holes 1', 2', 3', 4'	1.5	1650	909.09

EBP procedure makes it possible to detect objects as large as the wavelength and to discriminate defects twice the size of the wavelength. Figure 10 shows the tomogram of the reconstructed objects using projections acquired in the diffraction mode ($\theta=40^\circ$). As the theoretical analysis indicated, one reconstructs a composite object combining the two acoustical parameters (namely impedance and sound speed), each of which is weighted by its own directivity function. Figures 11 and 12 show the poor image quality of the near-field reconstruction obtained when plane waves are transmitted within the medium and when the standard filtered backprojection (FBP) algorithm is used. The scatterers are badly localized: Instead of the radial distance from a point-like transducer, the projected distance from the plane source surface is estimated. With far-field measurements, the distortions are reduced since, locally, the cylindrical waves may be approximated by plane waves.

In conclusion, these numerical simulations confirm that this method makes it possible to separate either of the acoustical parameter contributions from the near-field data. In addition, the directivity functions constitute useful *a priori* information for the optimal processing of near-field scattering data.

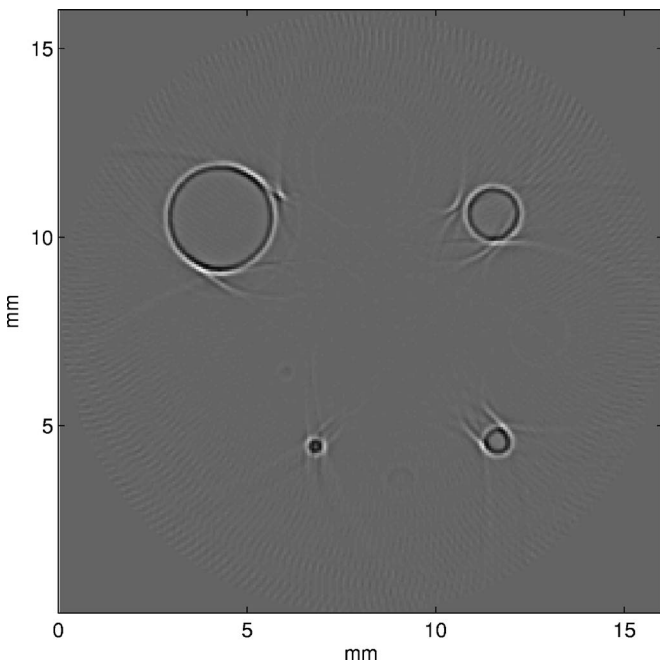


FIG. 8. EBP impedance reconstruction of the numerical phantom of Fig. 5. The acoustical parameters are separated according to the acquisition mode used: here pure reflection.

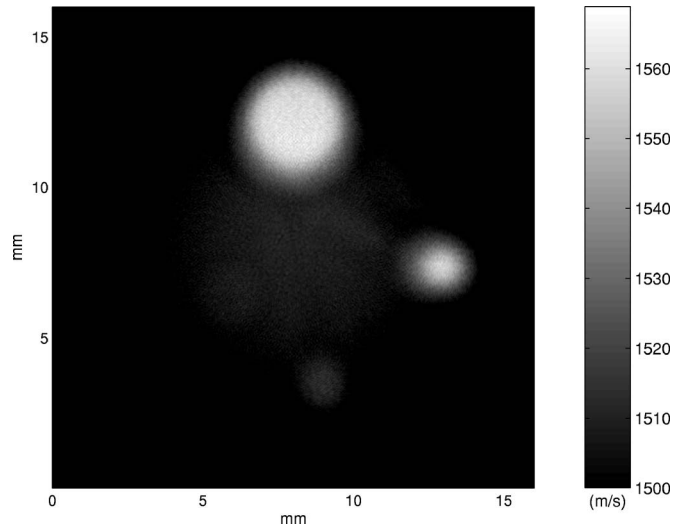


FIG. 9. EBP sound speed reconstruction of the numerical phantom of Fig. 5. The acoustical parameters are separated according to the acquisition mode used: here pure transmission.

VI. CONCLUSION

In quantitative (parametric) imaging and tissue characterization methods, the use of ultrasonic waves requires that the true nature of the waves be taken into consideration. Given the propagating distances generally employed, the shape of the actual wave front transmitted from an active probe element is practically spherical, and plane wave decomposition methods are therefore not optimal.

The derivation of the field scattered by a biological object (specified in terms of density and sound speed maps) shows three-dimensional (3D) ellipsoidal projections (depending on the time) on increasing ellipsoids; the foci of which coincide with the position of the transmitter and re-

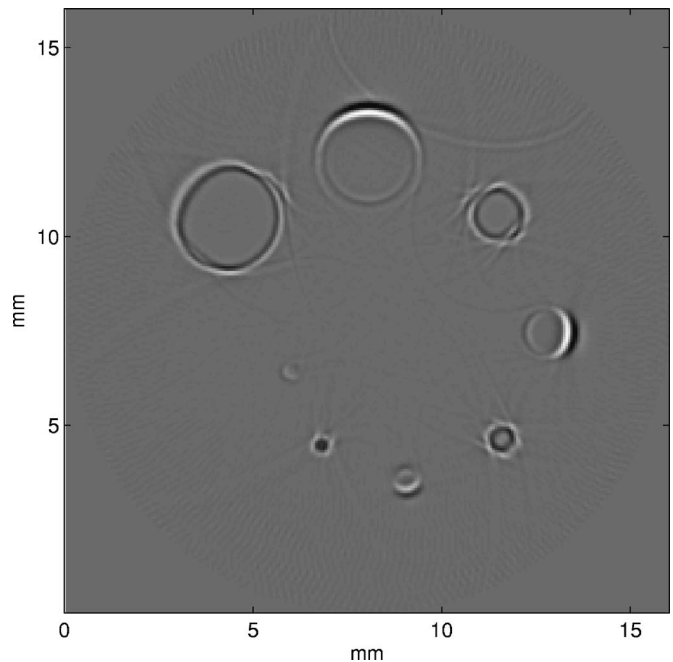


FIG. 10. Composite object reconstructed from diffraction measurements ($\theta=40^\circ$). The acoustical parameter contributions are combined.

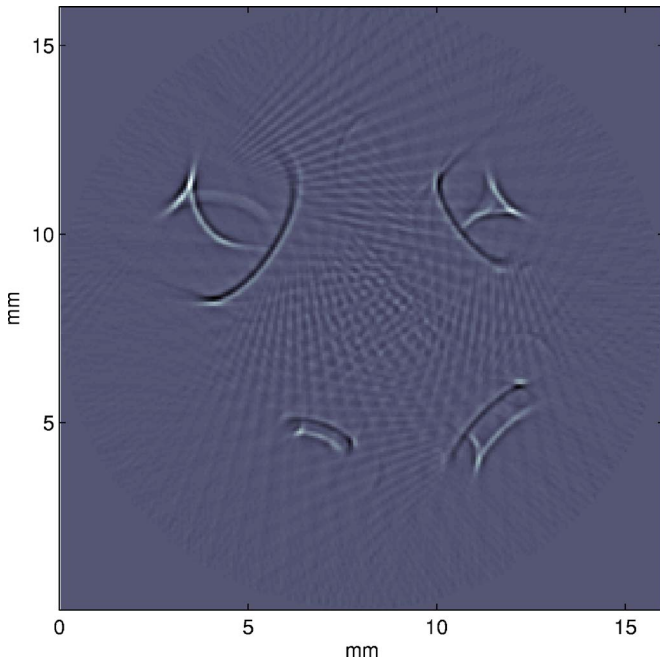


FIG. 11. Impedance reconstruction of the numerical phantom of Fig. 5: plane wave excitation (2.5 MHz central frequency), standard backprojection (FBP) algorithm in pure reflection.

ceiver. The amplitude of these projections is constrained by the fluctuations in impedance and in sound speed.

This ellipsoidal projection problem motivated the search for a harmonic ellipsoidal decomposition of the scattered field in line with the plane wave decomposition performed in Fourier spectral analysis. For this purpose, we defined an elliptical scalar product and an elliptical Fourier transform pair which enable us to model the scattered field spectrum in the form of an elliptical transfer function. This function de-

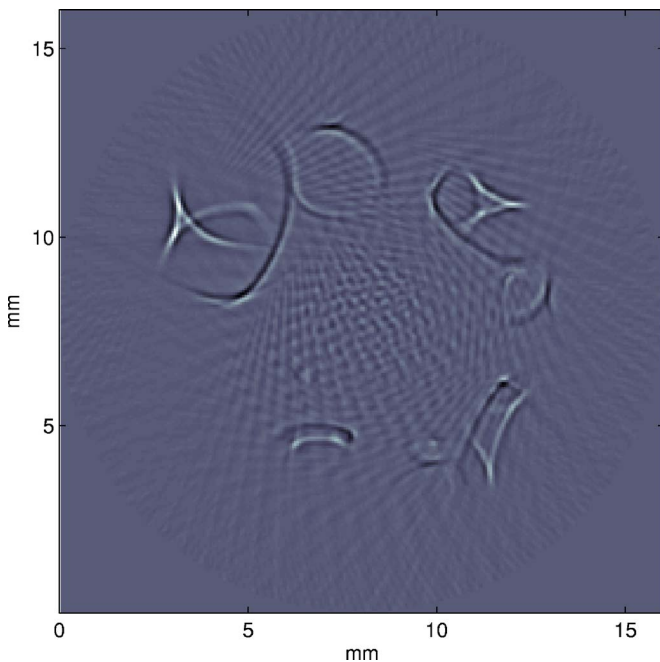


FIG. 12. Composite object reconstructed from diffraction measurements ($\theta=40^\circ$): plane wave excitation (2.5 MHz central frequency), standard backprojection (FBP) algorithm.

pend on the elliptical spectra of impedance and of sound speed, which give opposite local directivity diagrams: Impedance predominates in the reflection mode, whereas velocity predominates in the transmission mode. This makes it possible to reconstruct the impedance or the sound speed maps, depending on the data acquisition procedure used. Each of these parameters can thus be retrieved by applying an inverse 2D elliptical Fourier transform. However, it is better to adopt the optimized 1D fast Fourier transform algorithm used in the inverse elliptical Radon transform introduced here.

Numerical wave propagation simulations and the reconstruction of a tissue-mimicking phantom showing distinct contrast supports (for both impedance and sound speed fluctuations) confirm the validity of the present theory. This near-field scattering theory has many potential applications not only in the field of ultrasound imaging but also in the context of general wave theory.

ACKNOWLEDGMENTS

This study was supported by the City of Marseille, Region Provence-Alpes-Côte d'Azur, Conseil Général 13, CNRS-LMA. The authors would like to thank the Institute for Development and Resources in Intensive Scientific Computing (IDRIS) where the computations were performed. We are grateful to Chrysoula Tsogka (a researcher at CNRS) for placing her propagation code at our disposal.

APPENDIX A: ELLIPTICAL FOURIER TRANSFORMS

1. Dirac distributions

Let us consider the $I_{\mathbf{x}_0}(\mathbf{x})$ following integral of the variable $\mathbf{x} \in \mathcal{D}$ ($\mathcal{D} \subset \mathbb{R}^2$), where the fixed point \mathbf{x}_0 acts as a parameter:

$$I_{\mathbf{x}_0}(\mathbf{x}) = \frac{1}{(2\pi)^2} \int \frac{[\mathbf{x}_0]_{\mathbf{K}}}{[\mathbf{x}]_{\mathbf{K}}} e^{i\mathbf{K} \odot (\mathbf{x}_0 \ominus \mathbf{x})} d\mathbf{K}, \quad (\text{A1})$$

where $\mathbf{K} = k\mathbf{n}_{\phi, \theta}$ is a wave vector of the 2D elliptical Fourier domain, whose configuration angle θ is fixed. Then changing to polar coordinates, one obtains:

$$\begin{aligned} I_{\mathbf{x}_0}(\mathbf{x}) &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \frac{[\mathbf{x}_0]_{\phi}}{[\mathbf{x}]_{\phi}} \int e^{i\mathbf{k}\mathbf{n}_{\phi} \odot (\mathbf{x}_0 \ominus \mathbf{x})} k dk d\phi \\ &= \frac{i}{(2\pi)^2} \int_0^{2\pi} \mathbf{m}_{\phi}(\mathbf{x}) \nabla \left(\int e^{i\mathbf{k}\mathbf{n}_{\phi} \odot (\mathbf{x}_0 \ominus \mathbf{x})} dk \right) d\phi \\ &= \frac{i}{2\pi} \int_0^{2\pi} \mathbf{m}_{\phi}(\mathbf{x}) \delta'(\mathbf{n}_{\phi} \odot (\mathbf{x}_0 \ominus \mathbf{x})) d\phi, \quad (\text{A2}) \end{aligned}$$

$$\mathbf{m}_{\phi}(\mathbf{x}) = \frac{[\mathbf{x}_0]_{\phi}}{[\mathbf{x}]_{\phi}} \frac{\mathbf{n}_e + \mathbf{n}_r}{\|\mathbf{n}_e + \mathbf{n}_r\|^2}$$

is a differentiable (since $\mathbf{n}_e \neq -\mathbf{n}_r$) vector function and δ' is the first derivative of the Dirac distribution. The last integral may be written using the $\delta_{\mathcal{E}, \phi, \mathbf{x}_0}(\mathbf{x})$ distribution supported on the ellipse of ϕ direction, crossing the points \mathbf{x}_0 and \mathbf{x} , whose foci are on the \mathcal{S} circle:

$$I_{\mathbf{x}_0}(\mathbf{x}) = \frac{i}{2\pi} \int_0^{2\pi} \mathbf{m}_{\phi}(\mathbf{x}) \delta'_{\mathcal{E}, \phi, \mathbf{x}_0}(\mathbf{x}) d\phi. \quad (\text{A3})$$

Thus, $I_{\mathbf{x}_0}(\mathbf{x})$, as a sum of $\delta_{\mathcal{E}, \phi, \mathbf{x}_0}$ distributions as ϕ varies, is a distribution that can be approached by a discrete summation:

$$I_{\mathbf{x}_0}(\mathbf{x}) = \lim_{N \rightarrow \infty} \frac{i}{N} \sum_{n=0}^{N-1} \mathbf{m}_{\phi_n}(\mathbf{x}) \delta'_{\mathcal{E}, \phi_n, \mathbf{x}_0}(\mathbf{x}) \quad \text{with } \phi_n = \frac{2n\pi}{N}. \quad (\text{A4})$$

Let us now consider an arbitrary parameter function g whose support is included in \mathcal{D} . Here, we can distinguish two cases:

(1) $\mathbf{x} \neq \mathbf{x}_0$: then only two ellipses $\delta_{\mathcal{E}, \phi_1, \mathbf{x}_0}(\mathbf{x})$ and $\delta_{\mathcal{E}, \phi_2, \mathbf{x}_0}(\mathbf{x})$ crossing simultaneously \mathbf{x} and \mathbf{x}_0 and whose foci are on \mathcal{S} exist: For instance, in the incident direction defined by \mathbf{x}_0 , $\phi = \phi_0$, let us consider the ellipse crossing \mathbf{x}_0 and let us follow it as ϕ varies (in the positive and in the negative directions) until it also crosses \mathbf{x} . Therefore, since ϕ_1 and ϕ_2 are not necessarily rational numbers:

$$\begin{aligned} \langle I_{\mathbf{x}_0}(\mathbf{x}), g \rangle &\leq \lim_{N \rightarrow \infty} \frac{i}{N} [\mathbf{m}_{\phi_1}(\mathbf{x}) \langle \delta'_{\mathcal{E}, \phi_1, \mathbf{x}_0}(\mathbf{x}), g \rangle + \mathbf{m}_{\phi_2}(\mathbf{x}) \\ &\quad \times \langle \delta'_{\mathcal{E}, \phi_2, \mathbf{x}_0}(\mathbf{x}), g \rangle] = 0. \end{aligned} \quad (\text{A5})$$

The null value is obtained since the two elliptical projections of g' are finite quantities.

(2) $\mathbf{x} = \mathbf{x}_0$ the calculus of $I_{\mathbf{x}_0}(\mathbf{x} = \mathbf{x}_0)$ is straightforward: $I_{\mathbf{x}_0}(\mathbf{x}_0) = \infty$. To sum up:

$$I_{\mathbf{x}_0}(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0). \quad (\text{A6})$$

The 3D extension of this relation can be made according to a similar derivation:

$$\frac{1}{(2\pi)^N} \int \frac{[\mathbf{x}_0]_{\mathbf{K}}}{[\mathbf{x}]_{\mathbf{K}}} e^{i\mathbf{K} \odot (\mathbf{x}_0 \odot \mathbf{x})} d\mathbf{K} = \delta(\mathbf{x} - \mathbf{x}_0). \quad (\text{A7})$$

This relation enables us to validate the expression of the inverse elliptical transform (5). In addition one can show that

$$\delta(\mathbf{K} - \mathbf{Z}) = \frac{1}{(2\pi)^N} \int \frac{[\mathbf{x}]_{\mathbf{Z}}}{[\mathbf{x}]_{\mathbf{K}}} e^{-i(\mathbf{K} \odot \mathbf{Z}) \odot \mathbf{x}} d\mathbf{x}, \quad (\text{A8})$$

$$\delta(\mathbf{K} - \mathbf{Z} - \mathbf{Y}) = \frac{1}{(2\pi)^{2N}} \int \frac{[\mathbf{x}]_{\mathbf{Z}} [\mathbf{x}]_{\mathbf{Y}}}{[\mathbf{x}]_{\mathbf{K}}} e^{-i(\mathbf{K} \odot \mathbf{Z} \odot \mathbf{Y}) \odot \mathbf{x}} d\mathbf{x}. \quad (\text{A9})$$

2. Derivation rule

Let us calculate the value of the elliptical Fourier transform of the projected gradient (directional derivative), $\mathbf{n}_{e_{\mathbf{K}}} \cdot \nabla g(\mathbf{x})$, of the parameter function g defined over a compact support \mathcal{D} . This derivation is performed at the arbitrary elliptical spectral location $\mathbf{K} = k\mathbf{n}_{\phi, \theta = \theta_0}$:

$$\begin{aligned} \mathcal{F}_{\mathcal{E}}(\mathbf{n}_{e_{\mathbf{K}}} \cdot \nabla g)(\mathbf{K}) &= \int \nabla g(\mathbf{x}) \cdot \mathbf{n}_{e_{\mathbf{K}}}(\mathbf{x}) \varepsilon^{-i\mathbf{K} \odot \mathbf{x}} d\mathbf{x} \\ &= \int \nabla \left(\frac{1}{(2\pi)^N} \int \check{g}(\Gamma) \varepsilon^{i\Gamma \odot \mathbf{x}} d\Gamma \right) \\ &\quad \cdot \mathbf{n}_{e_{\mathbf{K}}}(\mathbf{x}) \varepsilon^{-i\mathbf{K} \odot \mathbf{x}} d\mathbf{x} \\ &= \int \frac{1}{(2\pi)^N} \left(\int \mathcal{K}_{\Gamma}(\mathbf{x}) \check{g}(\Gamma) \frac{[\mathbf{x}]_{\Gamma}}{[\mathbf{x}]_{\mathbf{K}}} \right. \\ &\quad \left. \times e^{-i(\mathbf{K} \odot \Gamma) \odot \mathbf{x}} d\Gamma \right) d\mathbf{x}, \end{aligned} \quad (\text{A10})$$

where for a probe wave having wave vector \mathbf{K} :

$$\begin{aligned} \mathcal{K}_{\Gamma}(\mathbf{x}) &= \left[\mathbf{n}_{e_{\Gamma}} \left(\frac{1}{\|\mathbf{x} - \mathbf{e}_{\Gamma}\|} - ik \right) \right. \\ &\quad \left. + \mathbf{n}_{r_{\Gamma}} \left(\frac{1}{\|\mathbf{x} - \mathbf{r}_{\Gamma}\|} - ik \right) \right] \cdot \mathbf{n}_{e_{\mathbf{K}}}. \end{aligned} \quad (\text{A11})$$

Then Eq. (A10) can be written as follows:

$$\begin{aligned} \mathcal{F}_{\mathcal{E}}(\mathbf{n}_{e_{\mathbf{K}}} \cdot \nabla g(\mathbf{x})) &= \int \check{g}(\Gamma) \int \check{\mathcal{K}}_{\Gamma}(\mathbf{Z}) \frac{1}{(2\pi)^{2N}} \\ &\quad \times \int \frac{[\mathbf{x}]_{\Gamma} [\mathbf{x}]_{\mathbf{Z}}}{[\mathbf{x}]_{\mathbf{K}}} e^{-i(\mathbf{K} \odot \Gamma \odot \mathbf{Z}) \odot \mathbf{x}} d\mathbf{x} \\ &= \int \check{g}(\Gamma) \int \check{\mathcal{K}}_{\Gamma}(\mathbf{Z}) \delta(\mathbf{K} - \Gamma - \mathbf{Z}) d\mathbf{Z} d\Gamma \\ &= \int \check{g}(\Gamma) \check{\mathcal{K}}_{\Gamma}(\mathbf{K} - \Gamma) d\Gamma. \end{aligned} \quad (\text{A12})$$

With probe-object distances of the order of 1 cm (or more) at usual ultrasonic frequencies (corresponding typically to the standard bandwidth of medical ultrasounds), the terms $1/\|\mathbf{x} - \mathbf{e}_{\Gamma}\|$ and $1/\|\mathbf{x} - \mathbf{r}_{\Gamma}\|$ are small compared with k ; so then

$$\mathcal{K}_{\Gamma}(\mathbf{x}) \simeq -ik[\mathbf{n}_{e_{\Gamma}}(\mathbf{x}) + \mathbf{n}_{r_{\Gamma}}(\mathbf{x})] \cdot \mathbf{n}_{e_{\mathbf{K}}}(\mathbf{x}). \quad (\text{A13})$$

However, as shown in Fig. 13, $\mathcal{K}_{\Gamma}(\mathbf{x})$ varies very slowly with \mathbf{x} , and so the bandwidth of $\check{\mathcal{K}}_{\Gamma}(\mathbf{K})$ is of the order of $c/2R$. It is about 300 times smaller than the central frequency of the transmitted wavelet. Therefore, the angle $(\mathbf{K}, \hat{\Gamma}) < 1^\circ$, and one has

$$\mathcal{K}_{\Gamma}(\mathbf{x}) \simeq \mathcal{K}_{\mathbf{K}}(\mathbf{x}) = -ik[1 + \mathbf{n}_{e_{\mathbf{K}}}(\mathbf{x}) \cdot \mathbf{n}_{r_{\mathbf{K}}}(\mathbf{x})]. \quad (\text{A14})$$

We take “ \otimes ” to denote the (spectral) elliptical convolution defined by

$$\check{g} \otimes \check{\mathcal{K}}_{\mathbf{K}}(\mathbf{K}) = \int \check{g}(\Gamma) \check{\mathcal{K}}_{\mathbf{K}}(\mathbf{K} - \Gamma) d\Gamma. \quad (\text{A15})$$

One obtains the following derivation rule:

$$\mathcal{F}_{\mathcal{E}}(\mathbf{n}_{e_{\mathbf{K}}} \cdot \nabla g)(\mathbf{K}) \simeq \check{g} \otimes \check{\mathcal{K}}_{\mathbf{K}}(\mathbf{K}). \quad (\text{A16})$$

Here, we have extended the classical derivation rule based on plane wave decomposition (using Fourier transforms). In the latter case, the inner integral would reduce to

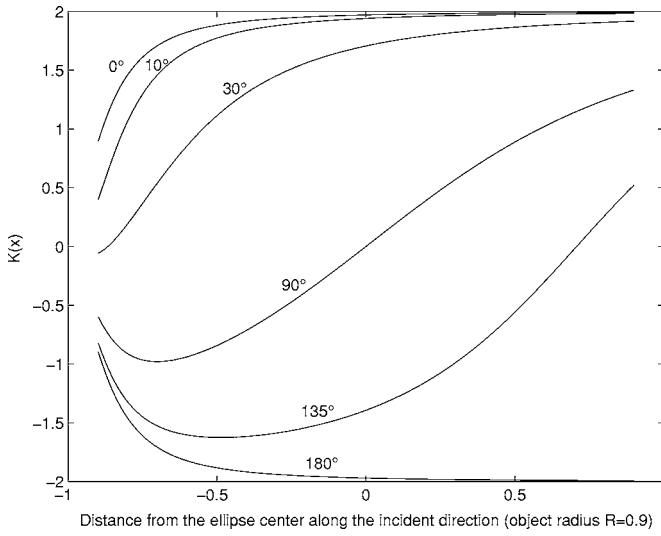


FIG. 13. Profiles of $\mathcal{K}_r(\mathbf{x})$ along the direction \mathbf{n}_ϕ at various angles (\mathbf{K}, Γ).

the distribution $-i2\mathbf{n}_\infty \cdot \mathbf{K} \delta(\mathbf{K} - \Gamma)$ giving the result $\mathcal{F}_\mathcal{E}(\mathbf{n}_\infty \cdot \nabla g)(\mathbf{K}) \approx 2F(\mathbf{n}_\infty \cdot \nabla g)(\mathbf{K}) = -i2\mathbf{n}_\infty \cdot \mathbf{K} F(g)(\mathbf{K})$; the transmitter-receiver being located at infinity. The factor 2 results from the fact that the wave performs a round-trip between the transmitter and the object. As regards the elliptical Fourier transform, the “derivation vector” $\mathcal{K}(\mathbf{x}) = \mathbf{n}_e / \|\mathbf{x} - \mathbf{e}\| + \mathbf{n}_r / \|\mathbf{x} - \mathbf{r}\| - ik(\mathbf{n}_e + \mathbf{n}_r)$ takes into account the curvature of the elliptical wave front.

In addition, if one operates in

$$\text{reflection } \mathbf{n}_{e\mathbf{K}} = \mathbf{n}_{r\mathbf{K}}, \quad \check{\mathcal{K}}_{\mathbf{K}} = -2ik\delta(\mathbf{K}),$$

$$\text{transmission } \mathbf{n}_{e\mathbf{K}} \approx -\mathbf{n}_{r\mathbf{K}}, \quad \check{\mathcal{K}}_{\mathbf{K}} \approx 0. \quad (\text{A17})$$

APPENDIX B: INVERSION OF THE ELLIPTICAL RADON TRANSFORM

Using the elliptical projection-slice theorem, Eq. (35) can be written:

$$f(\mathbf{x}) = \frac{1}{2} \times \frac{1}{\pi} \int_0^\pi \mathcal{F}_{\mathcal{E}|\phi}^{-1} \times [\text{Abs} \circ F \circ \mathcal{R}_\mathcal{E} \circ f](s = n_\phi \odot \mathbf{x}, \phi) d\phi, \quad (\text{B1})$$

where, for an arbitrary ϕ , $\mathcal{F}_{\mathcal{E}|\phi}^{-1}$ is the inverse elliptical Fourier transform:

$$\begin{aligned} & \mathcal{F}_{\mathcal{E}|\phi}^{-1}(\check{f}(S, \phi)|_{\phi \text{ fixed}})(s, \phi) \\ &= \frac{1}{2\pi} \int \check{f}(S, \phi)|_{\phi \text{ fixed}} e^{iS n_\phi \odot \mathbf{x}} [\mathbf{x}]_\phi dS. \end{aligned}$$

In order to obtain the inverse Radon transform, we take \mathcal{O} to denote the angular integration operator over the incident angle ϕ of any well-behaved function h mapping the cylindrical space $\mathcal{Y} = \{(s, \phi); s \in \mathbb{R}, 0 \leq \phi < \pi\}$:

$$\mathcal{O}[h](\mathbf{x}) = \frac{1}{2} \times \frac{1}{\pi} \int_0^\pi h(s = n_\phi \odot \mathbf{x}, \phi) d\phi \quad (\text{B2})$$

and from (Eq. (B1)):

$$f = \mathcal{O} \circ \mathcal{F}_{\mathcal{E}|\phi}^{-1} \circ \text{Abs} \circ F \circ \mathcal{R}_\mathcal{E} \circ f = \mathcal{O} \circ \mathcal{F}_{\mathcal{E}|\phi}^{-1} \circ \text{Abs} \circ \mathcal{F}_{\mathcal{E}|\phi} \circ f \quad (\text{B3})$$

In addition, one can note that

$$\begin{aligned} & [\mathcal{F}_{\mathcal{E}|\phi}^{-1} \circ \mathcal{F}_{\mathcal{E}|\phi} \circ f](\mathbf{x}) \\ &= \frac{1}{2\pi} \int \check{f}(S, \phi) e^{iS n_\phi \odot \mathbf{x}} [\mathbf{x}]_\phi dS \\ &= \int f(\mathbf{y}) \frac{[\mathbf{x}]_\phi}{[\mathbf{y}]_\phi} \frac{1}{2\pi} \int e^{iS n_\phi \odot (\mathbf{x} \odot \mathbf{y})} dS dy \\ &= \int f(\mathbf{y}) \frac{[\mathbf{x}]_\phi}{[\mathbf{y}]_\phi} \delta_{\mathcal{E}, \phi}(\mathbf{x} - \mathbf{y}) d\mathbf{y} = \pi_{f, \phi}(\mathbf{x}), \end{aligned} \quad (\text{B4})$$

where $\delta_{\mathcal{E}, \phi}(\mathbf{x})$ delineates the ellipse of foci ($\mathbf{e}_\phi, \mathbf{r}_\phi$) crossing the point \mathbf{x} . Then, $\pi_{f, \phi}(\mathbf{x})$ is the *elliptically* backprojected image of the elliptical projection $p_\phi(s)$. In addition we have

$$\begin{aligned} \pi_{f, \phi}(\mathbf{x}) &= \frac{1}{2\pi} \int \int \check{f}(S, \phi) e^{iS s} [\mathbf{x}]_\phi \delta(s - \mathbf{n}_\phi \odot \mathbf{x}) dS ds \\ &= [\mathbf{x}]_\phi \cdot [F^{-1} \circ \check{f}|_{\phi \text{ fixed}}](\mathbf{n}_\phi \odot \mathbf{x}) = [\mathcal{B}_\phi \circ F^{-1} \\ &\quad \circ \check{f}|_{\phi \text{ fixed}}](\mathbf{x}) \\ &= [\mathcal{B}_\phi \circ p_\phi](\mathbf{x}), \end{aligned} \quad (\text{B5})$$

where for any function h defined on \mathcal{Y} ,

$$\mathcal{B}_\phi[h(s, \phi)](\mathbf{x}) = [\mathbf{x}]_\phi \cdot h(\mathbf{n}_\phi \odot \mathbf{x}, \phi) \quad (\text{B6})$$

is the ϕ -backprojection²² operator on the ellipse perpendicular to the incident direction ϕ . To summarize, we have obtained the following relations:

$$[\mathcal{F}_{\mathcal{E}|\phi}^{-1} \circ \mathcal{F}_{\mathcal{E}|\phi} \circ f](\mathbf{x}) = \pi_{f, \phi}(\mathbf{x}) \quad (\text{B7})$$

$$\mathcal{F}_{\mathcal{E}|\phi}^{-1} = \mathcal{B}_\phi \circ F^{-1}. \quad (\text{B8})$$

The filtered elliptical projections Π_ϕ defined in Eq. (37) can be written:

$$\Pi_\phi = \mathcal{B}_\phi \circ F^{-1} \circ \text{Abs} \circ F \circ \mathcal{R}_\mathcal{E} \circ f = \mathcal{F}_{\mathcal{E}|\phi}^{-1} \circ \text{Abs} \circ \mathcal{F}_{\mathcal{E}|\phi} \circ f. \quad (\text{B9})$$

We retrieve the object function, Eq. (42):

$$\begin{aligned} f &= \mathcal{O} \circ \mathcal{F}_{\mathcal{E}|\phi}^{-1} \circ \text{Abs} \circ F \circ \mathcal{R}_\mathcal{E} \circ f \\ &= \mathcal{O} \circ \mathcal{B}_\phi \circ F^{-1} \circ \text{Abs} \circ F \circ \mathcal{R}_\mathcal{E} \circ f \\ &= \mathcal{B} \circ F^{-1} \circ \text{Abs} \circ F \circ \mathcal{R}_\mathcal{E} \circ f. \end{aligned} \quad (\text{B10})$$

Finally, the inverse elliptical Radon transform is defined by

$$\mathcal{R}_\mathcal{E}^{-1} = \mathcal{B} \circ F^{-1} \circ \text{Abs} \circ F. \quad (\text{B11})$$

¹S. J. Norton, “Reconstruction of a two-dimensional reflecting medium over a circular domain: Exact solution,” J. Acoust. Soc. Am. **67**, 1266–

- 1273 (1980).
- ²S. J. Norton and M. Linzer, "Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical and circular apertures," *IEEE Trans. Biomed. Eng.* **28**, 202–220 (1980).
- ³A. J. Devaney, "Generalized projection-slice theorem for fan beam diffraction tomography," *Ultrason. Imaging* **7**, 264–275 (1988).
- ⁴T. 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).
- ⁵P. Gan, R. Ludwig, and P. L. Levin, "Nonlinear diffractive inverse scattering for multiple scattering in inhomogeneous acoustic background media," *J. Acoust. Soc. Am.* **97**, 764–776 (1995).
- ⁶T. 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).
- ⁷M. P. André, H. S. Jarrée, G. P. Otto, P. J. Martin, and J. P. Jones, "Reduction of phase aberration in a diffraction tomography system for breast imaging," *Acoust. Imaging* **22**, 151–157 (1996).
- ⁸J. W. Wisikin, D. Borup, and S. A. Johnson, "Inverse scattering from arbitrary two-dimensional objects in stratified environments via a Green's operator," *J. Acoust. Soc. Am.* **102**, 853–864 (1997).
- ⁹S. Pourjavid and O. Tretiak, "Ultrasound imaging through time-domain diffraction tomography," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **38**, 74–85 (1991).
- ¹⁰C. Q. Lan and W. Xiong, "An iterative method of ultrasonic reflection-mode tomography," *IEEE Trans. Med. Imaging* **13**, 419–425 (1994).
- ¹¹T. D. Mast, "Wideband quantitative ultrasonic imaging by time-domain diffraction tomography," *J. Acoust. Soc. Am.* **106**, 3061–3071 (1999).
- ¹²M. H. Xu and L. H. Wang, "Time-domain reconstruction for thermoacoustic tomography in a spherical geometry," *IEEE Trans. Med. Imaging* **21**, 814–822 (2002).
- ¹³M. H. Xu, Y. Xu, and L. H. Wang, "Time-domain reconstruction and numerical simulations for thermoacoustic tomography in various geometries," *IEEE Trans. Med. Imaging* **50**, 1086–1099 (2003).
- ¹⁴G. Ambartsoumian and P. Kuchment, "On the injectivity of the circular Radon transform," *Inverse Probl.* **21**, 473–485 (2005).
- ¹⁵J. F. Greenleaf, S. A. Johnson, W. F. Wamoya, and F. A. Duck, "Algebraic reconstruction of spatial distributions of acoustics velocities in tissue from their time-of-flight profiles," in *Acoustical Holography* (Plenum, New York, 1975), pp. 71–90.
- ¹⁶S. J. Norton, "Generation of separate density and compressibility images in tissue," *Ultrason. Imaging* **5**, 240–252 (1983).
- ¹⁷A. J. Devaney, "Variable density acoustics tomography," *J. Acoust. Soc. Am.* **78**, 120–130 (1985).
- ¹⁸S. Mensah and J.-P. Lefebvre, "Enhanced compressibility tomography," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **44**, 1245–1252 (1997).
- ¹⁹J. P. Jones, "Impediography, a new technique for diagnostic medicine," *Ultrasound in Medicine*, edited by D. N. White (Plenum, New York, 1975), Vol. **1**, pp. 489–497.
- ²⁰S. J. Leeman, "Impediography equation," *Acoustical Imaging*, edited by A. F. Metherell (Plenum, New York, 1980), Vol. **8**, pp. 517–525.
- ²¹J.-P. Lefebvre, "A linearised inverse problem: Acoustic impedance tomography in biological media," in *Electromagnetic and Acoustic Scattering: Detection and Inverse Problem*, edited by C. Bourelly *et al.* (World Scientific, Singapore, 1988), pp. 287–301.
- ²²S. R. Deans, *The Radon Transform and Some of its Applications* (Wiley, New York, 1983), pp. 131–136.
- ²³C. Tsogka, "Modélisation mathématique et numérique de la propagation des ondes élastiques tridimensionnelles dans des matériaux fissurés," ("Numerical and mathematical modeling of elastic wave propagation in fractured materials"), Ph.D. thesis, Université Paris Dauphine—Paris IX, 1999.
- ²⁴E. Bécache, P. Joly, and C. Tsogka, "An analysis of new mixed finite elements for the approximation of wave propagation problems," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **37**, 1053–1084 (2000).
- ²⁵F. Collino and C. Tsogka, "Application of the pml absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media," *Geophysics* **66**, 294–305 (2001).
- ²⁶R. Waag and R. Fedewa, "A ring transducer system for medical ultrasound research," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53**, 1707–1718 (2006).

Acoustic power calibration of high-intensity focused ultrasound transducers using a radiation force technique

Subha Maruvada, Gerald R. Harris, and Bruce A. Herman

Center for Devices and Radiological Health, Food and Drug Administration Rockville, Maryland 20850

Randy L. King

King Acoustic Technologies, LLC, Washington, DC 20007

(Received 19 July 2006; revised 5 December 2006; accepted 6 December 2006)

To address the challenges associated with measuring the ultrasonic power from high-intensity focused ultrasound transducers via radiation force, a technique based on pulsed measurements was developed and analyzed. Two 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 applied power levels. Two absorbing target designs were used, and both gave comparable results and displayed no damage and minimal temperature rise if placed near the transducer and away from the focus. The method yielded reproducible results up to the maximum pulse power generated of approximately 230 W, thus allowing the radiated power to be calibrated in terms of the peak-to-peak voltage applied to the transducer. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2431332]

PACS number(s): 43.35.Yb [TDM]

Pages: 1434–1439

I. INTRODUCTION

Knowledge of the ultrasonic power radiated by therapeutic ultrasound transducers such as used in high-intensity focused ultrasound (HIFU) surgery is important from both an effectiveness and safety standpoint.^{1–3} The power produced by transducers used in biomedical ultrasound applications typically is found by radiation force means, in which the axial force (i.e., the force in the direction of propagation) on a target attached to a balance is measured.^{2–18} Two types of targets are used in a radiation force balance system: reflecting and absorbing⁷ (see Sec. II B). The relationship between the measured force (F) and the temporal-average acoustic power (P_{TA}) for the case of plane waves and a perfectly absorbing target is

$$P_{TA} = cF, \quad (1)$$

where c is the speed of sound in the propagation medium, usually water.^{7,19} For focused fields with beam convergence angle $\alpha = \sin^{-1}(d_S/2L)$,

$$P_{TA} = 2cF/(1 + \cos \alpha), \quad (2)$$

where d_S and L are the diameter and geometrical focal length (radius of curvature) of the source transducer, respectively.^{7,19}

Equations (1) and (2) have been used successfully in many measurement situations, but for high-power focused beams these measurements can be challenging.² For example, transducer damage can occur if the measurement time is longer than would be encountered in clinical use. Also, for absorbing targets, excessive heating can result in measurement error or target damage. Furthermore, bubble formation between the transducer and target at high powers can introduce measurement errors. Acoustic streaming is another potential source of measurement error that increases as both P_{TA} and ultrasonic frequency increase. To overcome these

problems, pulsed-mode instead of continuous-wave (cw) measurements can be made to lower the P_{TA} , from which the power during the pulse can be computed. The objective of this work was to analyze this approach for calibrating the transducer output power in terms of the applied peak-to-peak voltage.

II. MATERIALS AND METHODS

A. Transducers

Ultrasonic frequencies between 1 and 2 MHz commonly are employed in HIFU.²⁰ In this study two focused transducers operating in this range were used (1.50 MHz, 10-cm diameter, 15-cm focal length, ONDA Corp, Sunnyvale, CA; 1.11 MHz, 8-cm diameter, 10-cm focal length, King Acoustic Technologies, LLC, Washington, DC). The values of $\cos \alpha$ in Eq. (2) for these two transducers were 0.866 and 0.917, respectively. A third transducer (3.33 MHz, 6-cm diameter, 5-cm focal length, $\cos \alpha = 0.800$) was used to assess the effect of acoustic streaming at a higher frequency. All transducers were matched to an electrical impedance of 50 ohms.

B. Targets

For all reflecting targets, plane and conical, the force varies with the angle between the beam axis and the normal to the reflecting surface. Most reflecting targets are convex cones, for which precise positioning is necessary to achieve accurate measurements. Also, a disadvantage of conical reflectors in focused fields is that the radiation force is very sensitive to the beam convergence angle.¹⁸ Furthermore, with these targets more care is needed in tank design to avoid reflections. For absorbing targets these factors are much less critical. However, for high-power transducers, target heating can cause measurement instability, and for highly focused

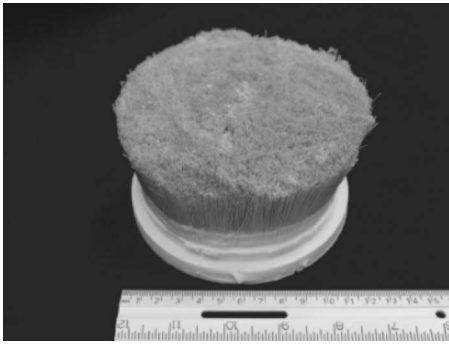


FIG. 1. Twelve-cm brush target.

beams, target damage can occur. In this work, three absorbing targets were employed and efforts were made to overcome the disadvantages.

Two “brush” targets⁸ were constructed having 10-cm (made in-house) and 12-cm (King Acoustic Technologies, LLC, Washington, DC) diameters with bristles of 4 and 6.5 cm length, respectively. Figure 1 shows a picture of the 12-cm brush target. Nylon bristles from common cleaning brushes were pulled and packed densely together and then potted in a base of rubber. The approximate diameters of the nylon brush bristles were 0.17 and 0.25 mm, respectively, for the 10- and 12-cm brush targets. The 10-cm brush was potted in a silicone elastomer (Sylgard 170, Dow Corning, Midland, MI) and the 12-cm brush was potted in a proprietary two-part rubber material designed to absorb ultrasound (King Acoustic Technologies, Washington, DC).

The third target, a commercially available absorbing design^{14,21} (model Ham A, Precision Acoustics Ltd., Dorchester, UK), is based on polyurethane rubber material. It is composed of two layers: a top layer whose acoustic impedance is matched to that of water and a backing layer that is partially air-loaded to increase transmission loss. The target has flat front and back faces and a thickness of 14 mm. The material was cut into a disk of 12 cm diameter.

According to Ref. 7, for unfocused transducers the target diameter should be at least 1.5 times the transducer diameter to intercept all significant parts of the field. For focused transducers, an equivalent specification based on simple ray acoustics can be expressed as

$$d_T \geq 1.5d_S(L-z)/L, \quad (3)$$

where d_T is the target diameter and z is the transducer-to-target distance.

The transducer-to-target distances stated herein refer to the distance between the plane defined by the transducer rim and either the tips of the brush bristles or the front face of the flat target. For the 1.11- and 1.50-MHz transducers, most measurements were made at a transducer-to-target distance of 5 cm to satisfy Eq. (3), but positions closer to the focus were used for comparison.

C. Radiation force balance system

The setup of the experiment is shown in Fig. 2. The absorbing target was suspended from the bottom of an electronic balance (AND HM-202, Bradford, MA), and the trans-

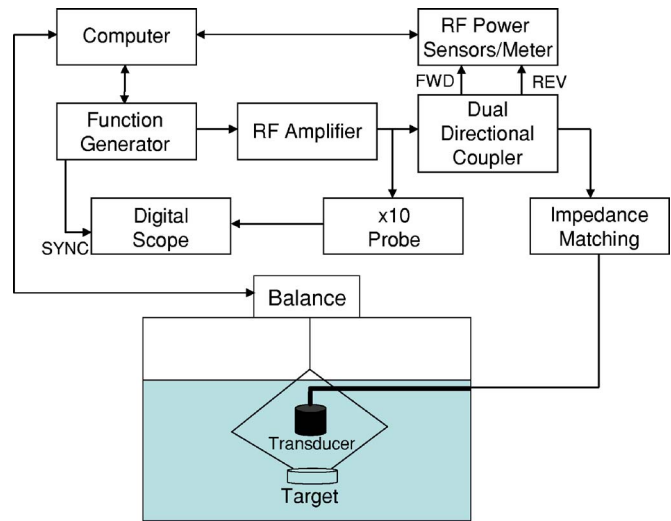


FIG. 2. Experimental setup of radiation force balance.

ducer was positioned so that it radiated downward onto the target. The driving electronics consisted of a function generator (Wavetek 81, Fluke Corp., Everett, WA) and power amplifier (ENI 2100L or ENI A-300, Rochester, NY). A 50-dB dual-directional coupler (Amplifier Research, model DC2000, Souderton, PA), two power sensors (model 8482A, Agilent Tech., Palo Alto, CA), and a power meter (model E4419B, Agilent Tech., Palo Alto, CA) were placed between the amplifier and transducer to monitor the forward and reversed electrical power to the transducer. Operation of the radiation force balance system was accomplished through MATLAB (The Mathworks Inc., Natick, MA). MATLAB was used to control the hardware (function generator, balance, and power meter) as well as to acquire and analyze the data.

The peak-to-peak transducer voltage was measured during the steady-state portion of the pulse with a digital oscilloscope (model 54622A, Agilent Tech., Palo Alto, CA). The peak-to-peak rather than rms pulse voltage was measured for convenience, although either would provide a robust calibration of the transducer and drive electronics combination until amplifier saturation at high drive levels becomes significant.

D. Effective duty factor (EDF)

Two pulse repetition frequencies, 500 and 1000 Hz, were used in this approach for circumventing the problems associated with measuring high temporal-average powers. For both, the pulse duration was chosen to give a duty factor (DF) equal to the product of the pulse repetition frequency and the pulse duration, of approximately 10%. The power during the pulse (P_{PA}) can be calculated from the measured temporal-average power as in Eq. (4),

$$P_{PA} = P_{TA}/DF. \quad (4)$$

Equation (4) is exact for a power pulse waveform having a true rectangular modulating envelope. However, turn-on and turn-off transients can lead to inaccuracies in Eq. (4) that will increase as the pulses duration decreases. Therefore, first an “effective duty factor” (EDF), defined as the ratio of pulsed to cw power, P_{PUL}/P_{CW} , was established by measuring

TABLE I. cw and pulsed-mode power values to establish an effective duty factor (EDF) for the 1.11-MHz transducer.

U_{pp} (V)	P_{cw} (W)	$P_{PUL}(500)$ (W)	$P_{PUL}(1000)$ (W)	EDF (500)	EDF (1000)
72	9.94	0.88	0.84	0.089	0.085
88	15.11	1.34	1.29	0.089	0.085
102	20.23	1.79	1.73	0.088	0.086
112	24.56	2.17	2.09	0.088	0.085
123	29.65	2.62	2.53	0.088	0.085

both P_{PUL} and P_{CW} with the transducer driven at the same peak-to-peak transducer voltage for five cw P_{TA} 's approximately evenly spaced over the range from 10 to 30 W. The EDF was the average valued obtained over this range. Then, in pulsed mode the transducer drive level was increased and the power during the pulse was calculated from

$$P_{PA} = P_{TA}/EDF. \quad (5)$$

E. Other measurement considerations

Each measurement comprised five on-off cycles under computer control per the procedure in Ref. 16. The on time was 9 s, a time sufficient to attain and process a stable balance reading, while the off time was 20 s to allow the target temperature to return to near baseline (see Sec. III E). Four independent measurements were made at each drive level to assess the type A (random) uncertainty.²²

The water was degassed to <2 ppm and measurements were made at room temperature (22 °C–25 °C). The speed of sound c in Eq. (2) was adjusted for its variation with temperature.²³

To determine where in the brush targets heating due to ultrasound absorption might arise, 250- μ m-diameter wire, copper-constantan thermocouples were placed centrally at three locations within the 10- and 12-cm brush targets: near the base of the bristles, in the center of the bristles, and 1 mm below the top of the bristles. Temperature measurements also were made in the flat target using the same type of thermocouple placed in the center and approximately 1 mm below the front surface.

To assess the effect of acoustic streaming, a 6.4- μ m-thick low-density polyethylene membrane 12.5 cm in diameter was inserted between the transducer and target in one set of measurements.⁷ For transducers with planar radiating surfaces, it is recommended that the antistreaming membrane be tilted with respect to the beam axis to eliminate the possibility of standing waves. To see if such angulation is critical for focused beams, the membrane angle was varied from approximately 0° to 5°, 0° corresponding to the plane normal to the beam axis. For the 3.3-MHz transducer used in the acoustic streaming measurements, the target was placed 2.5 cm from the transducer and also at an 8-cm distance to exacerbate the effects of streaming.

III. RESULTS AND DISCUSSION

A. Effective duty factor

Continuous wave and pulsed measurements to establish an EDF were performed on the 1.50- and 1.11-MHz transducers. The results for the 1.11-MHz transducer are given in Table I. $P_{PUL}(500)$ and $P_{PUL}(1000)$ are the pulsed mode P_{TA} 's at pulse repetition frequencies of 500 and 1000 Hz, respectively. The means and coefficients of variation for the five EDF values were 0.088 and 0.15% at 500 Hz and 0.085 and 0.47% at 1000 Hz. For the 1.50-MHz transducer, the means and coefficients of variation for the five EDF values were 0.094 and 0.7% at 500 Hz and 0.091 and 1.9% at 1000 Hz. The corresponding DFs [Eq. (4)] were 0.090 and 0.093 for the 1.11- and 1.50-MHz transducers, respectively, indicating a small but measurable effect of the turn-on and turn-off transients. (See also Sec. III G, below.)

B. High-output pulsed mode powers

Figure 3(a) shows the pulsed-mode power [P_{PA} from Eq. (5)] for the 1.11-MHz transducer using the 10-cm brush target vs transducer voltage for the two pulse repetition frequencies. The corresponding temporal-average powers ranged from 4 to 22 W. Peak acoustic powers during the pulse were calculated from Eq. (4) to be approximately 40 to 230 W, the upper value being the highest possible with the available amplifiers before noticeable distortion of the sinusoidal drive voltage was observed. Power-law fits to these data lead to calibration equations for the acoustic power of $P_{TA}=0.00166 U_{PP}^{2.037}$ (coefficient of determination $R^2=0.999$) and $P_{TA}=0.00168 U_{PP}^{2.038}$ ($R^2=0.999$) from the 500- and 1000-Hz repetition rate results, respectively, where the power, P_{TA} , has units of watts and the peak-to-peak voltage, U_{PP} , has units of volts.

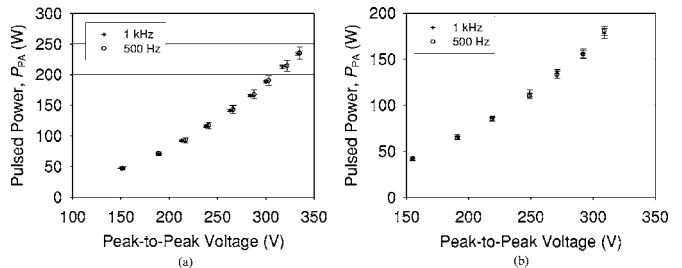


FIG. 3. Comparison of measured pulsed-mode power using the two different pulse repetition frequencies for the (a) 1.11-MHz transducer and (b) 1.50-MHz transducer.

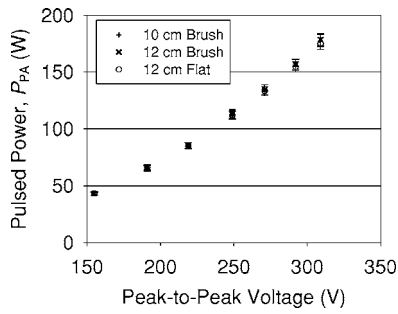


FIG. 4. Comparison of three absorbing targets.

Similar results were obtained with the 1.50-MHz transducer, for which pulsed-mode temporal-average acoustic powers from 4 to 16 W were measured using the 10-cm brush target [Fig. 3(b)]. Peak acoustic powers during the pulse were calculated to be approximately 40 to 180 W. The power law fits were $P_{TA}=0.00109 U_{PP}^{2.092}$ (500 Hz, $R^2=0.999$) and $P_{TA}=0.00132 U_{PP}^{2.058}$ (1000 Hz, $R^2=0.999$). Although the power-law fits appear to be different, the coefficients and exponents are such that the difference in P_{TA} values calculated for the two repetition rates are small and within the experimental error seen in Fig. 3(b).

C. Comparison of three targets

Pulsed-mode measurements for the 1.50-MHz transducer with all three targets are shown in Fig. 4. The distance between transducer and each target was 5 cm. Little variation (less than 2%) was found among the targets, the slight difference at the highest power being due to difficulty in recording the peak-to-peak voltage because of instability in the power amplifier at its maximum output. Similar results were seen with the 1.11-MHz transducer.

D. Effect of transducer-to-target distance

The distance between the transducer and target was varied using the 1.50-MHz transducer and the two brush targets. The distance was increased from 5 to 11 cm while the transducer was driven in cw mode to provide approximately 10- and 30-W temporal-average acoustic powers. As shown in Fig. 5, the measured acoustic power stayed nearly constant vs distance. In only one case, at 30 W with the 12-cm brush, was there a small but statistically significant drop in acoustic power of 2% at 11 cm versus 5 cm. The small attenuation in

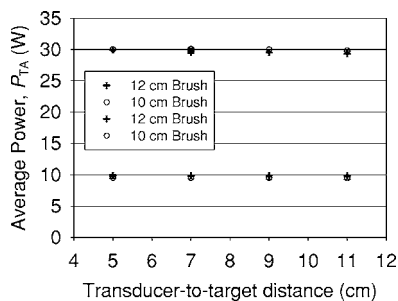


FIG. 5. Measured temporal-average power vs distance between transducer and 10- and 12-cm brush target using the 1.50-MHz transducer at 10 and 30 W.

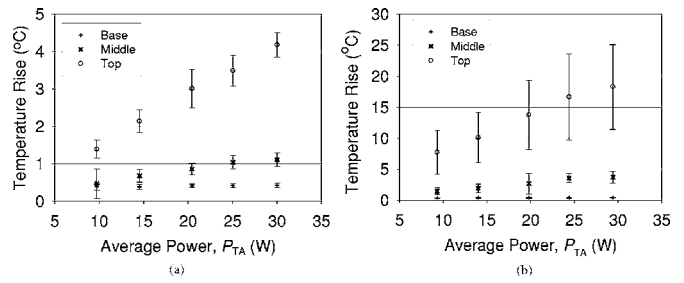


FIG. 6. Average temperature rise vs measured temporal-average acoustic power at three locations in 12-cm brush target at transducer-target distances of (a) 5 cm and (b) 10 cm using the 1.50-MHz transducer.

water over this range had no effect on these results. It should be noted, however, that the variation with distance may be greater at higher powers or frequencies where there will be more nonlinearity in the water and increased absorption and shock loss.

The same experiment was attempted with the 1.11-MHz transducer and 12-cm flat target while the transducer was driven at approximately 10-W temporal-average acoustic power. However, visible damage to the target occurred at a distance of 9 cm and the experiment was aborted.

E. Target temperature measurements

Using the 1.50-MHz transducer and 12-cm brush target, the temperature rise measured by the three thermocouples during each of the five on-off cycles was recorded and the results were averaged. The cycle off time did not allow the temperature to return to baseline, so the final temperature at the end of each cycle on time was greater than that of the previous cycle. However, the five temperature rises were approximately the same. The average temperature rise over the five on cycles was found for P_{TA} from 10 to 30 W for cw excitation. The temperature rise was recorded at two transducer-to-target distances of 5 and 10 cm. The results are plotted in Fig. 6. The temperature rise was much larger for the 10-cm distance as the target was closer to the transducer focus. The temperature rise was greatest near the top of the brush and decreased towards the base of the brush. The maximum temperatures at the end of the fifth cycle on time of the 30-W sonications were 30 °C, 27 °C, and 24 °C for the three thermocouple positions when the target was 5 cm from the transducer, and 43 °C, 28 °C, and 23 °C when the target was 10 cm from the transducer. At this maximum cw power, the maximum temperature rises were about 4 °C and 18 °C for transducer-to-target distances of 5 and 10 cm, respectively (see Fig. 6). Similar results were obtained with the 10-cm brush target.

Regarding the flat target, as mentioned above, thermal damage occurred at a distance of 9 cm. At 5 and 7 cm the power measured was 10.5 and 10.7 W while the average temperature rise was 6 °C and 13 °C, respectively. The maximum temperatures were 36 °C at 5 cm and 45 °C at 7 cm.

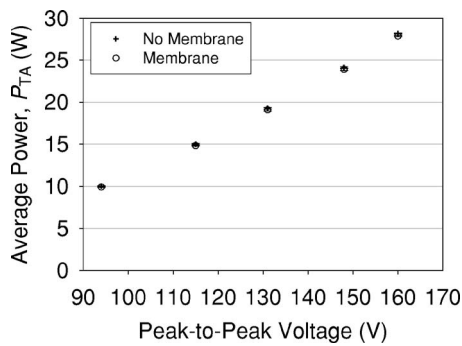


FIG. 7. Streaming measurements for 3.33-MHz transducer and 10-cm brush target. The transducer-to-target distance was 2.5 cm and the antistreaming membrane was placed 1 cm from the target.

F. Acoustic streaming

The attenuation through the antistreaming membrane, determined via hydrophone measurements, was found to be less than 0.15% (~ 0.01 dB) at 3.3 MHz. For the 3.3-MHz transducer, measurements were made with and without the membrane using cw excitation from 10 to 30 W at distances of 2.5 and 8 cm between the transducer and 10-cm brush target. Figure 7 shows the results for the 2.5-cm target distance both without the membrane and with the membrane placed 1 cm from the target. At 30 W the membrane caused a reduction in the measured power of 1.1%. (This same measurement for the 1.5-MHz transducer and a target distance of 5 cm resulted in a decrease of 0.2% at 30 W.)

Figure 8 shows the streaming effects on the measurements for the 3.33-MHz transducer when the brush target was placed 8 cm from the transducer. Two membrane distances were used: 1 cm from the transducer face and 1 cm from the target. The reduction in the measured power was 0.6% for the membrane near the transducer and 18% for the membrane near the target at 30 W. This latter result demonstrates that the power lost due to attenuation in the water is recaptured via streaming. In this regard, it is noted that an antistreaming membrane should be used only if the power at the target location is desired. To achieve a more accurate measurement of the total power radiated by the transducer, no membrane should be used. Also, because streaming is enhanced by nonlinear propagation, its effects will be reduced by making measurements at reduced temporal-average powers as in the present pulsed technique.

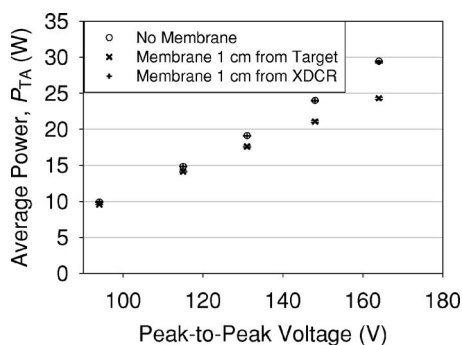


FIG. 8. Streaming measurements for 3.33-MHz transducer and 10-cm brush target. The transducer-to-target distance was 8 cm.

Finally, with the target 8 cm from the 3.3-MHz transducer, tilting the membrane from 0° – 5° had no discernible effect on the measurements at membrane positions both near the transducer and near the target.

G. Application of calibration method

Successful application of the strategy expressed in Eq. (5) relies on the constancy of the EDF as drive level is increased. That is, the shape of the pulsed transducer voltage waveform must remain unchanged with drive level. To validate this premise, the rms transducer voltage, normalized by the peak-to-peak voltage, was measured over the full range of drive levels used for the 1.11- and 1.50-MHz transducers. For both transducers, the coefficient of variation was found to be $<1\%$ for 13 measurements between 10 and 180 W. For the 1.11-MHz transducer, which was capable of higher output, the coefficient of variation was $<2\%$ for 16 measurements between 10 and 230 W.

Further, in using calibration data obtained under pulsed conditions, it should be recognized that when a HIFU transducer is driven at full (i.e., cw) output, self-heating due to $<100\%$ efficiency of operation can affect the acoustic output. However, other means can be used to evaluate possible transducer instability during continuous operation, such as monitoring the forward and reverse electrical power, a change in which would be indicative of a change in transducer impedance and output due to, for example, thermal drift in the matching network components. Also, one could observe the output over time of a hydrophone placed in a low-intensity region of the field where neither hydrophone damage nor saturation effects due to nonlinear propagation are expected.

IV. CONCLUSION

Temporal-average acoustic powers of approximately 4 to 22 W were measured from focused ultrasound therapy transducers in a pulsed mode. By determining the effective duty factor of the pulses at relatively low temporal-average power, it was possible to relate peak-to-peak transducer voltages to acoustic powers up to 230 W. Three absorbing targets of two designs were used, and powers agreed for all targets to within the Type A (random) measurement uncertainty at all drive levels (maximum coefficient of variation: $<2\%$ at 95% confidence level), indicating that all three are suitable as absorbing targets. No evidence of cavitation (e.g., balance instability) was observed at the 500- and 1000-Hz pulse repetition frequencies used and a duty factor of about 10%. Brush target temperature rises were less than about 4°C at a target distance of 5 cm. The temperature rises in the flat target were greater, with thermal damage occurring if it was placed near the focus. For the brush targets, no significant variation in measured power was seen for the transducer-to-target distances used. For these measurements a distance of 5 cm was found to be a reasonable compromise in that it was small enough to avoid significant focal heating, yet large enough to permit a practical target size according to Eq. (3).

Streaming effects were not observed at 1.50 MHz but were seen with the 3.33-MHz transducer. A membrane may be needed to eliminate streaming effects on power measurements if the power at the plane of the target is desired. However, at higher frequencies the membrane can introduce greater variation in the measured acoustic power, most likely due to reflection from the membrane. A membrane may be counterproductive, however, if the total power output is the quantity of interest, since unobstructed streaming allows some recapture, by the target, of the momentum in the ultrasound field lost to absorption in the transmitting medium.

These results suggest that a pulsed-mode approach using 500–1000-Hz pulse repetition frequencies, approximately 10% duty factor, absorbing target not near focus, and well-degassed water should be a suitable method for accurate voltage-power characterization that avoids transducer or target damage and bubble- or thermal-related measurement anomalies. This technique is appropriate for preclinical research, development, and testing of HIFU transducers and systems where monitoring and control of the transducer drive electronics is available. For further assessment of transducer performance under cw drive conditions, the forward and reverse electrical power can be measured, or the output vs time of a hydrophone placed in a low-intensity region of the field can be observed.

Note: The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the U.S. Department of Health and Human Services.

¹G. Fleury, R. Berriet, J. Y. Chapelon, G. ter Haar, C. Lafon, O. Le Baron, L. Chupin, F. Pichonnat, and J. Lenormand, "Safety issues for HIFU transducer design," in *4th International Symposium on Therapeutic Ultrasound Proceedings*, edited by G. R. ter Haar and I. Rivens (American Institute of Physics, Melville, NY, 2005), pp. 233–241.

²A. Shaw and G. ter Haar, "Requirements for Measurement Standards in High Intensity Focused Ultrasound (HIFU) Fields," National Physical Laboratory Report DQL AC 015 (National Physical Laboratory, Teddington, UK, February 2006).

³V. A. Khokhlova, M. R. Bailey, J. A. Reed, B. W. Cunitz, P. J. Kaczowski, and L. A. 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).

⁴G. Kossoff, "Balance technique for the measurement of very low ultrasonic power outputs," *J. Acoust. Soc. Am.* **38**, 880–881 (1965).

⁵K. Beissner, "Minimum target size in radiation force measurements," *J. Acoust. Soc. Am.* **76**, 1505–1510 (1984).

⁶R. A. Robinson, "Performance evaluation of a digital readout hyperthermia range ultrasonic wattmeter," *IEEE Trans. Sonics Ultrason.* **31**, 467–472 (1984).

⁷"Ultrasonic Power Measurement in Liquids in the Frequency Range 0.5 MHz to 25 MHz," Publ. IEC 61161 (International Electrotechnical Commission, Geneva, 1992).

⁸K. Hynynen, "Acoustic power calibrations of cylindrical intracavitary ultrasound hyperthermia applicators," *Med. Phys.* **20**, 129–134 (1993).

⁹K. Beissner, W. A. Oosterbaan, R. T. Hekkenberg, and A. Shaw, "European intercomparison of ultrasonic power measurements," *Acust. Acta Acust.* **82**, 450–458 (1996).

¹⁰S. E. Fick, "Ultrasound power measurement by pulsed radiation pressure," *Metrologia* **36**, 351–356 (1999).

¹¹R. T. Hekkenberg, K. Beissner, B. Zeqiri, R. A. Bezemer, and M. Hodnett, "Validated ultrasonic power measurements up to 20 W," *Ultrasound Med. Biol.* **27**, 427–438 (2001).

¹²G. S. K. Wong and L. Wu, "High power ultrasound standard," *J. Acoust. Soc. Am.* **111**, 1791–1799 (2002).

¹³K. Beissner, "IEC measurement standards for ultrasonic hydrophones and radiation force balances," *J. Acoust. Soc. Am.* **112**, 2342 (2002).

¹⁴Y. Sutton, A. Shaw, and B. Zeqiri, "Measurement of ultrasonic power using an acoustically absorbing well," *Ultrasound Med. Biol.* **29**, 1507–1513 (2003).

¹⁵P. A. Lewin, N. Barrie-Smith, M. Ide, K. Hynynen, and M. Macdonald, "Interlaboratory acoustic power measurement," *J. Ultrasound Med.* **22**, 207–213 (2003).

¹⁶"Acoustic Output Measurement Standard for Diagnostic Ultrasound Equipment, Rev. 3," NEMA Standards Publication UD-2-2004 (American Institute of Ultrasound in Medicine, Laurel, MD, and National Electrical Manufacturers Association, Rosslyn, VA, 2004).

¹⁷G. R. Harris, "Progress in medical ultrasound exosimetry," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **52**, 717–736 (2005).

¹⁸A. Shaw, "How to measure HIFU output power properly," in *5th International Symposium on Therapeutic Ultrasound Proceedings*, edited by G. T. Clement, N. J. MacDonnald, and K. Hynynen (American Institute of Physics, Melville, NY, 2006), pp. 628–632.

¹⁹K. Beissner, "Radiation force calculations," *Acustica* **62**, 255–263 (1987).

²⁰G. Clement, "Perspectives in clinical uses of high-intensity focused ultrasound," *Ultrasonics* **42**, 1087–1093 (2004).

²¹B. Zeqiri and C. J. Bickley, "A new anechoic material for medical ultrasonic applications," *Ultrasound Med. Biol.* **26**, 481–485 (2000).

²²M. C. Ziskin, "Specification of acoustic output level and measurement uncertainty in ultrasonic exosimetry," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50**, 1023–1034 (2003).

²³N. Bilaniuk and G. S. K. Wong, "Speed of sound in water as a function of temperature," *J. Acoust. Soc. Am.* **93**, 1609–1612 (1993).

Elastic waves excited by a plane source on the surface of a multilayered medium

Bixing Zhang^{a)}

Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100080, People's Republic of China

(Received 14 June 2006; revised 4 December 2006; accepted 16 December 2006)

Elastic waves excited by a plane piezoelectric source with an arbitrary shape on the surface of a multilayered medium have been studied for the first time in this paper. On the basis of Abo-zena [Geophys. J. R. Astron. Soc. **58**, 91–105 (1979)] and Menke [Geophys. J. R. Astron. Soc. **59**, 315–323 (1979)], the propagator matrix for the elastic wave field in multilayered medium is extended from two- to three-dimensional (3D) space. 3D elastic wave field is investigated and the displacement-stress response for the boundary conditions is obtained. The propagation of elastic wave in multilayered media is analyzed in 3D space in the frequency domain. The *P-SV* and *SH* modes corresponding to the poles are studied. The excitation and propagation of the modes are analyzed further. It is found that the propagation velocities of the *P-SV* and *SH* modes do not depend on the propagation azimuth θ in the plane parallel to the free surface of the multilayered medium while the displacement amplitudes are strongly dependent on the azimuth θ . The directional distribution functions of the modes are independent of the medium parameters and the modes are dependent on the shape and excitation fashion of the source. Finally, as an example, the displacement fields of the *P-SV* and *SH* modes excited by a rectangle source are analyzed. The displacement representation and numerical results of the directivity distribution functions for the *P-SV* and *SH* modes are obtained. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2434243]

PACS number(s): 43.40.At, 43.20.Gp, 43.20.Mv, 43.35.Cg [LLT]

Pages: 1440–1448

I. INTRODUCTION

Elastic wave propagation in multilayered media is an important research topic in nondestructive testing (NDT), ultrasonic imaging, film testing, geophysics, and other fields. Many studies have been carried out. The propagator matrix method is a very effective^{1–9} method for the elastic wave propagation in multilayered medium. This method is succinct and has been used by many researchers. The original propagator matrix has the problem of precision in the high frequency range. Abo-zena¹⁰ and Menke¹¹ developed a novel method which can increase the computable frequency range greatly. However, this method cannot avoid the problem of precision loss in the higher frequency range. Some other methods^{12–17} were presented for the problem of precision loss in the high frequency range. However, most of them lose the simplicity of the original propagator matrix. Another quite different method, the alternative method, e.g., the global matrix method, was proposed.^{18–20} Such technique can completely avoid the problem of precision loss in the high frequency range and be implemented simply but may be relatively slow to compute.

In ultrasonic NDT, elastic waves are usually generated by a piezoelectric transducer on the medium surface. The transducer is excited by an electric signal and vibrates along its thickness and/or width directions. Miller²¹ studied the elastic waves excited by a piezoelectric strip with infinite

length and a piezoelectric circular disk on the free surface of a homogeneous half space. This can be generalized to the multilayered media. The multilayered media is a one-dimensional structure. The elastic waves excited by an arbitrary source in multilayered media usually distribute in three-dimensional (3D) space. The previous studies on elastic waves excited by a piezoelectric source in the multilayered medium are all almost localized in a two-dimensional (2D) distribution field. Even for a 3D field, the previous studies are established on the condition that the system formed by the multilayered medium and the source has an axial symmetry and the field can be simply represented in the cylindrical coordinate oriented along the symmetric axis of the system.

Whichever methods of the propagator matrix have been used, even if the global matrix method, few authors opted to study the elastic waves excited by a plane piezoelectric transducer (for example, rectangular transducer) on the free surface of a multilayered medium. Although many papers about the elastic waves in multilayered medium have been published in recent years,^{22–26} there are no papers focused on the problem of the excitation of elastic waves by a plane source in a multilayered medium. The study of this problem is more and more important with the rapid development of sciences and technology and will become a more interesting research direction. The excitation theory by an arbitrary plane source on the free surface of a multilayered medium will provide an important theoretical foundation for NDT and other application fields.

^{a)}Electronic mail: zhbz@mail.ioa.ac.cn

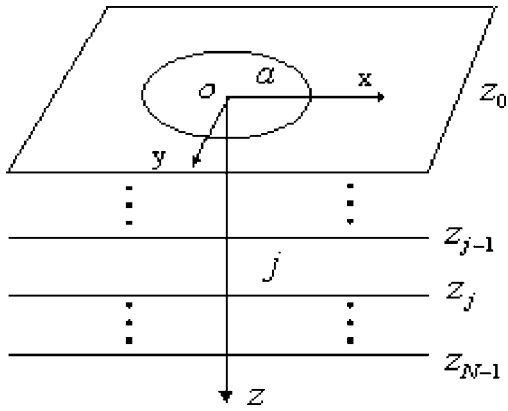


FIG. 1. Configuration of a multilayered medium.

The motivation for this paper is to give a study on the excitation theory to the plane source with arbitrary shape on the free surface of a multilayered medium. The propagator matrix method of Abo-zena,¹⁰ Menke,¹¹ and Zhang¹⁷ is adopted in this paper. Although other matrix methods can also be adopted, the methods of Abo-zena,¹⁰ Menke,¹¹ and Zhang¹⁷ are not only very simple and allow for fast calculation, but also can avoid the problem of precision loss in the general frequency range.

In Sec. II, the propagator matrix method of Abo-zena,¹⁰ Menke,¹¹ and Zhang¹⁷ is extended from 2D to 3D space. In Secs. III and IV, the displacement field is studied in the wave number and spatial domains, respectively. The *P*-*SV* and *SH* modes corresponding to the poles are investigated in Sec. V. Finally, as a typical example, the modes excited by a rectangular source are analyzed in Sec. VI.

II. PROPAGATOR MATRIX

A semi-infinite medium consisted of N parallel, homogeneous, isotropic, elastic solid layers are considered. A Cartesian coordinate system (x, y, z) is adopted. The positive direction of the z axis vertically points to the medium interior. $z=0$ is the free surface of the first layer of the medium (Fig. 1). The j th layer is bounded by the planes $z=z_{j-1}$ and $z=z_j$. For the j th layer with the thickness h_j , its properties are denoted by V_{pj} (*P* wave velocity), V_{sj} (*S* wave velocity), and ρ_j (density). The argument (j) will be omitted whenever it is possible without causing confusion.

The displacement potentials φ (*P* wave), ψ (*SV* wave), and χ (*SH* wave) are usually introduced by

$$\mathbf{u} = \nabla\varphi + \nabla \times (\chi \mathbf{e}_z) + \nabla \times \nabla \times (\psi \mathbf{e}_z). \quad (1)$$

The displacement potentials in the frequency domain satisfy (with the time-dependent factor $e^{-i\omega t}$)

$$(\nabla^2 + k_p^2)\varphi = 0, \quad (\nabla^2 + k_s^2)\chi = 0, \quad (\nabla^2 + k_s^2)\psi = 0, \quad (2)$$

where $k_p = \omega/V_p$, $k_s = \omega/V_s$, and ω is the angular frequency.

Introduce 2D Fourier transformation for x and y ,

$$\varphi(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(k_x, k_y, z) e^{i(k_x x + k_y y)} dk_x dk_y,$$

$$\psi(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(k_x, k_y, z) e^{i(k_x x + k_y y)} dk_x dk_y, \quad (3)$$

$$\chi(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(k_x, k_y, z) e^{i(k_x x + k_y y)} dk_x dk_y.$$

k_x and k_y are the wave numbers in the x and y directions. The elastic wave field will be analyzed in the wave number space (k_x, k_y) .

From Eqs. (2) and (3) it is easy to derive

$$\left(\frac{\partial^2}{\partial z^2} + k_{zp}^2 \right) \varphi(k_x, k_y, z) = 0,$$

$$\left(\frac{\partial^2}{\partial z^2} + k_{zs}^2 \right) \psi(k_x, k_y, z) = 0, \quad (4)$$

$$\left(\frac{\partial^2}{\partial z^2} + k_{zs}^2 \right) \chi(k_x, k_y, z) = 0,$$

where $k_{zp} = \sqrt{k_p^2 - k_r^2}$, $k_{zs} = \sqrt{k_s^2 - k_r^2}$, $k_r^2 = k_x^2 + k_y^2$, k_r is the wave number in the xoy plane, and k_{zp} and k_{zs} are the wave numbers of the *P* and *S* waves in the z axis.

Then, the displacement potentials can be written as

$$\varphi(k_x, k_y, z) = A e^{ik_{zp}z} + B e^{-ik_{zp}z} = \varphi^+ + \varphi^-,$$

$$\psi(k_x, k_y, z) = C e^{ik_{zs}z} + D e^{-ik_{zs}z} = \psi^+ + \psi^-, \quad (5)$$

$$\chi(k_x, k_y, z) = E e^{ik_{zs}z} + F e^{-ik_{zs}z} = \chi^+ + \chi^-,$$

where φ^+ , ψ^+ , and χ^+ represent waves propagating along the positive z axis, while φ^- , ψ^- , and χ^- represent waves propagating along the negative z axis.

Define the following vectors of *S* (motion stress) and Φ (displacement potentials):

$$S = \left(\frac{u_x}{ik_r}, \frac{u_z}{k_r}, \frac{\tau_{zz}}{\omega^2}, \frac{\tau_{zx}}{i\omega^2}, \frac{u_y}{ik_r}, \frac{\tau_{zy}}{i\omega^2} \right)^T, \quad (6)$$

$$\Phi = (\varphi^+, \varphi^-, \psi^+ k_r, \psi^- k_r, \chi^+, \chi^-)^T.$$

It can be found that the vector *S* and Φ satisfy

$$S = M\Phi, \quad (7)$$

$$M = \begin{pmatrix} l_x & l_x & -l_x \gamma_s & l_x \gamma_s & l_y & l_y \\ -\gamma_p & \gamma_p & 1 & 1 & 0 & 0 \\ \rho(\gamma-1) & \rho(\gamma-1) & -\rho\gamma\gamma_s & \rho\gamma\gamma_s & 0 & 0 \\ -\rho\gamma\gamma_p l_x & \rho\gamma\gamma_p l_x & \rho(\gamma-1)l_x & \rho(\gamma-1)l_x & -\rho\gamma\gamma_s l_y/2 & \rho\gamma\gamma_s l_y/2 \\ l_y & l_y & -l_y \gamma_s & l_y \gamma_s & -l_x & -l_x \\ -\rho\gamma\gamma_p l_y & -\rho\gamma\gamma_p l_y & \rho(\gamma-1)l_y & \rho(\gamma-1)l_y & \rho\gamma\gamma_s l_x/2 & -\rho\gamma\gamma_s l_x/2 \end{pmatrix}, \quad (8)$$

where $\gamma=2k_r^2/k_s^2$, $\gamma_p=\sqrt{k_r^2-k_p^2}/k_r$, $\gamma_s=\sqrt{k_r^2-k_s^2}/k_r$, $k_{zp}=ik_r\gamma_p$, $k_{zs}=ik_r\gamma_s$, $l_x=k_x/k_r$, $l_y=k_y/k_r$, and $l_x^2+l_y^2=1$. Equation (8) shows that the field distribution in 3D space is more complicated than that in 2D case.^{10,11,17,22}

If the wave propagates along the x axis ($l_y=0, l_x=1$), then

$$M_0 = M|_{l_y=0} = \begin{pmatrix} 1 & 1 & -\gamma_s & \gamma_s & 0 & 0 \\ -\gamma_p & \gamma_p & 1 & 1 & 0 & 0 \\ \rho(\gamma-1) & \rho(\gamma-1) & -\rho\gamma\gamma_s & \rho\gamma\gamma_s & 0 & 0 \\ -\rho\gamma\gamma_p & \rho\gamma\gamma_p & \rho(\gamma-1) & \rho(\gamma-1) & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & \rho\gamma\gamma_s/2 & -\rho\gamma\gamma_s/2 \end{pmatrix}, \quad (9)$$

where M_0 is the same as that in the 2D case.^{10,11,17}

It can be found that

$$M = aM_0, \quad (10)$$

$$a = \begin{pmatrix} l_x & 0 & 0 & 0 & -l_y & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & l_x & 0 & -l_y \\ l_y & 0 & 0 & 0 & l_x & 0 \\ 0 & 0 & 0 & l_y & 0 & l_x \end{pmatrix}. \quad (11)$$

This shows that the matrix M can be obtained from the matrix M_0 by the transformation which is to rotate the Cartesian coordinate system by an angle $\phi=\tan^{-1}(l_y/l_x)$ along the z axis. In this transformation, u_z and τ_{zz} are not changed while both (u_x, u_y) and (τ_{zx}, τ_{zy}) are linked by the coordinate trans-

formation matrix of the rotating angle ϕ along the z axis. Therefore, the inverse matrix a^{-1} should be the transformation rotating the angle $-\phi$,

$$a^{-1} = \begin{pmatrix} l_x & 0 & 0 & 0 & l_y & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & l_x & 0 & l_y \\ -l_y & 0 & 0 & 0 & l_x & 0 \\ 0 & 0 & 0 & -l_y & 0 & l_x \end{pmatrix}, \quad (12)$$

a and a^{-1} are not related to the parameters of the medium and only depend on the rotating angle.

The inverse matrix of M is then

$$M^{-1} = M_0^{-1}a^{-1}, \quad (13)$$

$$M_0^{-1} = \begin{pmatrix} \frac{\gamma}{2} & \frac{\gamma-1}{2\gamma_p} & -\frac{1}{2\rho} & -\frac{1}{2\rho\gamma_p} & 0 & 0 \\ \frac{\gamma}{2} & -\frac{\gamma-1}{2\gamma_p} & -\frac{1}{2\rho} & \frac{1}{2\rho\gamma_p} & 0 & 0 \\ \frac{\gamma-1}{2\gamma_s} & \frac{\gamma}{2} & -\frac{1}{2\rho\gamma_s} & -\frac{1}{2\rho} & 0 & 0 \\ -\frac{\gamma-1}{2\gamma_s} & \frac{\gamma}{2} & \frac{1}{2\rho\gamma_s} & -\frac{1}{2\rho} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{\rho\gamma\gamma_s} \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{\rho\gamma\gamma_s} \end{pmatrix}. \quad (14)$$

It can be seen that

$$\Phi(z_j) = \lambda \Phi(z_{j-1}), \quad (15)$$

where $\lambda = \text{diag}(P, P^{-1}, Q, Q^{-1}, Q, Q^{-1})$, $P = e^{-\gamma_p k_r h}$, $Q = e^{-\gamma_s k_r h}$, and $h = z_j - z_{j-1}$. So

$$\begin{aligned} S(z_j) &= M \Phi(z_j) = M \lambda \Phi(z_{j-1}) = M \lambda M^{-1} S(z_{j-1}) \\ &= p(z_j, z_{j-1}) S(z_{j-1}), \end{aligned} \quad (16)$$

$$p(z_j, z_{j-1}) = M \lambda M^{-1} = a p_0(z_j, z_{j-1}) a^{-1}, \quad (17)$$

$$p_0(z_j, z_{j-1}) = M_0 \lambda M_0^{-1}, \quad (18)$$

where $p(z_j, z_{j-1})$ and $p_0(z_j, z_{j-1})$ are the propagator matrices for 3D and 2D wave fields, respectively. Equation (17) is also valid in arbitrary two points in the multilayered medium because the matrix only depends on the rotating angle. It can be seen that $p_0(z_j, z_{j-1})$ is a quasideagonal matrix while the matrix $p(z_j, z_{j-1})$ is not in general.

The vectors S in the planes $z=0$ and $z=z_{N-1}$ are related by

$$S(z_{N-1}) = p(z_{N-1}, z_0) S(z_0) = a p_0(z_{N-1}, z_0) a^{-1} S(z_0). \quad (19)$$

III. DISPLACEMENT FIELD

The elastic wave field approaches zero as z becomes infinite. This demands that there are no waves propagating along the negative z axis in the N th layer medium. Therefore,

$$\begin{aligned} \begin{pmatrix} \phi^+ \\ 0 \\ \psi^+ k_r \\ 0 \\ \chi^+ \\ 0 \end{pmatrix}_{z=z_{N-1}+0} &= M_N^{-1} p(z_{N-1}, z_0) S(z_0) \\ &= (M_0)_N^{-1} p_0(z_{N-1}, z_0) a^{-1} S(z_0). \end{aligned} \quad (20)$$

It is convenient to introduce

$$\begin{pmatrix} \frac{u_x}{ik_r} \\ \frac{u_z}{k_r} \\ \frac{u_y}{ik_r} \end{pmatrix} = \begin{pmatrix} \frac{E_3^{(1)}}{E_6^{(1)}} l_x & -\frac{E_2^{(1)}}{E_6^{(1)}} l_x^2 - \frac{H_{66}}{H_{65}} l_y^2 & \left(-\frac{E_2^{(1)}}{E_6^{(1)}} + \frac{H_{66}}{H_{65}}\right) l_x l_y \\ \frac{E_5^{(1)}}{E_6^{(1)}} & \frac{E_3^{(1)}}{E_6^{(1)}} l_x & \frac{E_3^{(1)}}{E_6^{(1)}} l_y \\ \frac{E_3^{(1)}}{E_6^{(1)}} l_y & \left(-\frac{E_2^{(1)}}{E_6^{(1)}} + \frac{H_{66}}{H_{65}}\right) l_x l_y & -\frac{E_2^{(1)}}{E_6^{(1)}} l_y^2 - \frac{H_{66}}{H_{65}} l_x^2 \end{pmatrix} \times \begin{pmatrix} \frac{\tau_{zz}}{\omega^2} \\ \frac{\tau_{zx}}{i\omega^2} \\ \frac{\tau_{zy}}{i\omega^2} \end{pmatrix}. \quad (26)$$

The elements with denominators $E_6^{(1)}$ and H_{65} in the matrix correspond to the P - SV and SH waves, respectively. This

$$\begin{aligned} W &= a^{-1} S(z_0) \\ &= \begin{pmatrix} \frac{l_x u_x + l_y u_y}{ik_r}, \frac{u_z}{k_r}, \frac{\tau_{zz}}{\omega^2}, \frac{l_x \tau_{zx} + l_y \tau_{zy}}{i\omega^2}, \frac{l_x u_y - l_y u_x}{ik_r}, \frac{l_x \tau_{zy} - l_y \tau_{zx}}{i\omega^2} \end{pmatrix}^T, \end{aligned} \quad (21)$$

and

$$H = (M_0)_N^{-1} p_0(z_{N-1}, z_0). \quad (22)$$

The matrix H can be written as the following form:

$$H = \begin{pmatrix} H_{11} & H_{12} & H_{13} & H_{14} & 0 & 0 \\ H_{21} & H_{22} & H_{23} & H_{24} & 0 & 0 \\ H_{31} & H_{32} & H_{33} & H_{34} & 0 & 0 \\ H_{41} & H_{42} & H_{43} & H_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & H_{55} & H_{56} \\ 0 & 0 & 0 & 0 & H_{65} & H_{66} \end{pmatrix}, \quad (23)$$

where $H_{ij}(i \leq 4, j \leq 4)$ and $H_{ij}(i > 4, j > 4)$ correspond to the P - SV and SH waves, respectively.^{11,17} Equation (20) shows that

$$\begin{pmatrix} H_{21} & H_{22} & 0 \\ H_{41} & H_{42} & 0 \\ 0 & 0 & H_{65} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ W_5 \end{pmatrix} + \begin{pmatrix} H_{23} & H_{24} & 0 \\ H_{43} & H_{44} & 0 \\ 0 & 0 & H_{66} \end{pmatrix} \begin{pmatrix} W_3 \\ W_4 \\ W_6 \end{pmatrix} = 0. \quad (24)$$

The matrix H is the same as that in the 2D case. It can be verified by previous papers^{11,17} that

$$\begin{pmatrix} W_1 \\ W_2 \\ W_5 \end{pmatrix} = \begin{pmatrix} \frac{E_3^{(1)}}{E_6^{(1)}} & -\frac{E_2^{(1)}}{E_6^{(1)}} & 0 \\ \frac{E_5^{(1)}}{E_6^{(1)}} & \frac{E_3^{(1)}}{E_6^{(1)}} & 0 \\ 0 & 0 & -\frac{H_{66}}{H_{65}} \end{pmatrix} \begin{pmatrix} W_3 \\ W_4 \\ W_6 \end{pmatrix}, \quad (25)$$

where the meanings of $E_2^{(1)}$, $E_3^{(1)}$, $E_5^{(1)}$, and $E_6^{(1)}$ are given in Refs. 11 and 17. Hence, it can be obtained by Eq. (21) that

shows that the P - SV and SH wave fields are all contained in the displacement field.

Therefore, if the field is excited only by the component τ_{zz} of the stress, the displacement components on the free surface $z=0$ are

$$u_x = \frac{ik_r E_3^{(1)}}{\omega^2 E_6^{(1)}} l_x \tau_{zz}, \quad u_y = \frac{ik_r E_3^{(1)}}{\omega^2 E_6^{(1)}} l_y \tau_{zz}, \quad u_z = \frac{k_r E_5^{(1)}}{\omega^2 E_6^{(1)}} \tau_{zz}, \quad (27)$$

and if the field is excited only by τ_{zx} , then

$$u_x = - \left(\frac{E_2^{(1)}}{E_6^{(1)}} l_x^2 + \frac{H_{66}}{H_{65}} l_y^2 \right) \frac{k_r}{\omega^2} \tau_{zx},$$

$$u_y = \left(- \frac{E_2^{(1)}}{E_6^{(1)}} + \frac{H_{66}}{H_{65}} \right) l_x l_y \frac{k_r}{\omega^2} \tau_{zx}, \quad (28)$$

$$u_z = \frac{k_r E_3^{(1)}}{i \omega^2 E_6^{(1)}} l_x \tau_{zx}.$$

Equations (27) and (28) display that only the *P-SV* wave is contained in the field in the case of τ_{zz} excitation while the *P-SV* and *SH* waves are all contained in the case of τ_{zx} excitation.

The displacement and stress in an arbitrary point in the multilayered medium can be obtained by the propagator matrix when the displacement and stress on the free surface are determined. Only the displacement in the plane $z=0$ is studied in the following sections. The wave field in the case $z > 0$ has a similar result.

IV. THE SPATIAL DISTRIBUTION OF THE DISPLACEMENT FIELD

The displacement field considered in Sec. III is in the wave number domain. This section will study the displacement field in the spatial domain.

First, consider the case where the field is excited only by τ_{zz} . By Eq. (27), the components of the displacement in the spatial domain on the plane $z=0$ can be written as

$$u_x(x, y) = \frac{i}{\omega^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{k_r E_3^{(1)}}{E_6^{(1)}} l_x \tau_{zz} e^{i(k_x x + k_y y)} dk_x dk_y,$$

$$u_y(x, y) = \frac{i}{\omega^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{k_r E_3^{(1)}}{E_6^{(1)}} l_y \tau_{zz} e^{i(k_x x + k_y y)} dk_x dk_y, \quad (29)$$

$$u_z(x, y) = \frac{1}{\omega^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{k_r E_5^{(1)}}{E_6^{(1)}} \tau_{zz} e^{i(k_x x + k_y y)} dk_x dk_y.$$

It is convenient to find the integrations in the polar coordinate system of the wave number domain,

$$u_x(r, \theta) = \frac{i}{\omega^2} \int_0^{\infty} \frac{k_r^2 E_3^{(1)}}{E_6^{(1)}} dk_r \int_{-\pi}^{\pi} \tau_{zz}(k_r, \varphi) \times \cos \varphi e^{ik_r r \cos(\varphi - \theta)} d\varphi,$$

$$u_y(r, \theta) = \frac{i}{\omega^2} \int_0^{\infty} \frac{k_r^2 E_3^{(1)}}{E_6^{(1)}} dk_r \int_{-\pi}^{\pi} \tau_{zz}(k_r, \varphi) \times \sin \varphi e^{ik_r r \cos(\varphi - \theta)} d\varphi, \quad (30)$$

$$u_z(r, \theta) = \frac{1}{\omega^2} \int_0^{\infty} \frac{k_r^2 E_5^{(1)}}{E_6^{(1)}} dk_r \int_{-\pi}^{\pi} \tau_{zz}(k_r, \varphi) e^{ik_r r \cos(\varphi - \theta)} d\varphi,$$

where (r, θ) and (k_r, φ) are the polar coordinates in the spatial and wave number domains, respectively. The compressional wave, shear wave, and the modes corresponding to the poles in the integrand are all contained in Eq. (30).

It is convenient to change the lower limit of the integration k_r from 0 to $-\infty$ and constitute an integration path from $-\infty$ to ∞ for k_r , as given in the following:

$$u_r(r, \theta) = \frac{i}{\omega^2} \int_{-\pi/2}^{\pi/2} \cos \varphi d\varphi \int_{-\infty}^{\infty} \frac{k_r^2 E_3^{(1)}}{E_6^{(1)}} \tau_{zz}(k_r, \theta + \varphi) \times e^{ik_r r \cos \varphi} dk_r,$$

$$u_{\theta}(r, \theta) = \frac{i}{\omega^2} \int_{-\pi/2}^{\pi/2} \sin \varphi d\varphi \int_{-\infty}^{\infty} \frac{k_r^2 E_3^{(1)}}{E_6^{(1)}} \tau_{zz}(k_r, \theta + \varphi) \times e^{ik_r r \cos \varphi} dk_r, \quad (31)$$

$$u_z(r, \theta) = \frac{1}{\omega^2} \int_{-\pi/2}^{\pi/2} d\varphi \int_{-\infty}^{\infty} \frac{k_r^2 E_5^{(1)}}{E_6^{(1)}} \tau_{zz}(k_r, \theta + \varphi) e^{ik_r r \cos \varphi} dk_r.$$

It is easy to derive Eq. (31).

Then, the case of τ_{zx} excitation is studied. In this case, the spatial displacement on the plane $z=0$, which is transformed from Eq. (28), can be obtained by the similar procedure course and can be written as

$$u_x(r, \theta) = - \frac{1}{\omega^2} \int_{-\pi/2}^{\pi/2} d\varphi \int_{-\infty}^{\infty} \left(\frac{E_2^{(1)}}{E_6^{(1)}} \cos^2(\theta + \varphi) + \frac{H_{66}}{H_{65}} \sin^2(\theta + \varphi) \right) k_r^2 \tau_{zx}(k_r, \theta + \varphi) e^{ik_r r \cos \varphi} dk_r,$$

$$u_y(r, \theta) = - \frac{1}{\omega^2} \int_{-\pi/2}^{\pi/2} \cos(\theta + \varphi) \sin(\theta + \varphi) d\varphi \times \int_{-\infty}^{\infty} \left(\frac{E_2^{(1)}}{E_6^{(1)}} - \frac{H_{66}}{H_{65}} \right) k_r^2 \tau_{zx}(k_r, \theta + \varphi) e^{ik_r r \cos \varphi} dk_r, \quad (32)$$

$$u_z(r, \theta) = \frac{1}{i \omega^2} \int_{-\pi/2}^{\pi/2} \cos(\theta + \varphi) d\varphi \int_{-\infty}^{\infty} \frac{E_3^{(1)}}{E_6^{(1)}} k_r^2 \tau_{zx}(k_r, \theta + \varphi) \times e^{ik_r r \cos \varphi} dk_r.$$

Equation (32) shows again that the *P-SV* and *SH* fields all appeared in this case.

V. MODES CORRESPONDING TO THE POLES

In this section, the contribution of the modes corresponding to the poles in the integrand of the field equations

(31) and (32) is studied. The dispersion equations and the displacement components for all modes are obtained.

The integrations in Eqs. (31) and (32) for the horizontal wave number k_r are from $-\infty$ to ∞ , which is along the whole real axis in the complex plane k_r . Let f be the integrand in Eqs. (31) and (32) and construct a closed integration counter $\int_{-R}^R f dk_r + \int_{C_R} f dk_r$ on the real axis and in the upper half plane. Then, let $R \rightarrow \infty$ and C_R tend to infinity in the upper half plane. If the integration

$$\int_{C_R} f dk_r \rightarrow 0, \quad (33)$$

then the integration contribution of the modes corresponding to the poles in the upper half plane and on the real axis can be given by their residues.

It is easy to find by Eqs. (31) and (32) that condition (33) is satisfied in general because of the exponential factor $e^{ik_r \cos \varphi}$ ($\varphi \in [-\pi/2, \pi/2]$) in the integrand. Condition (33) should be satisfied in the physical sense for an excitation source distributed in a finite range.

In the τ_{zz} excitation, only the modes corresponding to the P - SV wave field exist and the displacement components can be obtained by the residue of the poles,

$$\begin{aligned} u_r^{P-SV}(r, \theta) &= -\frac{2\pi k_n^2 E_3^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_z(k_n, \theta + \varphi) \cos \varphi \\ &\quad \times e^{ik_n r \cos \varphi} d\varphi, \\ u_\theta^{P-SV}(r, \theta) &= -\frac{2\pi k_n^2 E_3^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_z(k_n, \theta + \varphi) \sin \varphi \\ &\quad \times e^{ik_n r \cos \varphi} d\varphi, \\ u_z^{P-SV}(r, \theta) &= -\frac{2\pi i k_n^2 E_5^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_z(k_n, \theta + \varphi) \\ &\quad \times e^{ik_n r \cos \varphi} d\varphi, \end{aligned} \quad (34)$$

where k_n ($n=1, 2, 3, \dots$) is a given value that is determined by the poles and related to the source frequency. That is to say, k_n is the n th root of

$$E_6^{(1)} = 0, \quad (35)$$

about the variable k_r . Equation (35) is the dispersion equation, which does not depend on the propagation azimuth angle in the horizontal plane, of the P - SV modes. It shows that the dispersion characteristic of the modes is not related to the propagation direction in the horizontal plane. However, the displacement of the modes, which is given by Eq. (34), is strongly dependent on the propagation azimuth angle θ .

In the case of the τ_{zx} excitation, it is indicated by Eq. (32) that it contains not only the P - SV modes but also the SH modes. For the P - SV modes, the dispersion equation is the same as Eq. (35) and the displacement components can be written as

$$\begin{aligned} u_r^{P-SV}(r, \theta) &= -\frac{2\pi i k_n^2 E_2^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_{zx}(k_n, \theta + \varphi) \\ &\quad \times \cos(\theta + \varphi) \cos \varphi e^{ik_n r \cos \varphi} d\varphi, \\ u_\theta^{P-SV}(r, \theta) &= -\frac{2\pi i k_n^2 E_2^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_{zx}(k_n, \theta + \varphi) \\ &\quad \times \cos(\theta + \varphi) \sin \varphi e^{ik_n r \cos \varphi} d\varphi, \\ u_z^{P-SV}(r, \theta) &= \frac{2\pi k_n^2 E_3^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_{zx}(k_n, \theta + \varphi) \\ &\quad \times \cos(\theta + \varphi) e^{ik_n r \cos \varphi} d\varphi. \end{aligned} \quad (36)$$

For the SH modes, the dispersion equation is

$$H_{65} = 0, \quad (37)$$

and the displacement components can be obtained by Eq. (32),

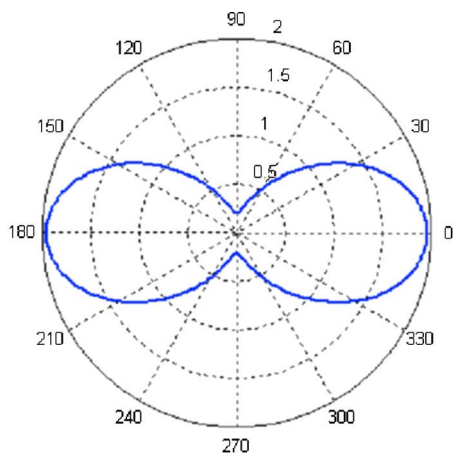
$$\begin{aligned} u_r^{SH}(r, \theta) &= -\frac{2\pi i k_n^2 H_{66}}{\omega^2 \partial H_{65} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_{zx}(k_n, \theta + \varphi) \\ &\quad \times \sin(\theta + \varphi) \sin \varphi e^{ik_n r \cos \varphi} d\varphi, \\ u_\theta^{SH}(r, \theta) &= \frac{2\pi i k_n^2 H_{66}}{\omega^2 \partial H_{65} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} \tau_{zx}(k_n, \theta + \varphi) \\ &\quad \times \sin(\theta + \varphi) \cos \varphi e^{ik_n r \cos \varphi} d\varphi, \\ u_z^{SH}(r, \theta) &= 0. \end{aligned} \quad (38)$$

It is also shown that the dispersion characteristic of the SH modes is not related to the propagation azimuth angle while the displacement is strongly dependent on the azimuth angle θ .

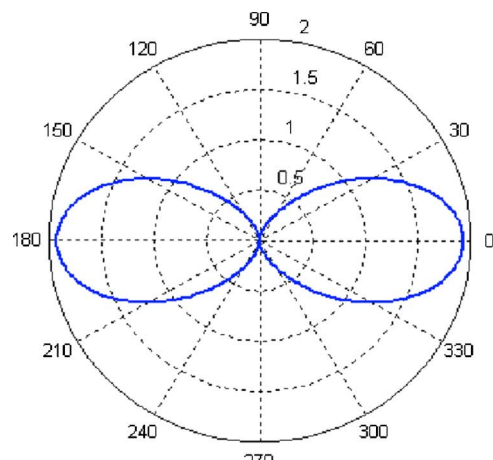
It can be seen by Eqs. (34), (36), and (38) that the displacements of the modes can be looked as a product of two factors. The first factor is the part before the integration sign and the second part is the integration to the text after the integration sign. The first factor is related to the parameters of the medium and the mode characteristic. The second factor is related to the excitation fashion and the shape of the source and independent of the parameters of the medium. For a given value of $k_n r$, the second factor is just the function of azimuth angle θ . So the second factor can be defined as the directional distribution function in the plane parallel to the free surface of the multilayered medium. The directional distribution function is independent of the parameters of the medium and the characteristic of the modes and dependent on the source.

VI. AN EXAMPLE: RECTANGULAR SOURCE EXCITATION

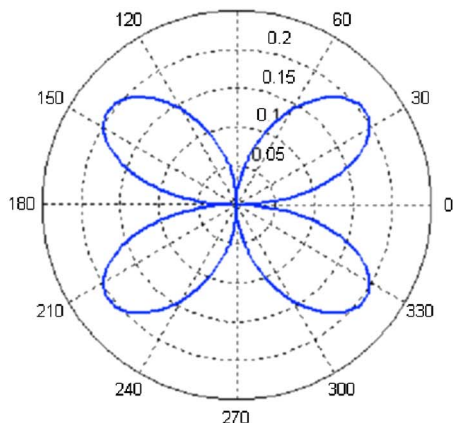
In this section, as an example, the displacement representations of the P - SV and SH modes excited by a rectangular source on the free surface of a multilayered medium are derived.



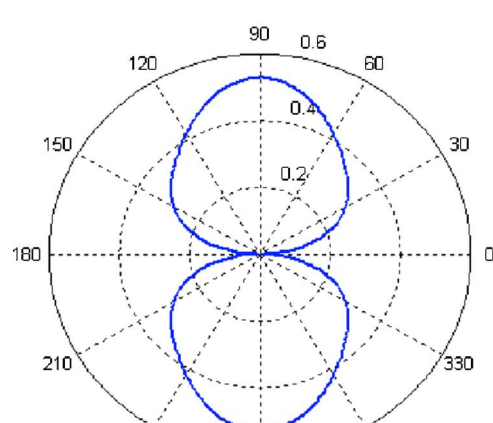
(a)



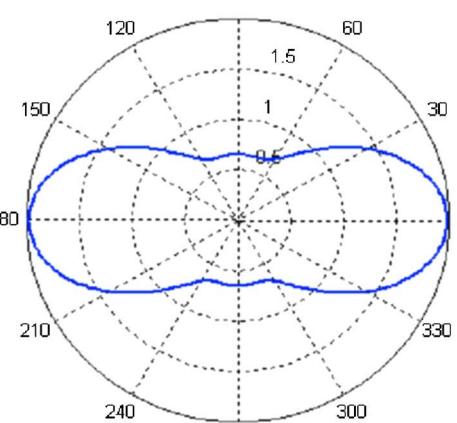
(a)



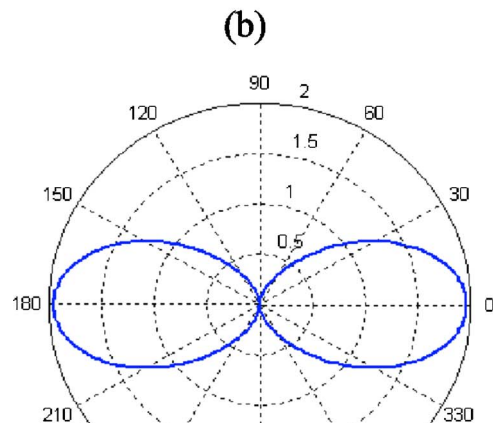
(b)



(b)



(c)



(c)

FIG. 2. (Color online) Directivity distributions of the displacement components u_r (a), u_θ (b), and u_z (c) of the P - SV modes excited by a rectangular source with τ_{zz} excitation. $k_n a=1$, $k_n b=3$, and $k_n r=10$.

A rectangular source with distribution range S_0 located in $|x| \leq a$ and $|y| \leq b$ is considered. In τ_{zz} excitation, it is assumed that the stress component τ_{zz} on the surface $z=0$ satisfies

$$\tau_{zz}(x,y) = \begin{cases} 1, & (x,y) \in S_0, \\ 0, & (x,y) \notin S_0, \end{cases} \quad (39)$$

and the other stress components are zeros on the free surface.

FIG. 3. (Color online) Directivity distributions of the displacement components u_r (a), u_θ (b), and u_z (c) of the P - SV modes excited by a rectangular source with τ_{zx} excitation. $k_n a=1$, $k_n b=3$, and $k_n r=10$.

It is easy to obtain

$$\tau_{zz}(k_r, \varphi) = \int_{-a}^a e^{-ik_x x} dx \int_{-b}^b e^{-ik_y y} dy = 4T(\varphi)/k_r^2, \quad (40)$$

where

$$T(\varphi) = \frac{\sin(k_r a \cos \varphi) \sin(k_r b \sin \varphi)}{\cos \varphi \sin \varphi}. \quad (41)$$

Therefore, the displacement components of the modes are

$$\begin{aligned} u_r^{P-SV}(r, \theta) &= -\frac{8\pi E_3^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \\ &\times \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \cos \varphi e^{ik_n r \cos \varphi} d\varphi, \\ u_\theta^{P-SV}(r, \theta) &= -\frac{8\pi E_3^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \\ &\times \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \sin \varphi e^{ik_n r \cos \varphi} d\varphi, \quad (42) \\ u_z^{P-SV}(r, \theta) &= \frac{8\pi i E_5^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) e^{ik_n r \cos \varphi} d\varphi. \end{aligned}$$

Equation (42) shows that the intensity of the modes is strongly dependent on the propagation direction θ . Figure 2 plots the directivity distribution functions of the displacement components u_r , u_θ , and u_z for the P - SV modes and illustrates the relation between the displacement and the propagation direction θ . In Fig. 2, $k_n a = 1$, $k_n b = 3$, and $k_n r = 10$. It can be seen that the displacement components are strongly dependent on the propagation direction and u_θ is zero in the directions $\theta = 0, \pi/2, \pi$, and $3\pi/2$.

In the case of τ_{zx} excitation, the stress component $\tau_{zx}(x, y)$ on the free surface is 1 inside the range S_0 and 0 outside S_0 , then $\tau_{zx}(k_r, \varphi) = 4T(\varphi)/k_r^2$. Therefore, the displacements of the P - SV and SH modes are

$$\begin{aligned} u_r^{P-SV}(r, \theta) &= -\frac{8\pi i E_2^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \\ &\times \cos(\theta + \varphi) \cos \varphi e^{ik_n r \cos \varphi} d\varphi, \\ u_\theta^{P-SV}(r, \theta) &= -\frac{8\pi i E_2^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \\ &\times \cos(\theta + \varphi) \sin \varphi e^{ik_n r \cos \varphi} d\varphi, \quad (43) \\ u_z^{P-SV}(r, \theta) &= \frac{8\pi E_3^{(1)}}{\omega^2 \partial E_6^{(1)} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \\ &\times \cos(\theta + \varphi) e^{ik_n r \cos \varphi} d\varphi, \end{aligned}$$

and

$$\begin{aligned} u_r^{SH}(r, \theta) &= -\frac{8\pi i H_{66}}{\omega^2 \partial H_{65} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \\ &\times \sin(\theta + \varphi) \sin \varphi e^{ik_n r \cos \varphi} d\varphi, \end{aligned}$$

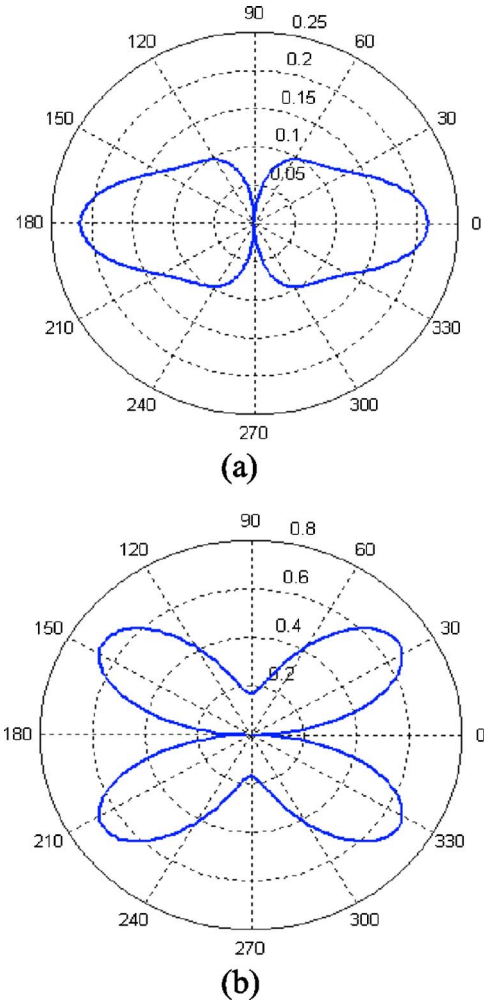


FIG. 4. (Color online) Directivity distributions of the displacement components u_r (a) and u_θ (b) of the SH modes excited by a rectangular source with τ_{zx} excitation. $k_n a = 1$, $k_n b = 3$, and $k_n r = 10$.

$$\begin{aligned} u_\theta^{SH}(r, \theta) &= \frac{8\pi i H_{66}}{\omega^2 \partial H_{65} / \partial k_r |_{k_r=k_n}} \int_{-\pi/2}^{\pi/2} T(\theta + \varphi) \\ &\times \sin(\theta + \varphi) \cos \varphi e^{ik_n r \cos \varphi} d\varphi, \quad (44) \end{aligned}$$

$$u_z^{SH}(r, \theta) = 0,$$

respectively. Figures 3 and 4 show the directivity distribution functions of the P - SV and SH modes, respectively. The parameters are the same as in Fig. 2. All the components of the displacement are dependent on the propagation direction. For the P - SV modes (Fig. 3), u_r and u_z are zeros when $\theta = \pi/2$ and $3\pi/2$ while u_θ is zero when $\theta = 0$ and π . For the SH modes (Fig. 4), u_r is zero when $\theta = \pi/2$ and $3\pi/2$ while u_θ is zero when $\theta = 0$ and π .

In τ_{zz} and τ_{zx} excitations, the directivity distribution of the P - SV and SH modes excited by the rectangular source is related to the value of $k_n r$. If the distance r from the receive point to the center of the source is changed (k_n is fixed), the different directivity distribution will be obtained. However, for the farfield (r is much larger than the wavelength and the size of the rectangular source), the directivity distribution will be independent of the distance r .

VII. CONCLUSIONS

Elastic wave field excited by a 2D plane source on the surface of a multilayered medium has been investigated. The propagator matrix for 3D elastic wave field is obtained and the propagation of elastic waves is analyzed further. Emphasis is given to analyzing P - SV and SH modes corresponding to the poles and giving the displacement representations for the modes excited by a rectangular source on the free surface of the multilayered medium.

It is found that the propagation velocities of the P - SV and SH modes do not depend on the propagation azimuth in the plane parallel to the free surface while the displacement amplitudes of the modes are strongly dependent on the propagation azimuth. The directional distribution functions of the modes are independent of the parameters of the medium and the modes and dependent on the source. It shows that the modes have greater intensities in some directions and are weak (maybe cannot be detected by receivers) in other directions. In order to receive the waves corresponding to the modes effectively, one should consider the azimuth of the receivers.

In NDT, the transducer is on the free surface of the tested object to excite elastic waves. In fact, the transducers are all 2D structure in many cases. However, the conventional studies about elastic waves in multilayered media are all focused on 1D transducer or axial symmetric sources. This work gives a method for studying the elastic wave field excited by a 2D transducer and provides a theoretical foundation for 2D ultrasonic array imaging in multilayered media.

ACKNOWLEDGMENTS

The author is very grateful to the reviewers for their comments. This work is supported by the National Natural Science Foundation of China under Grant No. 10374098.

¹W. T. Thomson, "Transmission of elastic waves through a stratified solid," *J. Appl. Phys.* **21**, 89–93 (1950).

²N. A. Haskell, "The dispersion of surface waves on multilayered media," *Bull. Seismol. Soc. Am.* **43**, 17–34 (1953).

³D. G. Harkrider, "Surface waves in multilayered elastic media, I. Rayleigh-Love waves from buried sources in a multilayered elastic half-space," *Bull. Seismol. Soc. Am.* **54**, 627–679 (1964).

⁴R. Kind, "Computation of reflection coefficients for layered media," *J.*

Geophys. **42**, 191–200 (1976).

⁵R. Kind, "The reflectivity method for a buried source," *J. Geophys.* **44**, 603–612 (1978).

⁶M. Sauto, "Excitation of free oscillations and surface waves by a point source in a vertically heterogeneous earth," *J. Geophys. Res.* **72**, 3689–3899 (1967).

⁷F. Gilbert and G. Backus, "Propagator matrices in elastic wave and vibration problems," *Geophysics* **31**, 326–332 (1966).

⁸E. N. Throver, "The computation of the dispersion of elastic waves in layered media," *J. Sound Vib.* **2**, 210–226 (1965).

⁹F. Schwab and L. Knoff, "Surface waves on multilayered anelastic media," *Bull. Seismol. Soc. Am.* **61**, 893–912 (1971).

¹⁰A. Abo-zena, "Dispersion function computation for unlimited frequency values," *Geophys. J. R. Astron. Soc.* **58**, 91–105 (1979).

¹¹W. Menke, "Comment on 'Dispersion function computation for unlimited frequency values' by A. Abo-zena," *Geophys. J. R. Astron. Soc.* **59**, 315–323 (1979).

¹²H. Schmidt and G. Tango, "Efficient global matrix approach to the computation of synthetic seismograms," *Geophys. J. R. Astron. Soc.* **84**, 331–359 (1986).

¹³A. Ben-Menahem and S. J. Singh, "Multipolar elastic field in a layered half-space," *Bull. Seismol. Soc. Am.* **58**, 1519–1572 (1968).

¹⁴J. A. Hudson, "A quantitative evaluation of seismic signals at teleseismic distance. II. Body waves and surface waves from an extended source," *Geophys. J. R. Astron. Soc.* **18**, 353–370 (1969).

¹⁵B. L. N. Kennet and N. J. Kerry, "Seismic waves in a stratified half space," *Geophys. J. R. Astron. Soc.* **57**, 557–583 (1979).

¹⁶B. L. N. Kennet and T. J. Clarke, "Rapid calculation of surface wave dispersion," *Geophys. J. R. Astron. Soc.* **72**, 619–631 (1983).

¹⁷B. Zhang, M. Yu, C. Q. Lan, and W. Xiong, "Elastic wave and excitation mechanism of surface waves in multilayered media," *J. Acoust. Soc. Am.* **100**, 3257–3538 (1996).

¹⁸L. Knopoff, "A matrix method for elastic wave problems," *Bull. Seismol. Soc. Am.* **54**, 431–438 (1964).

¹⁹T. Kundn and A. K. Mal, "Elastic wave in a multilayered solid due to a dislocation source," *Wave Motion* **7**, 459–471 (1985).

²⁰A. K. Mal, "Guided waves in layered solids with interface zones," *Int. J. Eng. Sci.* **26**, 873–881 (1988).

²¹G. F. Miller and H. Pursey, "The field and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid," *Proc. R. Soc. London, Ser. A* **223**, 521–541 (1954).

²²B. Zhang, W. Xiong, M. Yu, C. Q. Lan, and Long Li, "Study of energy distribution of guided waves in multilayered media," *J. Acoust. Soc. Am.* **103**, 125–135 (1998).

²³I. A. Kaibichev and V. G. Shavrov, "Transverse waves in inhomogeneous layer between two media," *Acoust. Phys.* **45**, 71–75 (1999).

²⁴N. Ryden and M. J. S. Lowe, "Guided wave propagation in three-layer pavement structures," *J. Acoust. Soc. Am.* **116**, 2902–2913 (2004).

²⁵D. De Nil, "Characteristics of surface waves in media with significant vertical variations in elastic-dynamic properties," *J. Environ. Eng. Geophys.* **10**, 263–274 (2005).

²⁶M. J. S. Lowe, "Matrix technique for modeling ultrasonic waves in multilayered media," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **42**, 525–542 (1995).

Simulation of guided waves in complex piping geometries using the elastodynamic finite integration technique

Kevin E. Rudd

Department of Applied Science, College of William and Mary, Williamsburg, Virginia 23187-8795

Kevin R. Leonard

Solers, Inc., 3811 Fairfax Drive, Suite 950, Arlington, Virginia 22203

Jill P. Bingham

Department of Applied Science, College of William and Mary, Williamsburg, Virginia 23187-8795

Mark K. Hinders^{a)}

Department of Applied Science, College of William and Mary, Williamsburg, Virginia 23187-8795

(Received 4 May 2006; revised 13 November 2006; accepted 8 December 2006)

Although many technologies exist for inspecting piping systems, they are most successful on straight pipes and are often unable to accommodate the added complexities of pipe elbows, bends, twists, and branches, particularly if the region of interest is inaccessible. This paper presents a numerical technique based on the elastodynamic finite integration technique for simulating guided elastic wave propagation in piping systems. Comparisons show agreement between experimental and simulated data, and guided wave interaction with flaws, focusing, and propagation in pipe bends are presented. These examples demonstrate the ability of the simulation method to be used to study elastic wave propagation in piping systems which include three-dimensional pipe bends, and suggest its potential as a design tool for designing pipe inspection hardware and ultrasonic signal processing algorithms. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431335]

PACS number(s): 43.40.Le [RLW]

Pages: 1449–1458

INTRODUCTION

Ultrasonic guided waves have been used successfully for nondestructively inspecting a wide variety of structures.^{1–3} Using ultrasonic guided wave methods to inspect piping systems holds great promise^{4–8} but to successfully develop an ultrasonic guided wave system for remotely inspecting beyond pipe elbows it is necessary to understand in detail how elastic waves propagate through and beyond pipe bends. With sufficient computational resources, this can be accomplished efficiently using three-dimensional (3D) numerical simulations. We present a simulation method based on the elastodynamic finite integration technique (EFIT) that can model guided elastic wave propagation in pipe-like structures including 3D pipe bends. This method also allows us to accurately and systematically simulate the interaction of guided elastic waves with arbitrary flaws in these structures in order to optimize ultrasonic guided wave pipe inspection protocols.

Several simulation techniques exist for modeling elastic waves in pipe-like structures. Gsell *et al.* developed a finite-difference technique for modeling elastic waves in straight pipe-like structures based on the displacement-equations of motion.⁹ One advantage of this technique is its ability to model elastic wave interaction with subtle flaws due to its fine grid spacing. Leutenegger *et al.* showed how this method could be used to assist in locating defects in cylindrical structures.¹⁰ Hayashi *et al.* developed a semianalytical

finite-element (SAFE) technique for modeling elastic wave propagation in pipe-like structures including pipe bends.^{11,12} Their technique uses a relatively large spatial discretization which leads to fast computational times and allows for long pipe sections with multiple bends to be modeled. Unfortunately, this technique as described cannot be used to model guided wave interactions with subtle flaws. The finite integration technique is a powerful, accurate, and stable time-domain method for numerically solving partial differential equations. It has been used to model 2D, axially symmetric (2.5D) and full 3D elastic waves in the Cartesian and cylindrical coordinate systems.^{13–15}

We present a finite integration method for modeling elastic waves in pipe-like structures and pipe bends. We then show that this method compares well to experimental results. We also show how the fine spatial discretization allows guided elastic wave scattering from subtle flaws to be modeled. This simulation method can be used to design complicated hardware devices such as phased array transducer belts to focus the elastic wave energy on straight pipe sections and beyond pipe bends and to generate systematic data to test signal processing algorithms.

I. 3D CYLINDRICAL ELECTRODYNAMIC FINITE INTEGRATION TECHNIQUE (3DCEFIT)

We first describe how the finite integration technique is used to simulate 3D elastic waves in complex pipe-like structures. We present the equations necessary to simulate elastic waves in a pipe-bend and show how they can easily be adapted to model straight pipe sections.

^{a)}Corresponding Author. Electronic mail: hinders@as.wm.edu

A. Finite integration procedure

We begin with the nine equations for elastic wave propagation in solids using the cylindrical coordinate system:¹⁶

$$\rho \dot{v}_r = \frac{\partial T_{rr}}{\partial r} + \frac{1}{r} \frac{\partial T_{r\phi}}{\partial \phi} + \frac{\partial T_{rz}}{\partial z} + \frac{T_{rr} - T_{\phi\phi}}{r} + \rho f_r, \quad (1)$$

$$\rho \dot{v}_z = \frac{\partial T_{rz}}{\partial r} + \frac{1}{r} \frac{\partial T_{\phi z}}{\partial \phi} + \frac{\partial T_{zz}}{\partial z} + \frac{1}{r} T_{rz} + \rho f_z, \quad (2)$$

$$\rho \dot{v}_\phi = \frac{\partial T_{r\phi}}{\partial r} + \frac{1}{r} \frac{\partial T_{\phi\phi}}{\partial \phi} + \frac{\partial T_{\phi z}}{\partial z} + \frac{2}{r} T_{r\phi} + \rho f_\phi, \quad (3)$$

$$\dot{T}_{rr} = (\lambda + 2\mu) \frac{\partial v_r}{\partial r} + \lambda \left(\frac{v_r}{r} + \frac{1}{r} \frac{\partial v_\phi}{\partial \phi} + \frac{\partial v_z}{\partial z} \right), \quad (4)$$

$$\dot{T}_{\phi\phi} = (\lambda + 2\mu) \left(\frac{v_r}{r} + \frac{1}{r} \frac{\partial v_\phi}{\partial \phi} \right) + \lambda \left(\frac{\partial v_r}{\partial r} + \frac{\partial v_z}{\partial z} \right), \quad (5)$$

$$\dot{T}_{zz} = (\lambda + 2\mu) \frac{\partial v_z}{\partial z} + \lambda \left(\frac{v_r}{r} + \frac{1}{r} \frac{\partial v_\phi}{\partial \phi} + \frac{\partial v_r}{\partial r} \right), \quad (6)$$

$$\dot{T}_{r\phi} = \mu \left(\frac{1}{r} \frac{\partial v_r}{\partial \phi} + \frac{\partial v_\phi}{\partial r} - \frac{v_\phi}{r} \right), \quad (7)$$

$$\dot{T}_{rz} = \mu \left(\frac{\partial v_z}{\partial r} + \frac{\partial v_r}{\partial z} \right), \quad (8)$$

$$\dot{T}_{z\phi} = \mu \left(\frac{\partial v_\phi}{\partial z} + \frac{1}{r} \frac{\partial v_z}{\partial \phi} \right). \quad (9)$$

The variables v and T are the components of the velocity vector and stress tensor, respectively. The material parameters are density ρ and the Lamé constants λ and μ . The body force components, f_r , f_z , and f_ϕ are manually driven to excite the simulation space. To describe elastic waves in pipe bends, we use a modified cylindrical coordinate system where the z axis is shifted and then curved to follow the center of a pipe bend as shown in Fig. 1. The only effect this ultimately has on Eqs. (1)–(9) is that dz is now replaced with the following:

$$dz = r \sin(\varphi - 0.5\pi + \varphi_{\text{curvature}}) + r_{\text{curvature}}. \quad (10)$$

Here $r_{\text{curvature}}$ is the radius of curvature of the pipe bend and $\varphi_{\text{curvature}}$ is the angle that points toward the inside of the pipe bend. For the finite integration method, the nine unknown variables are placed on a staggered grid. A single control volume with the variable distribution for the cylindrical coordinate system is shown in Fig. 2. Many millions of these grid cells are stacked together to create a computational space that accurately represents a pipe-like structure. Using the finite integration technique in this curved-cylindrical coordinate system, we arrive at the following nine equations. A detailed derivation of Eq. (11) for a straight pipe section is found in the Appendix:

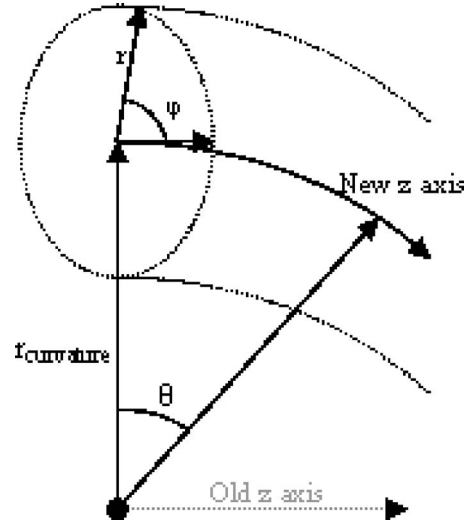


FIG. 1. Curved cylindrical coordinate system. The z axis is shifted and curved to follow the middle of a pipe bend.

$$\begin{aligned} p \dot{v}_r = & \frac{(r^{(o)} c^{(o)} T_{rr}^{(o)} - r^{(i)} c^{(i)} T_{rr}^{(i)})}{r^{(c)} c^{(c)} \Delta r} + \frac{(T_{r\phi}^{(+)} - T_{r\phi}^{(-)})}{r^{(c)} \Delta \phi} \\ & + \frac{(T_{rz}^{(i)} - T_{rz}^{(b)})}{c^{(c)} \Delta \theta} + \frac{T_{rr}^{(o)} + T_{rr}^{(i)} - T_{\phi\phi}^{(o)} - T_{\phi\phi}^{(i)}}{2r^{(c)}} + \rho f_r, \end{aligned} \quad (11)$$

$$\begin{aligned} p \dot{v}_z = & \frac{(r^{(o)} c^{(o)} T_{rz}^{(o)} - r^{(i)} c^{(i)} T_{rz}^{(i)})}{r^{(c)} c^{(c)} \Delta r} + \frac{(T_{\phi z}^{(+)} - T_{\phi z}^{(-)})}{r^{(c)} \Delta \phi} \\ & + \frac{(T_{zz}^{(i)} - T_{zz}^{(b)})}{c^{(c)} \Delta \theta} + \frac{(T_{rz}^{(o)} + T_{rz}^{(i)})}{2r^{(c)}} + \rho f_z, \end{aligned} \quad (12)$$

$$\begin{aligned} p \dot{v}_\phi = & \frac{(r^{(o)} c^{(o)} T_{r\phi}^{(o)} - r^{(i)} c^{(i)} T_{r\phi}^{(i)})}{r^{(c)} c^{(c)} \Delta r} + \frac{(T_{\phi\phi}^{(+)} - T_{\phi\phi}^{(-)})}{r^{(c)} \Delta \phi} \\ & + \frac{(T_{z\phi}^{(i)} - T_{z\phi}^{(b)})}{c^{(c)} \Delta \theta} + \frac{(T_{r\phi}^{(o)} + T_{r\phi}^{(i)})}{r^{(c)}} + \rho f_\phi, \end{aligned} \quad (13)$$

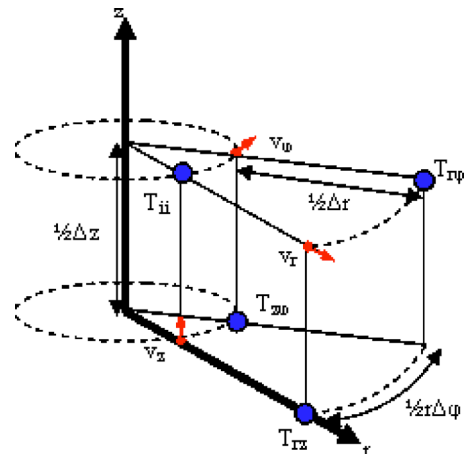


FIG. 2. (Color online) Staggered distribution of the nine unknown variables on a single 3D cylindrical computational cell.

$$\begin{aligned} \dot{T}_{rr} = & \frac{(\lambda + 2\mu)(r^{(o)}c^{(o)}v_r^{(o)} - r^{(i)}c^{(i)}v_r^{(i)})}{r^{(c)}c^{(c)}\Delta r} + \frac{\lambda(v_\varphi^{(+)} - v_\varphi^{(-)})}{r^{(c)}\Delta\varphi} \\ & + \frac{\lambda(v_z^{(i)} - v_z^{(b)})}{c^{(c)}\Delta\theta} + \frac{\lambda(v_r^{(o)} + v_r^{(i)})}{2r^{(c)}}, \end{aligned} \quad (14)$$

$$\begin{aligned} \dot{T}_{zz} = & \frac{(\lambda + 2\mu)(v_z^{(i)} - v_z^{(b)})}{c^{(c)}\Delta\theta} + \frac{\lambda(v_\varphi^{(+)} - v_\varphi^{(-)})}{r^{(c)}\Delta\varphi} \\ & + \frac{\lambda(r^{(o)}c^{(o)}v_r^{(o)} - r^{(i)}c^{(i)}v_r^{(i)})}{r^{(c)}c^{(c)}\Delta r} + \frac{\lambda(v_r^{(o)} + v_r^{(i)})}{2r^{(c)}}, \end{aligned} \quad (15)$$

$$\begin{aligned} \dot{T}_{\varphi\varphi} = & \frac{(\lambda + 2\mu)(v_\varphi^{(+)} - v_\varphi^{(-)})}{r^{(c)}\Delta\varphi} + \frac{\lambda(c^{(o)}r^{(o)}v_r^{(o)} - c^{(i)}r^{(i)}v_r^{(i)})}{r^{(c)}c^{(c)}\Delta r} \\ & + \frac{\lambda(v_z^{(i)} - v_z^{(b)})}{c^{(c)}\Delta\theta} + \frac{(\lambda + 2\mu)(v_r^{(o)} + v_r^{(i)})}{2r^{(c)}}, \end{aligned} \quad (16)$$

$$\dot{T}_{rz} = \frac{\mu(r^{(o)}c^{(o)}v_z^{(o)} - r^{(i)}c^{(i)}v_z^{(i)})}{r^{(c)}c^{(c)}\Delta r} + \frac{\mu(v_r^{(i)} - v_r^{(b)})}{c^{(c)}\Delta\theta}, \quad (17)$$

$$\begin{aligned} \dot{T}_{r\varphi} = & \frac{\mu(v_r^{(+)} - v_r^{(-)})}{r^{(c)}\Delta\varphi} + \frac{\mu(r^{(o)}c^{(o)}v_\varphi^{(o)} - r^{(i)}c^{(i)}v_\varphi^{(i)})}{r^{(c)}c^{(c)}\Delta r} \\ & - \frac{\mu(v_\varphi^{(o)} + v_\varphi^{(i)})}{2r^{(c)}}, \end{aligned} \quad (18)$$

$$\dot{T}_{z\varphi} = \frac{\mu(v_\varphi^{(i)} - v_\varphi^{(b)})}{c^{(c)}\Delta\theta} + \frac{\mu(v_z^{(+)} - v_z^{(-)})}{r^{(c)}\Delta\varphi}. \quad (19)$$

Here $r^{(i)}$, $r^{(c)}$, and $r^{(o)}$ are the radial distances measured from the curved z axis to the inside, center, and outside of the control volume. The variables $c^{(i)}$, $c^{(c)}$, and $c^{(o)}$ are the distances measured from the axis of curvature to the inside, center, and outside of the control volume. These values are a function of the current radial position, the radius of curvature $r_{\text{curvature}}$, and the angle pointing toward the inside of the bend $\varphi_{\text{curvature}}$:

$$c^{(i)} = [r^{(i)}\sin(\varphi - 0.5\pi + \varphi_{\text{curvature}}) + r_{\text{curvature}}], \quad (20)$$

$$c^{(c)} = [r^{(c)}\sin(\varphi - 0.5\pi + \varphi_{\text{curvature}}) + r_{\text{curvature}}], \quad (21)$$

$$c^{(o)} = [r^{(o)}\sin(\varphi - 0.5\pi + \varphi_{\text{curvature}}) + r_{\text{curvature}}]. \quad (22)$$

To simulate a straight pipe section, we set $c^{(i)}=c^{(c)}=c^{(o)}=1$ and $\Delta\theta=\Delta z$. With these equations, it is straightforward to simulate elastic waves in pipe segments containing combinations of straight sections and bends in any direction, including multiple out-of-plane bends. After having discretized the equations in space using the finite integration technique, we approximate the time derivatives using the standard central difference

$$v_i^{(n)} = v_i^{(n-1)} - \dot{v}_i^{(n-1/2)}\Delta t, \quad (23)$$

$$T_{ij}^{(n+1/2)} = T_{ij}^{(n-1/2)} - \dot{T}_{ij}^{(n)}\Delta t. \quad (24)$$

Here the superscript n represents the time step. This leads to a temporal discretization that is staggered in time. With these

equations, we can update each value in our simulation space based on the neighboring values.

B. Stability criteria

The spatial discretization in the radial direction is identical to the one derived by Schubert in his axially symmetric cylindrical finite integration technique.¹⁴ We assign 10 grid points to the shear wavelength:

$$\Delta r \approx \left(\frac{1}{10}\right)\frac{c_s}{f_{\text{max}}}. \quad (25)$$

Here c_s is the shear wave speed and f_{max} is the maximum frequency present in the simulation. The exact radial step size is adjusted as needed to simulate the correct pipe wall thickness. Next, we choose $\Delta\varphi$ so that the grid spacing on the outer circumference of the pipe ($r_{\text{outer}}\Delta\varphi$) is equal to or smaller than Δr and that the total number of grid points in the φ direction sweeps an angle of exactly 2π . First we find the number of grid points in the φ -direction N_φ :

$$N_\varphi = \left\lceil \frac{2\pi r_{\text{outer}}}{\Delta r} \right\rceil. \quad (26)$$

Now we find the exact discretization angle $\Delta\varphi$:

$$\Delta\varphi = \frac{2\pi}{N_\varphi}. \quad (27)$$

Again, we choose $\Delta\theta$ so that the grid spacing on the outside of a bend is equal to or less than Δr . First we find the number of grid points in the z direction through the curve that sweeps an angle of S :

$$N_\theta = \left\lceil \frac{S(r_{\text{outer}} + r_{\text{curve}})}{\Delta r} \right\rceil. \quad (28)$$

Now we find $\Delta\theta$ as

$$\Delta\theta = \frac{S}{N_\theta}. \quad (29)$$

If a straight section is being simulated, then we set $\Delta\theta = \Delta z \leq \Delta r$. The temporal discretization is found using the fastest wave speed and the smallest spatial grid sizes. In the φ direction, this is found on the inside circumference of the pipe. In the θ direction, this is found on the outer edge of the pipe on the inside of the pipe bend. We use the grid spacing at these locations to compute our time step Δt :

$$\Delta t \leq \frac{1}{c_l \sqrt{\frac{1}{\Delta r^2} + \frac{1}{[(r_{\text{curve}} - r_{\text{outer}})\Delta\theta]^2} + \frac{1}{(r_{\text{inner}}\Delta\varphi)^2}}}. \quad (30)$$

If a simulation models a pipe segment containing multiple bends of different curvatures, then the smallest curvature must be used to determine the temporal step size.

C. Boundary conditions

Here we derive the stress-free boundary conditions at the surfaces of the pipe. Even with added complexity of the

full 3D space and the curved z axis, we arrive at the same boundary conditions as Schubert¹⁴ in the axially symmetric cylindrical case. We begin by enforcing that the velocity components be placed on the surface of the pipe. At the inner and outer surface of the pipe, we want the stress components T_{rr} , T_{rz} , and $T_{r\varphi}$ to be zero. Since the T_{rz} and $T_{r\varphi}$ stress components are on the surface, we simply set them to zero. To enforce that T_{rr} is zero at the surface, we set $T_{rr}^{(o)} = -T_{rr}^{(i)}$ at the outer surface of the pipe and $T_{rr}^{(i)} = -T_{rr}^{(o)}$ at the inner surface of the pipe. We then extrapolate to find the $T_{\varphi\varphi}$ term that is outside the boundary of the pipe. This leads to the following algorithm for finding ν_r .

At the outer surface of the pipe ($r=r_{\max}$)

We set $T_{rr}^{(o)} = -T_{rr}^{(i)}$, $T_{\varphi\varphi}^{(o)} = 2T_{\varphi\varphi}^{(i)} - T_{\varphi\varphi}^{(ii)}$, $T_{rz} = T_{rz} = T_{r\varphi} = 0$:

$$p\dot{\nu}_r = -\frac{2T_{rr}^{(i)}}{\Delta r} - \frac{3T_{\varphi\varphi}^{(i)} - T_{\varphi\varphi}^{(ii)}}{2r^{(c)}} + f_r. \quad (31)$$

At the inner surface of the pipe ($r=r_{\min}$)

We set $T_{rr}^{(i)} = -T_{rr}^{(o)}$, $T_{\varphi\varphi}^{(i)} = 2T_{\varphi\varphi}^{(o)} - T_{\varphi\varphi}^{(oo)}$, $T_{rz} = T_{rz} = T_{r\varphi} = 0$:

$$p\dot{\nu}_r = \frac{2T_{rr}^{(o)}}{\Delta r} - \frac{3T_{\varphi\varphi}^{(o)} - T_{\varphi\varphi}^{(oo)}}{2r^{(c)}} + f_r. \quad (32)$$

The same procedure is carried for the z -velocity components ν_z equations on the pipe ends.

At the ends of the pipe ($z=z_{\min}$)

We set $T_{zz}^{(b)} = -T_{zz}^{(t)}$, $T_{zz} = T_{rz} = T_{z\varphi} = 0$:

$$p\dot{\nu}_z = -\frac{2T_{zz}^{(t)}}{c^{(c)}\Delta\theta} + f_z. \quad (33)$$

At the ends of the pipe ($z=z_{\max}$)

We set $T_{zz}^{(t)} = -T_{zz}^{(b)}$, $T_{zz} = T_{rz} = T_{z\varphi} = 0$:

$$p\dot{\nu}_z = \frac{2T_{zz}^{(b)}}{c^{(c)}\Delta\theta} + f_z. \quad (34)$$

D. Absorbing boundary layers

In most simulations, it is convenient to eliminate or significantly reduce wave reflections from the pipe ends, i.e., when the actual pipe is longer than the simulation space allows. This is accomplished by adding absorbing boundary regions to the end(s) of the simulated pipe. To do this, the velocities in the absorbing region are computed as usual but with a small damping factor. The damping factor is a function of the distance from the inside position in the absorbing layer a and the percentage p in which each layer removes from the velocity values. We replace (22) with the following equation:

$$\nu_i^{(n)} = \nu_i^{(n-1)} - D\nu_i^{(n-1/2)}\Delta t, \quad (35)$$

where the damping factor D is given by

$$D = (1 - p^* a). \quad (36)$$

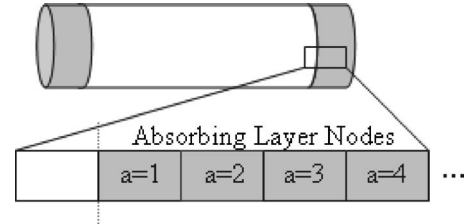


FIG. 3. Absorbing boundary layers are added to the pipe ends to significantly reduce reflections.

The damping factor is zero at the inside of the region and steadily increases as one moves to the outside of the absorbing region as shown in Fig. 3. This technique can be used to significantly reduce reflections on all acoustic and elastic finite integration methods. For the finite integration technique described in this paper, we found that an absorbing region 40–70 nodes thick with a damping percentage of 0.2% ($p=0.002$) worked well for reducing reflections from the artificial pipe ends.

II. PARALLEL IMPLEMENTATION

While some results can be obtained using the 3DCEFIT technique on a standard desktop computer, substantial improvements in computational time and model complexity are achieved with a parallel implementation. A parallel version of the 3DCEFIT has been implemented on William and Mary's high performance computational cluster, the SciClone.¹⁷ At the time of this work, the SciClone was composed of 311 computer processors with 236 GB of physical memory and 15.1 TB of disk capacity, and with a peak performance of 362 billion floating point operations per second.

The parallel algorithm uses a straightforward domain decomposition approach^{18,19} to divide the simulation space across many computers. After every half time step of the simulation, each computer swaps the appropriate boundary values with neighboring computers to create a large and seamless simulation space. Most, if not all, message passing interfaces allow blocking and nonblocking routines to send and receive data. We use a combination of these routines to achieve an optimized parallel algorithm. The parallel algorithm is given below.

Parallel algorithm

- (1) Compute boundary velocity values.
- (2) Send the new boundary values to appropriate neighboring computers (using a MPI non-blocking send).
- (3) Compute the rest of the velocity values.
- (4) Receive the boundary velocity values from neighbors (using a MPI blocking receive).
- (5) Repeat these four steps with the stress values.

Simulations that take many hours to complete on a single high-end desktop PC take just minutes using this parallel implementation.

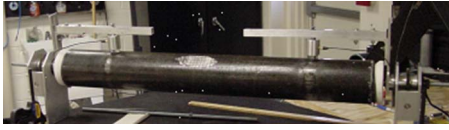


FIG. 4. (Color online) A photograph of the experimental apparatus. The steel pipe segment is 3 feet long, has an inner diameter of 4 inches, and a wall thickness of 0.25 inches.

III. COMPARISON WITH EXPERIMENTAL DATA

To validate this simulation method, we compared simulated and experimental ultrasonic guided waveforms. This was done by performing pitch-catch measurements and corresponding simulations on a straight pipe section mounted in a laboratory scanner²⁰⁻²³ as shown in Fig. 4. A-line waveforms were directly compared for three different catch transducer locations located at 0°, 90°, and 180° and at a separation distance of 60 cm from the pitch transducers. The longitudinal contact transducers are 3 cm in diameter and the pitch transducer is driven with a short 200 kHz toneburst. Figure 5 shows several snapshots in time from this simulation. The gray-scale intensity of these plots is proportional to the radial displacement on the outer surface of the pipe. Absorbing boundary conditions were not included in this simulation in order to closely match the experimental setup, i.e., a short segment of pipe. It can be seen from this figure how the presence of multiple modes, end reflections, and wrap-around waves can make signal interpretation quite difficult. Figure 6 shows normalized A-line waveforms from the simulation and experiment plotted together. The A-line waveforms match well at all three locations which gives us confidence that our simulation accurately describes guided wave propagation in pipe-like structures.

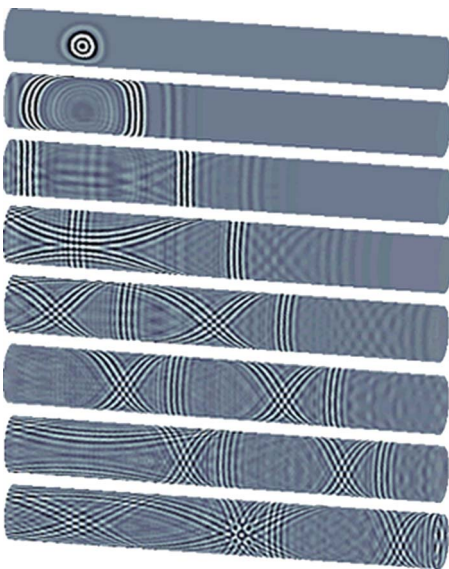


FIG. 5. Snapshots of a three-dimensional 200-kHz pipe simulation. A single 3-cm transducer is driven with a short 200-kHz tone burst. The gray scale color intensity is proportional to the radial displacement on the outer surface of the pipe.

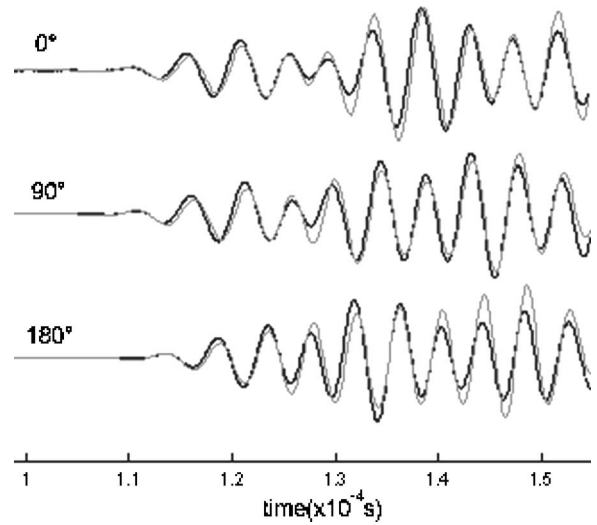


FIG. 6. Comparison between the simulated (gray) and the experimental (black) A-line data recorded at three different locations on the pipe showing very good agreement.

IV. EXAMPLE PIPE SIMULATIONS

A. Simulated flaw scattering

We next present simulation results from a pipe segment with a small thinning area located in the center of the pipe. The pipe dimensions are identical to the pipe simulated in the previous section. Figure 7 shows several snapshots from the simulation, while Fig. 8 shows two recorded A-lines taken from a clean and flawed pipe segment, respectively. The differences in the two A-lines are apparent at the beginning of the signals where one of the guided wave modes has shifted in time and changed amplitude. A systematic study of the interaction of guided waves with flaws and the resulting changes in the recorded A-line measurements will greatly benefit the development of automatic flaw detection algorithms.

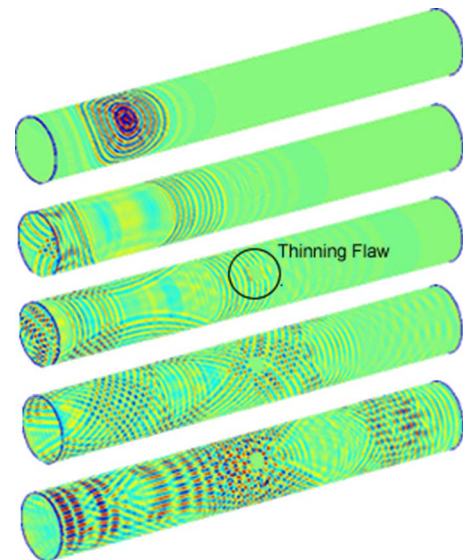


FIG. 7. (Color online) Snapshots of a three-dimensional pipe simulation with a small thinning flaw located in the center of the pipe. The shading is proportional to the radial displacement on the outer surface of the pipe.

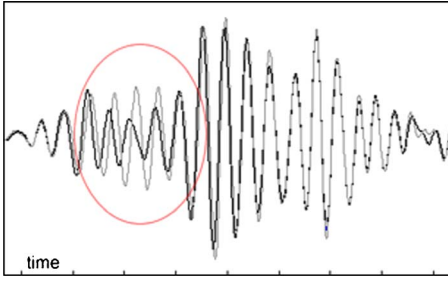


FIG. 8. (Color online) Comparison of A-line data from a clean (black) and flawed (gray) pipe segment. The circled region shows where one of the guided wave modes has shifted and changed in amplitude because of the interaction with the flaw.

Figure 8 demonstrates how small the differences are in the recorded waveforms between a corrosive-type flaw and unflawed pipe sample. Corrosive flaws are gradual thinnings that do not reflect strongly compared to other types of flaws such as saw cuts or flat-bottomed holes. The problem is compounded if the flaw is small and located in a hard-to-reach area such as underground or beyond pipe bends. For these reasons, it is advantageous to be able to focus guided wave energy at long distances and beyond pipe bends. The focal spot can be walked about the circumference of the pipe and down the axis to inspect the structure completely for flaws.

B. Guided wave focusing

In this section we present two focusing techniques to demonstrate that the 3DCEFIT simulation method can be used to develop and test hardware and software systems, and show agreement between simulated and experimental results.

1. Phased array transducer belts: Focusing with hardware

Guided wave focusing in pipes is typically done with phased array transducer belts.²⁴ The timing and amplitude of each excitation waveform are adjusted such that the desired guided-wave mode from each transducer arrives at the focal spot at the same time. The following algorithm will focus a single Lamb wave mode when using a phased array of normal-incidence contact transducers, but it is important to note that in focusing one mode of a multimode signal the other modes are not usually focused at the same desired location. The delay of each transducer d can be found given the velocity v of the desired guided wave mode and the shortest distance s between the given transducer and focal point. The delay d is given by the following equation:

$$d = \frac{s}{v}. \quad (37)$$

For a straight pipe section, the shortest distance between the transducer and the focal point is given by

$$s = \sqrt{f_z^2 + \gamma r^2}, \quad (38)$$

where f_z is the axial distance between the transducer and the focal point, r is the radius of the pipe, and γ is the smallest angle between the transducer and focal point. This angle γ is given by the following equation:

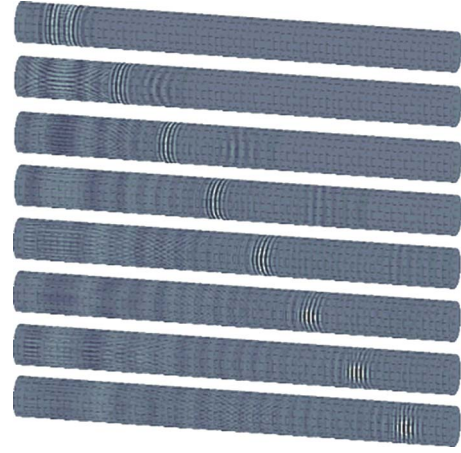


FIG. 9. Snapshots of a three-dimensional 200-kHz pipe simulation. A transducer belt containing 32 omnidirectional transducers. The transducers fire out of phase such that the desired guided wave becomes focused one meter from the transducer belt.

$$\gamma = \min \begin{cases} (a_\varphi - f_\varphi + 2\pi)^2, \\ (a_\varphi - f_\varphi)^2, \\ (a_\varphi - f_\varphi - 2\pi)^2. \end{cases} \quad (39)$$

Here, a_φ and f_φ are the angles in radians of the transducer and the focal point, respectively. The following simulation demonstrates phased array focusing on a straight pipe section. A phased array of thirty-two 1-cm-diameter normal-incidence contact transducers evenly spaced along the circumference of a 4-in.-inner-diameter steel pipe with a wall thickness of 0.25 in. is simulated. All the transducers are driven with the same 200 kHz tone-burst excitation except they are delayed according to Eq. (37) such that the desired guided wave mode arrives at the focal point, 1 m from the transducer belt, at the same time. Figure 9 shows several snapshots from this simulation. Figure 10 compares the energy distribution on the circumference of the pipe at one meter for the focused simulation (black) and an axially symmetric wave (gray). The energy of the axially symmetric wave is evenly distributed across the circumference while in the focused simulation, the energy is concentrated at 180°.

2. The synthetic aperture focusing technique (SAFT): Focusing in software

The synthetic aperture focusing technique (SAFT) is a numerical method for focusing ultrasound waves. SAFT was originally developed for radar applications and has been adapted by the NDE community for improving lateral resolution and imaging quality.²⁵ The SAFT technique has been used successfully with Lamb waves for locating and identifying flaws in plate-like structures.²⁶ Here we implement a time-domain SAFT technique for focusing Lamb waves in pipe structures.

One of the advantages of this technique is that no complicated phased array hardware is required. Instead, each transducer in the array fires individually while A-lines are recorded at each of the catch transducer locations. These A-lines are stored and then later combined using the SAFT

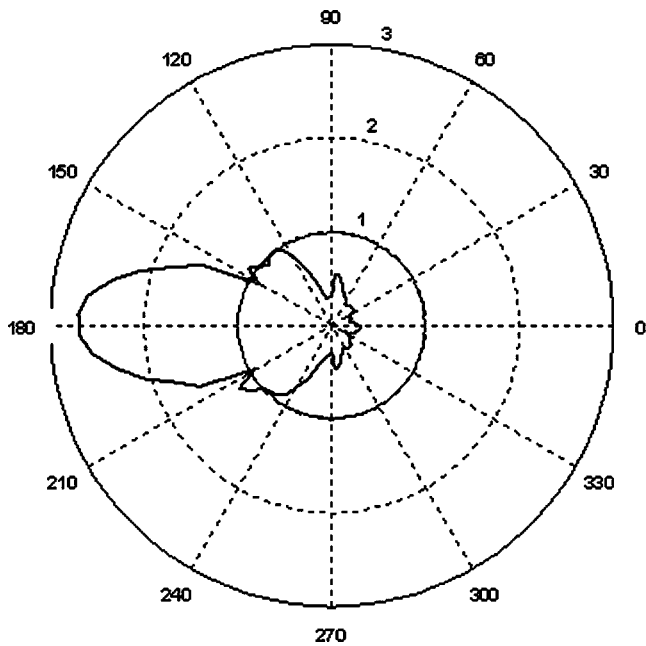


FIG. 10. Polar plot comparing the energy distribution on the circumference of the pipe at one meter for an axial symmetric wave (gray circle) and the focused wave (black). For an axial symmetric wave, the energy is uniformly distributed around the pipe. Using a phased array transducer belt and the focusing algorithm, the energy becomes concentrated at 180°.

algorithm to synthetically focus the ultrasound waves onto any location on the pipe. The A-line waveforms are shifted in time and summed up such that the guided waves arrived at the focal point at the same time. The time shift of each waveform is identical to the time shift we previously used for phased array focusing. The following equation produces a new A-line A_c at a given transducer location c by combining all the A-lines recorded at that location:

$$SA_c(t) = \sum_{n=1}^N A_{n,c}(t + d). \quad (40)$$

Here N is the number of pitch transducers, $SA_c(t)$ is the new SAFT constructed A-line at catch transducer location c , $A_{n,c}(t)$ is the recorded A-line from catch transducer c when the pitch transducer n fired, and d is the same delay we computed in Eq. (37). It should be noted that there are frequency domain versions of the SAFT algorithm that are more computationally efficient.²⁴ A simulation was performed to validate the SAFT algorithm, using a one meter pipe with 32 pitch and 32 catch transducers separated by 60 centimeters. Each pitch transducer fires individually while all 32 catch transducers record the radial pipe displacements over time (A-lines). This creates 32 A-lines recorded at each catch transducer location with 1024 A-lines in total. Figure 11 shows SAFT A-lines where the focal point was chosen to be at the location of one of the catch transducers. The focused guided wave mode is clearly visible in the new A-line data.

The SAFT technique also works well with experimental data. The same setup as above was performed experimentally. The same pitch and catch transducer locations were recorded experimentally so that the results could be com-

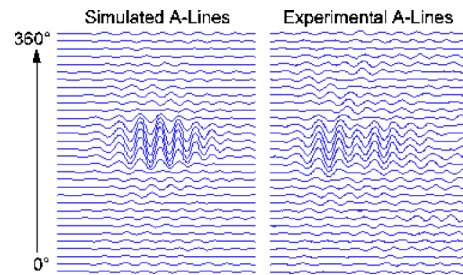


FIG. 11. A-lines produced by SAFT with the focal point at the catch transducer located at 180°. Results are shown for simulated data (left) and experimental data (right). In both cases, the S_0 mode is focused at 180°.

pared to the simulation results. Figure 11 shows the SAFT results with the experimental A-lines. Figure 12 shows a polar plot comparing the energy distribution on the pipe for the simulated and experimental data. In these examples, the focal point was chosen to be located along the catch transducer ring. In practice, the focal point would be swept across the entire surface of the pipe or sections of interest to image the pipe for flaws.

There is good agreement between the experimental and simulation results, giving an example of how simulations of this type can be used to develop signal processing techniques that may be difficult to refine using experimental data alone. With the SAFT algorithm, the focal point can be swept along the pipe in software to inspect it for flaws, which can substantially reduce the cost and complexity of the experimental apparatus.

V. PIPE BEND SIMULATION EXAMPLES

Most piping systems contain bends which make inspecting them problematic. The 3DCEFIT technique as described above can simulate elastic guided waves propagation in pip-

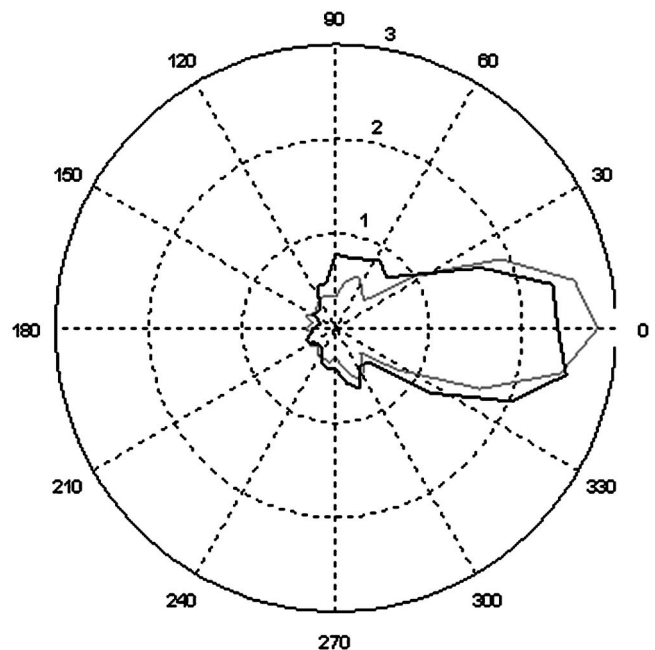


FIG. 12. Polar plot showing the energy distribution on the circumference of the pipe at the focal point using the SAFT algorithm. There is very good agreement between the experimental (black) and simulated (gray) results.



FIG. 13. (Color online) A 50-kHz guided elastic wave propagates through a 90° pipe bend. The guided waves naturally focus at the back of the bend.

ing systems that contain combinations of bends of various curvature and in any direction. Figure 13 shows snapshots from an initially axial-symmetric 50-kHz guided wave propagating through a 90° pipe bend with radius of curvature of 6 in. The axial symmetric wave was created with a transducer belt consisting of 32 omnidirectional 1-cm-diameter transducers. As expected,¹⁰ the guided waves naturally focus at the back of the pipe bend.

To validate the simulation method for pipes with bends, we next compare our 3DCEFIT simulation results directly to results obtained with the commercial finite element software package COMSOL. The pipe is symmetrically excited with 5 cycle 100 kHz shear transducers on the outer radius of the pipe. The pipe segment contains a 90° bend, of the same dimensions as in the previous example, located 1.5 m from

the transducer belt. The shear displacement field on the outer surface of the pipe, after propagating through the bend, was recorded over time. The shear displacements obtained using COMSOL and 3DCEFIT at two time instances are shown in Fig. 14 and agree well. For comparison, this simulation took roughly 18 h to run using COMSOL on a single desktop computer (2 GHz processor). The 3DCEFIT simulation took 7 min on a 64 node parallel computer (650 MHz processor per node).

Figure 15 shows a 100-kHz guided wave propagating through an S-bend and Fig. 16 shows a 100-kHz guided wave propagating through a series of 3D pipe bends. These figures demonstrate the ability to simulate complex piping systems containing multiple bends. This simulation method can also be used to simulate guided waves in pipe coils that are routinely found in heating and cooling systems and power plants.

VI. CONCLUSIONS

We have developed and implemented a 3D simulation method based on the finite integration technique for modeling guided elastic waves in pipe-like structures including pipe bends. Comparisons show agreement between simulated and experimental data, and we have shown that the finite integration technique is well suited for modeling elastic wave propagation and interactions with flaws. This simulation method can be used as a design tool for developing complicated inspection hardware and signal processing algorithms. Future work will focus on systematically studying guided wave interaction with varying flaw types and sizes. Discretizing the material parameters will also allow one to model piping systems which contain coatings and welds. Overall, the 3DCEFIT technique provides an accurate method for simulating guided elastic waves in complex piping systems.

ACKNOWLEDGMENTS

Support for this work was provided in part by the US Navy under Contract No. N68335-05-C-0189.

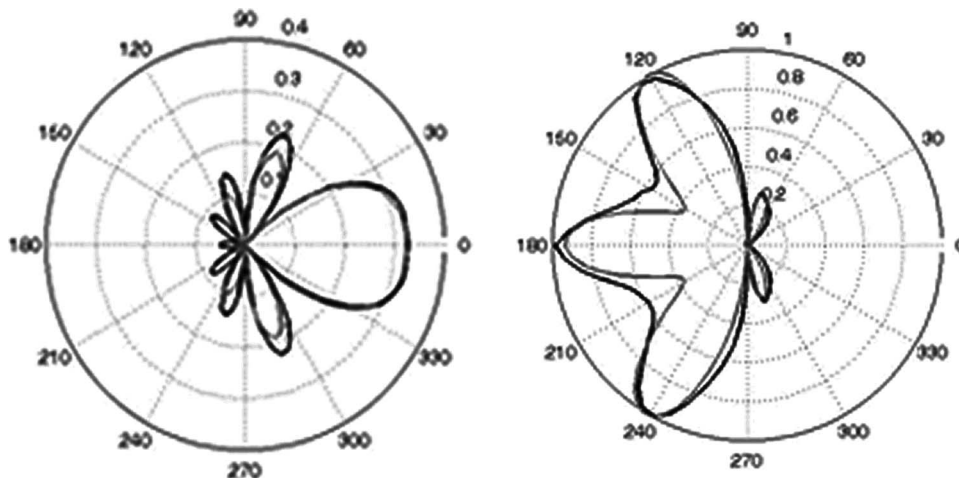


FIG. 14. Polar plots of the shear displacements obtained from a COMSOL finite element simulation (black) and a 3DCEFIT simulation (gray). The shear displacements were recorded at 361 μ s (left) and 390 μ s (right) from the initial transducer excitation.

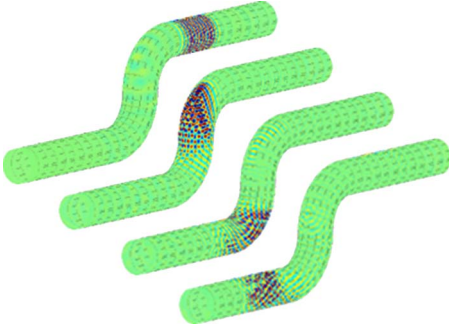


FIG. 15. (Color online) A 100-kHz guided elastic wave propagates through a pipe S-bend.

APPENDIX: SAMPLE 3DCEFIT DERIVATION

Here we describe in detail how to discretize (1), which will illustrate how to find the radial velocity v_r component at each time step for a straight pipe section. To model a pipe bend, we simply make the substitution in (11). We begin by integrating both sides of Eq. (1) over a cylindrical control volume centered on the unknown variable v_r :

$$\iiint \rho \dot{v}_r r dr dz d\varphi = \iiint \left(\frac{\partial T_{rr}}{\partial r} + \frac{1}{r} \frac{\partial T_{r\varphi}}{\partial \varphi} + \frac{\partial T_{rz}}{\partial z} + \frac{T_{rr} - T_{\varphi\varphi}}{r} + \rho f_r \right) r dr dz d\varphi.$$

Now we invoke the divergence theorem which replaces part of the volume integrals on the right-hand side with a surface integral:

$$\begin{aligned} \iiint \rho \dot{v}_r r dr dz d\varphi = & \oint_{\partial v} (r T_{rr} dz d\varphi + T_{r\varphi} dr dz \\ & + r T_{rz} dr d\varphi) + \iiint \left(\frac{T_{rr} - T_{\varphi\varphi}}{r} \right. \\ & \left. + \rho f_r \right) r dr dz d\varphi. \end{aligned}$$

Next we approximate the integrals by multiplying the integrand by the volume or surface of each integral. Unlike the Cartesian method, the inner and outer surfaces of the control volume have different surface areas. These surface areas are defined by $r\Delta z\Delta\varphi$, where the distance r is different for the inner and outer surfaces of the control volume. Thus we denote $r^{(i)}$ and $r^{(o)}$ as the radial distance to the inner and outer surfaces, respectively. We will also introduce superscripts $+$, $-$, t , and b to indicate the direction of the variables relative to the unknown variable v_r . The superscripts $+$ and $-$ represent the direction of the variables in the positive and

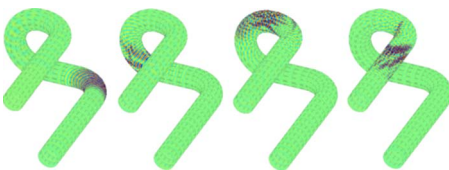


FIG. 16. (Color online) A 100-kHz guided elastic wave propagates through many 3D pipe bends. The simulations provide a way to predict the path of the guided waves in complicated piping systems.

negative φ directions and superscripts t and b represent the direction of the variables in the positive and negative z directions:

$$\begin{aligned} \rho \dot{v}_r r \Delta r \Delta z \Delta \varphi = & (r^{(o)} T_{rr}^{(o)} - r^{(i)} T_{rr}^{(i)}) \Delta z \Delta \varphi + (T_{r\varphi}^{(+)} \\ & - T_{r\varphi}^{(-)}) \Delta r \Delta z + (T_{rz}^{(t)} - T_{rz}^{(b)}) r \Delta r \Delta \varphi \\ & + \left(\frac{T_{rr}^{(o)} + T_{rr}^{(i)} - T_{\varphi\varphi}^{(o)} - T_{\varphi\varphi}^{(i)}}{2r} \right. \\ & \left. + \rho f_r \right) r \Delta r \Delta z \Delta \varphi. \end{aligned}$$

Next, we divide both sides by $r\Delta r\Delta z\Delta\varphi$:

$$\begin{aligned} \rho \dot{v}_r = & \frac{(r^{(o)} T_{rr}^{(o)} - r^{(i)} T_{rr}^{(i)})}{r \Delta r} + \frac{(T_{r\varphi}^{(+)} - T_{r\varphi}^{(-)})}{r \Delta \varphi} + \frac{(T_{rz}^{(t)} - T_{rz}^{(b)})}{\Delta z} \\ & + \frac{T_{rr}^{(o)} + T_{rr}^{(i)} - T_{\varphi\varphi}^{(o)} - T_{\varphi\varphi}^{(i)}}{2r} + \rho f_r. \end{aligned}$$

We have now spatially discretized our equation. Next we approximate the time derivative on the left-hand side using a standard central difference. This yields our final equation:

$$\begin{aligned} v_r = v_r + \frac{\Delta t}{\rho} \left[\frac{(r^{(o)} T_{rr}^{(o)} - r^{(i)} T_{rr}^{(i)})}{r \Delta r} + \frac{(T_{r\varphi}^{(+)} - T_{r\varphi}^{(-)})}{r \Delta \varphi} \right. \\ \left. + \frac{(T_{rz}^{(t)} - T_{rz}^{(b)})}{\Delta z} + \frac{T_{rr}^{(o)} + T_{rr}^{(i)} - T_{\varphi\varphi}^{(o)} - T_{\varphi\varphi}^{(i)}}{2r} + \rho f_r \right]. \end{aligned}$$

This equation tells us how to update each of the v_r values on the grid given its neighboring stress values at past half-time step.

- ¹J. L. Rose, "A baseline and vision of ultrasonic guided wave inspection potential," *J. Pressure Vessel Technol.* **124**, 273–282 (2002).
- ²J. L. Rose, "Standing on the shoulders of giants: An example of guided wave inspection," *Mater. Eval.* **60**, 53–59 (2002).
- ³J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, 1999).
- ⁴A. H. Harker, *Elastic Waves in Solids—With Applications to Nondestructive Testing of Pipelines* (Institute of Physics Publishing, Bristol, UK, British Gas 1988).
- ⁵J. L. Rose, D. Jiao, and J. Spanner, Jr., "Ultrasonic guided wave NDE for piping," *Mater. Eval.* **54**, 1310–1313 (1996).
- ⁶M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "Defect detection in pipes using guided waves," *Ultrasonics* **36**, 147–154 (1998).
- ⁷D. N. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," *Mater. Eval.* **55**, 504–508 (1997).
- ⁸D. Alleyne *et al.*, "The Lamb wave inspection of chemical plant pipework," in *Review of Progress in QNDE* (Plenum, New York, 1997), pp. 1269–1276.
- ⁹D. Gsell, T. Leutenegger, and J. Dual, "Modeling three-dimensional elastic wave propagation in circular cylindrical structures using a finite-difference approach," *J. Acoust. Soc. Am.* **116**, 3284–3293 (2004).
- ¹⁰T. Leutenegger and J. Dual, "Detection of defects in cylindrical structures using a time reverse method and a finite-difference approach," *Ultrasonics* **40**, 721–725 (2002).
- ¹¹T. Hayashi, K. Kawashima, Z. Q. Sun, and J. L. Rose, "Analysis of flexural mode focusing by a semianalytical finite element method," *J. Acoust. Soc. Am.* **113**, 1241–1248 (2003).
- ¹²T. Hayashi, K. Kawashima, Z. Q. Sun, and J. L. Rose, "Guided wave propagation mechanics across a pipe elbow," *J. Pressure Vessel Technol.* **127**, 322–327 (2005).
- ¹³P. Fellingner, R. Marklein, K. J. Langenberg, and S. Klaholz, "Numerical modeling of elastic wave propagation and scattering with EFIT—elastodynamic finite integration technique," *Wave Motion* **21**, 47–66 (1995).
- ¹⁴F. Schubert, A. Peiffer, and B. Kohler, "The elastodynamic finite integration technique for waves of cylindrical geometries," *J. Acoust. Soc. Am.* **104**, 2604–2614 (1998).
- ¹⁵F. Schubert, "Numerical time-domain modeling of linear and nonlinear

ultrasonic wave propagation using finite integration techniques—theory and applications,” *Ultrasonics* **42**, 221–229 (2004).

¹⁶K. F. Graff, *Wave Motion In Elastic Solids* (Dover, New York, 1991).

¹⁷The SciClone Cluster Project <http://www.compsci.wm.edu/SciClone/>.

¹⁸R. Marklein, “Numerical simulation of fields and waves in nondestructive testing” *9th European Conference on Non-Destructive Testing* Berlin, Germany (2006).

¹⁹T. Bohlen, “Parallel 3-D viscoelastic finite difference seismic modeling,” *Comput. Geosci.* **28**, 887–899 (2002).

²⁰K. R. Leonard and M. K. Hinders, “Guided wave helical ultrasonic tomography of pipes,” *J. Acoust. Soc. Am.* **114**, 767–774 (2003).

²¹K. R. Leonard and M. K. Hinders, “Lamb wave tomography of pipe-like structures,” *Ultrasonics* **44**, 574–583 (2005).

²²M. K. Hinders and K. R. Leonard, “Lamb wave tomography of pipes and tanks using frequency compounding,” *Rev. Prog. Quant. Nondestr. Eval.*, edited by D.O. Thompson and D.E. Chimenti, **24**, 867–874 (2005).

²³K. R. Leonard and M. K. Hinders, “Lamb wave helical ultrasonic tomography,” *Rev. Prog. Quant. Nondestr. Eval.*, edited by D.O. Thompson and D.E. Chimenti, **23**, 173–180 (2004).

²⁴T. Hayashi, K. Kawashima, Z. Sun, and J. Rose, “Guided wave focusing mechanics in pipe,” *J. Pressure Vessel Technol.* **127**, 317–321 (2005).

²⁵V. Schmitz, S. Chakhlov, and W. Müller, “Experiences with synthetic aperture focusing technique in the field,” *Ultrasonics* **38**, 731–738 (2000).

²⁶R. Sicard, J. Goyette, and D. Zellouf, “A SAFT algorithm for Lamb wave imaging of isotropic plate-like structures,” *Ultrasonics* **39**, 487–494 (2002).

A moving zone of quiet for narrowband noise in a one-dimensional duct using virtual sensing

Cornelis D. Petersen,^{a)} Anthony C. Zander,^{b)} Ben S. Cazzolato,^{c)} and Colin H. Hansen^{d)}
*Active Noise and Vibration Control Group, School of Mechanical Engineering,
The University of Adelaide, SA 5005, Australia*

(Received 15 September 2005; revised 4 December 2006; accepted 13 December 2006)

A frequent problem in active noise control is that the zone of quiet created at the error sensor tends to be very small. This means that the error sensor generally needs to be located close to an observer's ear, which might not always be a convenient or feasible solution. Virtual sensing is a method that can move the zone of quiet away from the error sensor to a desired location that is spatially fixed. This method has been investigated previously, and has shown potential to improve the performance of an active noise control system. However, it is very likely that the desired location of the zone of quiet is not spatially fixed. An active noise control system incorporating a virtual sensing method thus has to be able to create a moving zone of quiet that tracks the observer's ears. This paper presents a method for creating a moving zone of quiet based on the LMS virtual microphone technique. To illustrate the proposed method, it is implemented in an acoustic duct and narrowband control results are presented. These results show that a moving zone of quiet was effectively created inside the duct for narrowband noise. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431583]

PACS number(s): 43.50.Ki [KAC]

Pages: 1459–1470

I. INTRODUCTION

The aim of a local active noise control system is to create a zone of quiet at a specific location, for instance at the passenger's ear inside a vehicle cabin. Generally, the greatest noise reduction is achieved at the error sensor location, and the created zone of quiet tends to be very small. Elliott *et al.*¹ have shown both analytically and experimentally that in a diffuse sound field, the zone of quiet in which the noise is reduced by 10 dB or more typically has the shape of a sphere with a diameter of one-tenth of an acoustic wavelength. This means that the error sensor usually has to be placed close to an observer's ear, which might not always be a convenient or feasible solution. Virtual sensing^{2–5} is a method that has been developed in order to overcome these problems that are often encountered in local active noise control systems. This method requires a nonintrusive sensor which is placed remotely from the desired location of maximum noise attenuation. The nonintrusive sensor is used to provide an estimate of the pressure at the desired location, which is spatially fixed. The estimated pressure can then be minimized by a local active noise control system such that the zone of quiet is moved away from the physical location of the transducers to the spatially fixed desired location of maximum attenuation, such as a person's ear.

The concept of virtual sensing has been shown to improve the performance of a local active noise control system.^{2–20} However, it is very likely that the desired location of maximum attenuation is not spatially fixed. For in-

stance, a passenger inside a vehicle cabin will move their head, thereby changing the desired location of the zone of quiet. This means that an active noise control system incorporating a virtual sensing method has to be able to create a moving zone of quiet that tracks the passenger's head. Although active noise control at a spatially fixed virtual location has been investigated previously by a number of authors, the concept of creating a zone of quiet at a moving virtual location based on virtual sensing has not been investigated to the authors' knowledge.

Elliott and David² were the first to suggest a virtual sensing method called the *virtual microphone arrangement*. An important assumption made in this method is that the primary pressures at the physical and virtual microphones are equal. Furthermore, a preliminary system identification step is required in which models of the transfer paths between the control source and the physical and virtual microphones are estimated. The assumption of equal primary pressures and knowledge of these transfer paths allow the pressure at the virtual microphone to be estimated. The virtual microphone arrangement has been thoroughly investigated by a number of authors.^{3,10}

A virtual sensing method called the *remote microphone technique* was suggested by both Popovich¹¹ and Roure and Albarrazin.⁴ The remote microphone technique requires the estimation of two transfer paths and one filter, which are usually estimated in a preliminary system identification step. The two transfer paths are the secondary transfer paths also needed in the virtual microphone arrangement.² However, an additional filter needs to be estimated in the remote microphone technique, that computes an estimate of the primary pressure at the virtual microphone from the primary pressure at the physical microphone. The virtual microphone arrangement assumes this filter to be unity, and is therefore a sim-

^{a)}Electronic mail: cornelis.petersen@mecheng.adelaide.edu.au

^{b)}Electronic mail: anthony.zander@mecheng.adelaide.edu.au

^{c)}Electronic mail: ben.cazzolato@mecheng.adelaide.edu.au

^{d)}Electronic mail: colin.hansen@mecheng.adelaide.edu.au

plified version of the remote microphone technique.

Cazzolato⁵ suggested a virtual sensing method based on *forward difference prediction techniques*. In this approach, the pressure at the virtual microphone is estimated by summing the weighted pressures from a number of microphones in an array. The weights for each of the elements in the microphone array are determined using forward difference prediction techniques. Kestell¹² and Munn¹³ investigated the forward difference prediction approach for simple sound fields, namely an acoustic duct and a free field. Experiments showed that a linear prediction method, which uses a two-microphone array, proved to be better in practice than a quadratic prediction method, which uses a three-microphone array.^{14–16} The reason for this was the high sensitivity of the quadratic prediction method to short wavelength extraneous noise. In an effort to overcome this problem, higher-order virtual microphone arrays, which were thought to be able to spatially filter out the extraneous noise, were investigated.¹³ A higher number of microphones than the order of the prediction algorithm is used in this method, which results in an overconstrained problem that is solved by a least-squares approximation. Unfortunately, the accuracy of these higher-order prediction algorithms was shown to be very much affected by the phase and sensitivity mismatches and relative position errors between the microphones in the array.¹⁷ These mismatches and position errors are generally unavoidable, especially if the number of microphones used is large.

Instead of using forward difference prediction techniques to determine the weights for each of the elements in the microphone array, Cazzolato¹⁸ suggested that the optimal values for the weights could be determined using the LMS algorithm.²¹ A similar method was also suggested by Gawron and Schaff,¹⁹ who applied it to local active noise control inside a car cabin. The *adaptive LMS virtual microphone technique* involves placing a microphone at the virtual location, after which the microphone weights are adapted by the LMS algorithm so as to optimally predict the sound pressure at this location. After the weights have converged, the microphone is removed from the virtual location and the weights are fixed to their optimal value. The adaptive LMS virtual microphone technique was theoretically investigated for an acoustic duct.¹³ The results indicated that the adaptive LMS algorithm was able to completely compensate for relative position errors and sensitivity mismatches between the microphone elements, and partly compensate for phase mismatches between the microphone elements. The adaptive LMS virtual microphone technique proved to outperform the forward difference prediction technique in real-time control experiments conducted in an acoustic duct.^{13,20}

The virtual sensing methods discussed so far were all developed with the aim to create a zone of quiet at a spatially fixed virtual microphone. The primary aim of this paper is to present a method for creating a zone of quiet at a moving virtual microphone inside an acoustic duct based on the adaptive LMS virtual microphone technique.¹⁸ In previous work,^{13,18} this technique has only been applied to the case of using *one spatially fixed* virtual microphone, and it is therefore initially extended here to the case of using *multiple spatially fixed* virtual microphones. The analysis is based on a

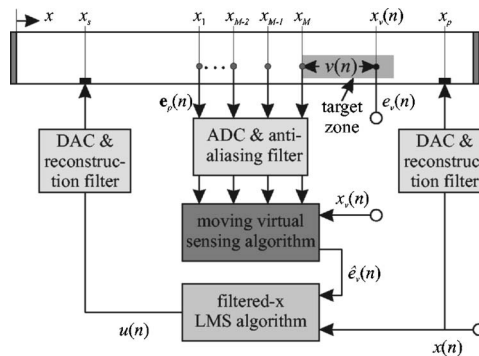


FIG. 1. Schematic diagram of the one-dimensional acoustic duct arrangement under consideration.

state-space model of the active noise control system under consideration, because this allows an easy extension to the case of a *moving* virtual microphone instead of a spatially fixed virtual microphone. The developed moving virtual microphone method is then combined with the filtered-x LMS algorithm.²² To illustrate the proposed method, it is implemented in a one-dimensional acoustic duct with the aim to create a moving zone of quiet for narrowband noise. Experimental results are presented and compared to experimental results for narrowband active noise control at a spatially fixed virtual microphone, and a spatially fixed physical microphone.

In a practical application, an important issue that would need to be addressed is how to determine the desired location of maximum attenuation, i.e., the moving virtual location. This could for instance be done using a three-dimensional head tracking system based on camera vision, or a head tracking system based on ultrasonic position sensing such as the Logitech[®] head tracker. However, the issue of determining the desired location of the zone of quiet is beyond the scope of this paper, and the focus here is on the development of a method for creating a zone of quiet at a moving virtual location inside an acoustic duct, while assuming that an exact determination of this location is available.

In Sec. II, the acoustic duct arrangement used to illustrate the proposed method is introduced. In Sec. III, the LMS virtual microphone technique is extended to include *multiple spatially fixed* virtual microphones, given a state-space model of the considered active noise control system. In Sec. IV, the theory discussed in Sec. III is extended to develop a *moving* virtual sensing algorithm that computes an estimate of the error signal at a moving virtual microphone. In Sec. V, it is explained how the moving virtual sensing algorithm is combined with the filtered-x LMS algorithm to create a zone of quiet at a moving virtual microphone. In Sec. VI, the proposed algorithm is implemented in real-time on an acoustic duct arrangement, and narrowband control results are presented.

II. PROBLEM DEFINITION

A. Acoustic duct arrangement

Figure 1 shows a schematic diagram of the one-dimensional acoustic duct arrangement considered here. The

rigid-walled acoustic duct has a length L , a primary source located at x_p , and a control source located at x_s . The primary source is excited by the disturbance source signal $x(n)$, and the control source by the control signal $u(n)$. Both signals are first passed through a DA-converter and a reconstruction filter before being sent to their respective sources. It is assumed here that the disturbance source signal $x(n)$ is a stationary process, and that the primary sound field created inside the acoustic duct is therefore stationary as well.

The aim of the active noise control system is to minimize the virtual error signal $e_v(n)$ at a moving virtual microphone located inside the acoustic duct, indicated by the black dot in Fig. 1. This moving virtual microphone tracks the desired location of the zone of quiet defined by the moving virtual location $x_v(n)=x_M+v(n)$, with $v(n)$ the *moving virtual distance* shown in Fig. 1. It will be assumed that the moving virtual microphone stays within a confined region inside the acoustic duct called the *target zone*, which has been indicated in Fig. 1 by the gray box. The virtual error signal at the moving virtual microphone is given by

$$e_v(n) = d_v(n) + y_v(n), \quad (1)$$

where $d_v(n)$ is the virtual primary disturbance that needs to be attenuated, and $y_v(n)$ is the virtual secondary disturbance created by the control source, such that

$$y_v(n) = G_{vu}(n)u(n), \quad (2)$$

with $G_{vu}(n)$ the *linear time-varying* virtual secondary transfer path. This transfer path is time-varying due to the movement of the virtual microphone.

The virtual primary disturbance $d_v(n)$ is a *nonstationary* process due to the movement of the virtual microphone, even though it is assumed that the primary sound field inside the acoustic duct is stationary. When the primary disturbance is nonstationary, the common approach in active noise control is to use an adaptive control algorithm. Adaptive control algorithms are able to track the changes in the statistical properties of the primary disturbance and adjust the controller accordingly.²² Here, the filtered-x LMS algorithm²² is used to adapt a FIR control filter that generates the control signal $u(n)$ in Fig. 1. A feedforward control approach is adopted as it is assumed that a feedforward reference signal $x(n)$ is available for control purposes. As illustrated in Fig. 1, this signal is equal to the disturbance source signal that excites the primary source.

B. Implementing the filtered-x LMS algorithm

If the virtual error signal $e_v(n)$ was directly measured by a physical microphone, and the virtual location was spatially fixed, the filtered-x LMS algorithm could be implemented as²²

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \mathbf{r}_v(n)e_v(n), \quad (3)$$

where μ is the convergence coefficient, $\mathbf{w}(n) \in \mathbb{R}^I$ is a vector containing the filter coefficients

$$\mathbf{w}(n) = [w_0(n) w_1(n) \cdots w_{I-1}(n)]^T, \quad (4)$$

and $e_v(n)$ is the virtual error signal directly measured by the spatially fixed physical microphone. The vector $\mathbf{r}_v(n) \in \mathbb{R}^I$ in Eq. (3) is given by

$$\mathbf{r}_v(n) = [r_v(n) r_v(n-1) \cdots r_v(n-I+1)]^T, \quad (5)$$

where $r_v(n)=G_{vu}x(n)$ is the virtual filtered-reference signal generated by filtering the reference signal $x(n)$ with the *linear time-invariant* virtual secondary transfer path G_{vu} . However, the virtual error signal $e_v(n)$ is not directly measured in the active noise control system considered here, and the virtual microphone is not spatially fixed but moving, such that the virtual secondary transfer path is linear time-varying, as indicated in Eq. (2). To implement the filtered-x LMS algorithm, these two issues thus need to be addressed.

1. Moving virtual sensing algorithm

As indicated by Eq. (3), the filtered-x LMS algorithm needs the feedback information contained in the virtual error signal $e_v(n)$ to update the control filter coefficients $\mathbf{w}(n)$. This signal is not directly measured by a physical microphone in the problem considered here. A *moving virtual sensing* method thus needs to be developed that is able to compute an estimate $\hat{e}_v(n)$ of the virtual error signal at the moving virtual location $x_v(n)$. As illustrated in Fig. 1, the moving virtual sensing algorithm uses information about the sound field provided by a physical microphone array, and knowledge of the moving virtual location $x_v(n)$ to compute an estimate $\hat{e}_v(n)$ of the virtual error signal. The physical microphone array, indicated by the gray dots in Fig. 1, measures the physical error signals

$$\mathbf{e}_p(n) = [e_1(n) e_2(n) \cdots e_M(n)]^T, \quad (6)$$

at the spatially fixed physical locations

$$\mathbf{x} = [x_1 x_2 \cdots x_M]^T. \quad (7)$$

The physical error signals $\mathbf{e}_p(n)$ are first passed through an antialiasing filter and an AD-converter before being used by the moving virtual sensing algorithm to compute an estimate $\hat{e}_v(n)$ of the virtual error signal. As illustrated in Fig. 1, instead of the unavailable virtual error signal $e_v(n)$, the filtered-x LMS algorithm uses this estimate to update the control filter coefficients.

2. Generating the virtual filtered-reference signal

The second issue that needs to be addressed to implement the filtered-x LMS algorithm in Eq. (3) is that the virtual microphone is moving. For a spatially fixed virtual microphone, the virtual secondary transfer path G_{vu} is *linear time-invariant*, and the virtual filtered-reference signal can be generated as usual, such that $r_v(n)=G_{vu}x(n)$. However, for a moving virtual microphone, the virtual secondary transfer path is *linear time-varying*, and the virtual filtered-reference signals are then generated as

$$r_v(n) = G_{vu}(n)x(n). \quad (8)$$

To generate the virtual filtered-reference signal, a model of the time-varying virtual secondary transfer path $G_{vu}(n)$ is thus needed in a practical implementation for every sample n . This is generally not possible in a practical situation, and a method for generating estimates of the virtual filtered-reference signals is therefore presented in Sec. V.

C. State-space model of system

In this section, a state-space model²³ of the acoustic duct arrangement in Fig. 1 is introduced. This state-space system models the transfer paths between the input signals into the primary and control sources, and the output signals at the physical microphones located at \mathbf{x} , the moving virtual microphone located at $x_v(n)$, and a number of spatially fixed virtual microphones located within the target zone at

$$\mathbf{x}_v = [x_{v1} \ x_{v2} \ \cdots \ x_{v\bar{M}_v}]^T. \quad (9)$$

The state-space system that models these transfer paths is given by

$$\begin{aligned} \mathbf{z}(n+1) &= \mathbf{A}\mathbf{z}(n) + \mathbf{B}_u u(n) + \mathbf{B}_x x(n), \\ \mathbf{e}_p(n) &= \mathbf{C}_p \mathbf{z}(n) + \mathbf{v}_p(n), \\ \bar{\mathbf{e}}_v(n) &= \bar{\mathbf{C}}_v \mathbf{z}(n) + \bar{\mathbf{v}}_v(n), \end{aligned} \quad (10)$$

$$e_v(n) = \mathbf{C}_v(n)\mathbf{z}(n) + v_v(n),$$

with $\mathbf{z}(n) \in \mathbb{R}^N$ the state of the plant, $u(n)$ the control signal, $x(n)$ the disturbance source signal, $\mathbf{e}_p(n) \in \mathbb{R}^{M_p}$ the physical error signals at the spatially fixed physical microphones, $\bar{\mathbf{e}}_v(n) \in \mathbb{R}^{\bar{M}_v}$ the virtual error signals at the virtual microphones spatially fixed at \mathbf{x}_v within the target zone, with

$$\bar{\mathbf{e}}_v(n) = [\bar{e}_{v1}(n) \ \bar{e}_{v2}(n) \ \cdots \ \bar{e}_{v\bar{M}_v}(n)]^T, \quad (11)$$

and $e_v(n)$ the virtual error signal at the moving virtual location $x_v(n)$. The signals $\mathbf{v}_p(n) \in \mathbb{R}^{M_p}$, $\bar{\mathbf{v}}_v(n) \in \mathbb{R}^{\bar{M}_v}$, and $v_v(n)$ in Eq. (10) are the measurement noise signals on the physical microphones, spatially fixed virtual microphones, and moving virtual microphone, respectively. The state-space matrices in Eq. (10) are real-valued and of appropriate dimensions.

One may wonder why measurement noise on the virtual microphones has been included in Eq. (10). The reason is that physical microphones are usually located at the virtual locations in a preliminary identification stage of the plant, and the measurement noise on these microphones thus has to be included in the analysis.

III. SPATIALLY FIXED VIRTUAL MICROPHONE

The LMS virtual microphone technique¹⁸ can be used to compute an estimate of the virtual error signals $\bar{\mathbf{e}}_v(n)$ at the spatially fixed virtual locations \mathbf{x}_v in Eq. (9), which are located throughout the target zone shown in Fig. 1. However, this algorithm¹⁸ has only been derived for the case of one

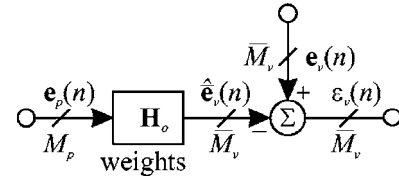


FIG. 2. Block diagram of the LMS virtual microphone technique using M_p physical microphones, and \bar{M}_v spatially fixed virtual microphones.

spatially fixed virtual microphone. In this section, this algorithm is therefore extended to include multiple spatially fixed virtual microphones.

A. LMS virtual microphone technique

A block diagram of the LMS virtual microphone technique using multiple spatially fixed virtual microphones is shown in Fig. 2.

An estimate $\hat{\mathbf{e}}_v(n)$ of the virtual error signals is calculated by summing the weighted physical error signals $\mathbf{e}_p(n)$, which can be expressed as

$$\hat{\mathbf{e}}_v(n) = \mathbf{H}^T \mathbf{e}_p(n), \quad (12)$$

where the matrix $\mathbf{H} \in \mathbb{R}^{M_p \times \bar{M}_v}$ contains the physical microphone weights given by

$$\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \cdots \ \mathbf{h}_{\bar{M}_v}]^T, \quad (13)$$

with $\mathbf{h}_i \in \mathbb{R}^{M_p}$ the weights used to compute the estimate $\bar{e}_{vi}(n) = \mathbf{h}_i^T \mathbf{e}_p(n)$. The physical error signals $\mathbf{e}_p(n) \in \mathbb{R}^{M_p}$ in Eq. (12) are defined as

$$\mathbf{e}_p(n) = \mathbf{d}_p(n) + \mathbf{y}_p(n), \quad (14)$$

with $\mathbf{d}_p(n)$ the physical primary disturbances, and $\mathbf{y}_p(n)$ the physical secondary disturbances. The aim of the filter problem illustrated in Fig. 2 is to find a set of real weights \mathbf{H} that minimize the virtual output error $\bar{\mathbf{e}}_v(n)$ given by

$$\bar{\mathbf{e}}_v(n) = \mathbf{e}_v(n) - \hat{\mathbf{e}}_v(n). \quad (15)$$

These weights will be called the optimal physical microphone weights, and are derived in the following.

B. Modified LMS virtual microphone technique

Using Eq. (14), the estimated virtual error signal in Eq. (12) can also be written as

$$\hat{\mathbf{e}}_v(n) = \mathbf{H}^T (\mathbf{d}_p(n) + \mathbf{y}_p(n)) = \hat{\mathbf{d}}_v(n) + \hat{\mathbf{y}}_v(n), \quad (16)$$

where $\hat{\mathbf{d}}_v(n)$ and $\hat{\mathbf{y}}_v(n)$ are the estimated virtual primary and secondary disturbances, respectively. Equation (16) illustrates that the microphone weights \mathbf{H} are applied to both the primary and secondary disturbances at the physical microphones. The underlying assumption made by previous researchers¹³ is therefore that the optimal microphone weights for the estimation of both $\bar{\mathbf{d}}_v(n)$ and $\bar{\mathbf{y}}_v(n)$ are equal, which might not always be true. For example, for active noise control in the near-field of the secondary source, the spatial characteristics of the primary and secondary fields can be very different.² This difference is accounted for in the

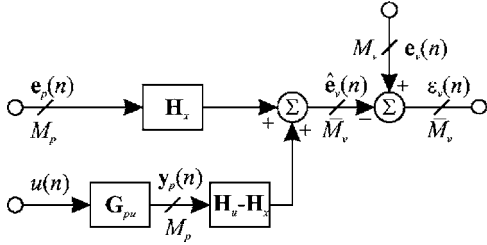


FIG. 3. Block diagram of the modified LMS virtual microphone technique using M_p physical microphones, and M_v spatially fixed virtual microphones.

virtual microphone arrangement,² where it is assumed that the primary field changes relatively little between the physical and virtual microphones, such that the primary disturbances at the two microphones are assumed equal. The secondary disturbances at the physical and virtual microphones are, however, not assumed equal due to the near-field properties of the secondary source. For the LMS virtual microphone technique, the described situation will result in different optimal microphone weights for the primary and secondary sound fields. In conclusion, it is thus important to find optimal microphone weights for the estimation of both the primary and secondary disturbances at the virtual microphone. If these weights are equal, the LMS virtual microphone technique can be implemented as illustrated in Fig. 2. If these weights are not equal, the virtual sensing algorithm needs to separate the physical error signals into their primary and secondary components. The optimal weights for each component can then be applied to obtain optimal estimates of $\bar{\mathbf{d}}_v(n)$ and $\bar{\mathbf{y}}_v(n)$, which can then be superimposed to get an optimal estimate $\hat{\mathbf{e}}_v(n)$ of the virtual error signal. The separation of the physical error signals into their primary and secondary components can be achieved as illustrated in Fig. 3.

The estimated virtual error signal in this figure is given by

$$\hat{\mathbf{e}}_v(n) = \mathbf{H}_x^T \mathbf{e}_p(n) + (\mathbf{H}_u - \mathbf{H}_x)^T \mathbf{G}_{pu} u(n), \quad (17)$$

where $\mathbf{H}_x \in \mathbb{R}^{M_p \times M_v}$ are the weights for the primary field, $\mathbf{H}_u \in \mathbb{R}^{M_p \times M_v}$ the weights for the secondary field, and \mathbf{G}_{pu} the physical secondary transfer path matrix between the control source and the M_p physical microphones. It can be seen that the block diagram in Fig. 3 reduces to the block diagram in Fig. 2 if the weights for the primary and secondary field are equal, such that $\mathbf{H}_x = \mathbf{H}_u$.

C. Optimal microphone weights

In this section, an optimal solution for the physical microphone weights is derived given the state-space model in Eq. (10). By setting $x(n)=0$ in this equation, optimal weights for the secondary field can be derived. Similarly, optimal weights for the primary field can be derived by setting $u(n)=0$. First, consider the case where the disturbance source signal $x(n)$ is set to zero. An estimate of the virtual secondary disturbances $\bar{\mathbf{y}}_v(n)$ is now computed as

$$\hat{\bar{\mathbf{y}}}_v(n) = \mathbf{H}_u^T \mathbf{y}_p(n). \quad (18)$$

From Eqs. (10) and (18), a state-space system that models the virtual output error $\bar{\mathbf{e}}_v(n) = \bar{\mathbf{y}}_v(n) - \hat{\bar{\mathbf{y}}}_v(n)$ is then given by

$$\mathbf{z}(n+1) = \mathbf{A}\mathbf{z}(n) + \mathbf{B}_u u(n), \quad (19)$$

$$\bar{\mathbf{e}}_v(n) = \bar{\mathbf{C}}_\varepsilon \mathbf{z}(n) + \bar{\mathbf{v}}_v(n) - \mathbf{H}_u^T \mathbf{v}_p(n),$$

where the matrix $\bar{\mathbf{C}}_\varepsilon \in \mathbb{R}^{M_v \times N}$ is defined as

$$\bar{\mathbf{C}}_\varepsilon = \bar{\mathbf{C}}_v - \mathbf{H}_u^T \mathbf{C}_p. \quad (20)$$

The optimal weights $\mathbf{H}_{uo} \in \mathbb{R}^{M_v \times M_p}$ are now defined as the weights that minimize the cost function J_ε defined as

$$J_\varepsilon = \text{tr} \mathbb{E}[\bar{\mathbf{e}}_v(n) \bar{\mathbf{e}}_v(n)^T]. \quad (21)$$

It is assumed that the signal $u(n)$ in Eq. (19) is a zero-mean white and stationary random process during identification of the weights. Furthermore, the measurement noise signals $\mathbf{v}_p(n)$ and $\bar{\mathbf{v}}_v(n)$ are assumed to be zero-mean white and stationary random processes that are uncorrelated to $u(n)$, such that the following covariance matrices can be defined:

$$\mathbb{E} \begin{pmatrix} u(n) \\ \mathbf{v}_p(n) \\ \bar{\mathbf{v}}_v(n) \\ \mathbf{z}(0) \end{pmatrix} \begin{bmatrix} u(k) & \mathbf{v}_p(k)^T & \bar{\mathbf{v}}_v(k)^T & \mathbf{z}(0)^T & 1 \end{bmatrix} = \begin{bmatrix} Q_u & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{R}_p & \bar{\mathbf{R}}_{pv} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{R}}_{pv}^T & \bar{\mathbf{R}}_v & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Pi}_0 & \mathbf{0} \end{bmatrix} \delta_{nk}, \quad (22)$$

with δ_{nk} the Kronecker delta function,²³ such that $\delta_{nk}=1$ if $n=k$, $\mathbf{z}(0) \in \mathbb{R}^N$ the initial state, \mathbf{R}_p the covariance matrix of the measurement noise on the physical microphones, $\bar{\mathbf{R}}_v$ the covariance matrix of the measurement noise on the spatially fixed virtual microphones, and $\bar{\mathbf{R}}_{pv}$ the cross-covariance matrix between the measurement noise on the physical and spatially fixed virtual microphones. In the Appendix, it is shown that the cost function in Eq. (21) can then be written as

$$J_\varepsilon = \text{tr}(\mathbf{H}_u^T \mathbf{R}_u \mathbf{H}_u - 2\mathbf{H}_u^T \bar{\mathbf{p}}_u + \bar{\mathbf{c}}_u), \quad (23)$$

with

$$\mathbf{R}_u = \mathbb{E}[\mathbf{y}_p(n) \mathbf{y}_p(n)^T] = \mathbf{C}_p \bar{\mathbf{\Pi}}_u \mathbf{C}_p^T + \mathbf{R}_p, \quad (24)$$

$$\bar{\mathbf{p}}_u = \mathbb{E}[\mathbf{y}_p(n) \bar{\mathbf{y}}_v(n)^T] = \mathbf{C}_p \bar{\mathbf{\Pi}}_u \bar{\mathbf{C}}_v^T + \bar{\mathbf{R}}_{pv}, \quad (25)$$

$$\bar{\mathbf{c}}_u = \mathbb{E}[\bar{\mathbf{y}}_v(n) \bar{\mathbf{y}}_v(n)^T] = \bar{\mathbf{C}}_v \bar{\mathbf{\Pi}}_u \bar{\mathbf{C}}_v^T + \bar{\mathbf{R}}_v, \quad (26)$$

where $\bar{\mathbf{\Pi}}_u > 0$ is the solution to the discrete-time Lyapunov equation²³

$$\bar{\mathbf{\Pi}}_u = \mathbf{A} \bar{\mathbf{\Pi}}_u \mathbf{A}^T + \mathbf{B}_u \mathbf{Q}_u \mathbf{B}_u^T. \quad (27)$$

The optimal weights can now be found by differentiating the cost function J_ε in Eq. (23) with respect to the weights \mathbf{H}_u ,

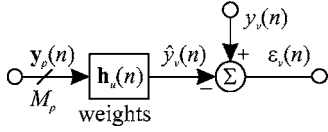


FIG. 4. Block diagram for deriving the optimal time-varying weights for the secondary field.

and setting all of the resulting derivatives to zero, which results in²²

$$\mathbf{H}_{uo} = \mathbf{R}_u^{-1} \bar{\mathbf{p}}_u. \quad (28)$$

Similarly, optimal weights $\mathbf{H}_{xo} = \mathbf{R}_x^{-1} \bar{\mathbf{p}}_x$ for the primary field can be computed, with

$$\mathbf{R}_x = E[\mathbf{d}_p(n) \mathbf{d}_p(n)^T] = \mathbf{C}_p \bar{\mathbf{\Pi}}_x \mathbf{C}_p^T + \mathbf{R}_p, \quad (29)$$

$$\bar{\mathbf{p}}_x = E[\mathbf{d}_p(n) \bar{\mathbf{d}}_v(n)^T] = \mathbf{C}_p \bar{\mathbf{\Pi}}_x \mathbf{C}_v^T + \bar{\mathbf{R}}_{pv}, \quad (30)$$

where $\bar{\mathbf{\Pi}}_x > 0$ is the solution to the discrete-time Lyapunov equation²³

$$\bar{\mathbf{\Pi}}_x = \mathbf{A} \bar{\mathbf{\Pi}}_x \mathbf{A}^T + \mathbf{B}_x \mathbf{Q}_x \mathbf{B}_x^T. \quad (31)$$

The above-presented derivations show that the weights for the primary and secondary field are equal if $\bar{\mathbf{\Pi}}_u = \bar{\mathbf{\Pi}}_x$.

IV. MOVING VIRTUAL MICROPHONE

In this section, the LMS virtual microphone technique introduced in Sec. III is extended to a moving virtual sensing algorithm. This algorithm can be used in the acoustic duct arrangement shown in Fig. 1 to estimate the virtual error signal at the moving virtual location $x_v(n)$. The LMS *moving virtual microphone technique* is derived in Sec. IV A assuming that the matrix $\mathbf{C}_v(n)$ in Eq. (10) is known at every sample n . This will not be the case in a practical situation, and a more practical implementation of the developed moving virtual sensing algorithm is therefore presented in Sec. IV B. This implementation is based on linear spatial interpolation between the estimated virtual error signals $\hat{\mathbf{e}}_v(n)$ at the spatially fixed virtual microphones located within the target zone at \mathbf{x}_v .

A. LMS moving virtual microphone technique

In this section, the analysis of the LMS virtual microphone technique presented in Sec. III is extended to develop the LMS moving virtual microphone technique. Because the virtual microphone is now moving, a *time-varying* optimal solution for the physical microphone weights is derived given the state-space model in Eq. (10). Here, time-varying optimal weights for the secondary field are derived by setting $x(n)=0$ in Eq. (10). A block diagram of the optimal filter problem for this case is shown in Fig. 4.

An estimate of the virtual secondary disturbance $y_v(n)$ at the moving virtual location $x_v(n)$ is now computed as

$$\hat{y}_v(n) = \mathbf{h}_u(n)^T \mathbf{y}_p(n), \quad (32)$$

with $\mathbf{h}_u(n) \in \mathbb{R}^{M_p}$ the time-varying weights for the secondary field. Using Eqs. (10) and (32), a state-space system that

models the virtual output error $\varepsilon_v(n) = y_v(n) - \hat{y}_v(n)$ is now given by

$$\mathbf{z}(n+1) = \mathbf{A} \mathbf{z}(n) + \mathbf{B}_u \mathbf{u}(n), \quad (33)$$

$$\varepsilon_v(n) = \mathbf{C}_\varepsilon(n) \mathbf{z}(n) + v_v(n) - \mathbf{h}_u(n)^T \mathbf{v}_p(n),$$

where the matrix $\mathbf{C}_\varepsilon(n) \in \mathbb{R}^{1 \times N}$ is defined as

$$\mathbf{C}_\varepsilon(n) = \mathbf{C}_v(n) - \mathbf{h}_u(n)^T \mathbf{C}_p. \quad (34)$$

The time-varying optimal weights $\mathbf{h}_{uo}(n) \in \mathbb{R}^{M_p}$ are defined as the weights that minimize the cost function $J_\varepsilon(n)$ at every sample n , with

$$J_\varepsilon(n) = E[\varepsilon_v(n)^2]. \quad (35)$$

Following the derivations presented in Sec. III C, it can be shown that this cost function can be written, similar to Eq. (23), as

$$J_\varepsilon(n) = \mathbf{h}_u(n)^T \mathbf{R}_u \mathbf{h}_u(n) - 2 \mathbf{h}_u(n)^T \mathbf{p}_u(n) + c_u(n), \quad (36)$$

with \mathbf{R}_u as defined in Eq. (24), and

$$\mathbf{p}_u(n) = E[\mathbf{y}_p(n) y_v(n)^T] = \mathbf{C}_p \bar{\mathbf{\Pi}}_u \mathbf{C}_v(n)^T + \mathbf{R}_{pv}, \quad (37)$$

$$c_u(n) = E[y_v(n) y_v(n)^T] = \mathbf{C}_v \bar{\mathbf{\Pi}}_u \mathbf{C}_v(n)^T + R_v, \quad (38)$$

where $\bar{\mathbf{\Pi}}_u > 0$ is the solution to the discrete-time Lyapunov equation in Eq. (27), R_v is the covariance of the measurement noise on the moving virtual microphone, and \mathbf{R}_{pv} the cross-covariance matrix between the measurement noise on the physical microphones and the moving virtual microphone, such that

$$\mathbf{R}_{pv} = E[\mathbf{v}_p(n) v_v(n)], \quad R_v = E[v_v(n)^2]. \quad (39)$$

The time-varying optimal weights can now be found by differentiating the cost function $J_\varepsilon(n)$ in Eq. (36) with respect to the weights $\mathbf{h}_u(n)$, and setting all of the resulting derivatives to zero at every sample n , which results in²²

$$\mathbf{h}_{uo}(n) = \mathbf{R}_u^{-1} \mathbf{p}_u(n). \quad (40)$$

Similarly, time-varying optimal weights for the primary field can be computed as $\mathbf{h}_{xo}(n) = \mathbf{R}_x^{-1} \mathbf{p}_x(n)$, with \mathbf{R}_x as defined in Eq. (29), and

$$\mathbf{p}_x(n) = E[\mathbf{d}_p(n) d_v(n)] = \mathbf{C}_p \bar{\mathbf{\Pi}}_x \mathbf{C}_v(n)^T + \mathbf{R}_{pv}, \quad (41)$$

where $\bar{\mathbf{\Pi}}_x > 0$ is the solution to the discrete-time Lyapunov equation in Eq. (31).

To implement the derived moving virtual sensing algorithm, the cross-covariance vectors $\mathbf{p}_u(n)$ and $\mathbf{p}_x(n)$ in Eqs. (37) and (41) need to be known at every sample n . This is generally not possible in practice, and a more practical implementation of the developed moving virtual sensing algorithm is therefore presented in the next section.

B. Practical implementation using spatial interpolation

The assumption that the moving virtual location $x_v(n)$ stays within the target zone shown in Fig. 1 is now used to arrive at a more practical implementation of the moving vir-

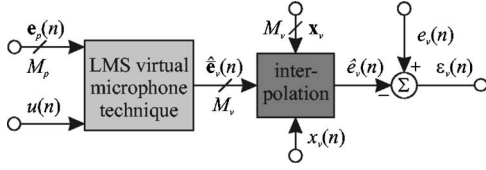


FIG. 5. Block diagram of the practical implementation of the LMS moving virtual microphone technique based on spatial interpolation, using M_p physical microphones, M_v virtual microphones spatially fixed at \mathbf{x}_v , and one moving virtual microphone located at $x_v(n)$.

tual sensing algorithm derived in the previous section. A block diagram of the proposed implementation based on spatial interpolation is shown in Fig. 5.

The approach taken here is to first compute an estimate $\hat{\mathbf{e}}_v(n)$ of the virtual error signals at each of the M_v spatially fixed virtual microphones located at \mathbf{x}_v defined in Eq. (9). These spatially fixed virtual microphones are positioned throughout the target zone in Fig. 1, and the estimate of the virtual error signals at each of these locations can be computed using the LMS virtual microphone technique described in Sec. III. As shown in Fig. 5, an estimate $\hat{e}_v(n)$ of the virtual error signal at the moving virtual location $x_v(n)$ is then computed using linear spatial interpolation between the error signals $\hat{\mathbf{e}}_v(n)$. Figure 5 shows that the information needed to compute the linear spatial interpolation are the estimated virtual error signals $\hat{\mathbf{e}}_v(n)$, the spatially fixed virtual locations \mathbf{x}_v at which these virtual error signals are estimated, and the moving virtual location $x_v(n)$, which is assumed known here.

It can be shown that an equivalent estimate of the virtual error signal can be computed as

$$\hat{e}_v(n) = \hat{\mathbf{h}}_{x_o}(n)^T \mathbf{e}_p(n) + (\hat{\mathbf{h}}_{u_o}(n) - \hat{\mathbf{h}}_{x_o}(n))^T \mathbf{G}_{pu} u(n), \quad (42)$$

where the weights $\hat{\mathbf{h}}_{u_o}(n) \in \mathbb{R}^{M_p}$ and $\hat{\mathbf{h}}_{x_o}(n) \in \mathbb{R}^{M_p}$ are estimates of the optimal time-varying weights defined in the previous section. These estimates are obtained using linear spatial interpolation between the weights \mathbf{H}_{u_o} and \mathbf{H}_{x_o} derived in Sec. III C for the spatially fixed virtual locations \mathbf{x}_v . As an example, the m th component $\hat{h}_{u_o}^m(n)$ of $\hat{\mathbf{h}}_{u_o}(n)$ is computed as

$$\hat{h}_{u_o}^m(n) = \frac{x_v(n) - x_{v(i+1)}}{x_{vi} - x_{v(i+1)}} \mathbf{H}_{u_o}^{(m,i)} + \frac{x_{vi} - x_v(n)}{x_{vi} - x_{v(i+1)}} \mathbf{H}_{u_o}^{(m,i+1)}, \quad (43)$$

with the moving virtual location $x_v(n)$ such that $x_{vi} \leq x_v(n) \leq x_{v(i+1)}$, and with x_{vi} the spatially fixed virtual locations defined in Eq. (9). If the weights for the primary and secondary field are equal, Eq. (42) reduces to

$$\hat{e}_v(n) = \hat{\mathbf{h}}_o(n)^T \mathbf{e}_p(n), \quad (44)$$

with $\hat{\mathbf{h}}_o(n) = \hat{\mathbf{h}}_{u_o}(n) = \hat{\mathbf{h}}_{x_o}(n)$. This is the implementation used in the experiments presented in Sec. VI.

V. FILTERED-x LMS ALGORITHM

A. Implementation

As discussed previously, to implement the filtered-x LMS algorithm in Eq. (3), the virtual filtered-reference signals $r_v(n) = G_{vu}(n)x(n)$ need to be generated, with $G_{vu}(n)$ the time-varying virtual secondary transfer path. In a practical situation, this transfer path will not be available for every sample n , and a method for generating an estimate $\hat{r}_v(n)$ of the virtual filter-reference signal needs to be developed. Here, an estimate of the virtual filtered-reference signal is generated as

$$\hat{r}_v(n) = \hat{\mathbf{h}}_{uo}(n)^T \mathbf{r}_p(n), \quad (45)$$

with $\hat{\mathbf{h}}_{uo}(n)$ the time-varying optimal weights defined in the previous section, and $\mathbf{r}_p(n) \in \mathbb{R}^{M_p}$ the vector of filtered-reference signals for the physical microphones, which is generated as

$$\mathbf{r}_p(n) = \mathbf{G}_{pu} x(n), \quad (46)$$

with \mathbf{G}_{pu} the physical secondary transfer path between the control source and the physical microphones. Substituting Eq. (46) into Eq. (45) gives

$$\hat{r}_v(n) = \hat{G}_{vu}(n)x(n), \quad (47)$$

with $\hat{G}_{vu}(n) = \hat{\mathbf{h}}_{uo}(n)^T \mathbf{G}_{pu}$. Effectively, this method thus provides an estimate of the time-varying virtual secondary transfer path $G_{vu}(n)$. The filtered-x LMS algorithm can now be implemented as

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \hat{\mathbf{r}}_v(n) \hat{e}_v(n), \quad (48)$$

with $\hat{e}_v(n)$ the estimated virtual error signal at the moving virtual location $x_v(n)$, and $\hat{\mathbf{r}}_v(n) \in \mathbb{R}^I$ given by

$$\hat{\mathbf{r}}_v(n) = [\hat{r}_v(n) \hat{r}_v(n-1) \cdots \hat{r}_v(n-I+1)]^T. \quad (49)$$

This is the implementation used in the experiments presented in Sec. VI.

B. Tracking

The aim of the filtered-x LMS algorithm in Eq. (48) is to track the changes in the statistical properties of the estimated virtual primary disturbance at the moving virtual location $x_v(n)$, and adjust the control filter coefficients $\mathbf{w}(n)$ accordingly. The amount and speed of tracking needed is dependent on both the temporal rate of change of the moving virtual location, and the spatial rate of change of the relative magnitude and phase between the primary and secondary sound field over the target zone. This spatial rate of change determines how much the filter coefficients need to be adjusted over the spatial region through which the virtual microphone is moving. This information, together with the temporal rate of change of the moving virtual location, determines the amount and speed of the tracking that is needed to successfully create a moving zone of quiet. However, as stated by Ref. 24, the tracking details of a time-varying system are very problem specific. As a result, the tracking behavior of

the suggested algorithm will be problem specific as well, and general statements on how fast the virtual microphone can be moved cannot be made.

As discussed in Ref. 24, the convergence rate and the tracking capability of an adaptive algorithm are generally two different properties. Whereas convergence is a *transient* phenomenon, tracking is a *steady-state* phenomenon. For an adaptive algorithm to exercise its tracking capability, it must therefore first pass from the transient mode to the steady-state mode.²⁴ In the experiments presented in Sec. VI, the performance obtained with the proposed method was measured after the filtered-x LMS algorithm in Eq. (48) had passed from the transient mode to the steady-state mode. Thus, in the experiments presented in Sec. VI, it is the *tracking* capability of the filtered-x LMS algorithm that was investigated.

VI. ACOUSTIC DUCT EXPERIMENTS

The algorithms introduced in the previous sections were implemented on the acoustic duct arrangement shown in Fig. 1. The experimental arrangement is now described in more detail, after which experimental results are presented and discussed.

A. Experimental arrangement

The rectangular acoustic duct in Fig. 1 that was used in the real-time experiments was of length $L=4.830$ m, width 0.205 m, and height 0.205 m. A loudspeaker located at $x_p=4.730$ m was used as a primary source, and another loudspeaker located at $x_s=0.500$ m as a control source. The primary loudspeaker was excited by a tonal disturbance source signal $x(n)$. As shown in Fig. 1, this signal was also used as a feedforward reference signal in the filtered-x LMS algorithm. A physical microphone array consisting of two electret microphones was located at $\mathbf{x}=[1.425 \ 1.475]$ m. A traversing microphone mounted on a cable wrapped around pulleys at each end of the duct was located inside the duct. This cable was wound onto a 0.150m-diam spool, which was mounted on the shaft of a dc servo-motor encoder unit. This encoder provided dual track TTL output signals of 500 pulses per revolution, which enabled accurate position control of the traversing microphone. In a preliminary identification procedure, the traversing microphone was placed at a number of spatially fixed virtual locations \mathbf{x}_v in order to determine optimal microphone weights for these locations. Furthermore, the traversing microphone was position controlled to measure the primary and controlled sound pressure at the moving virtual location $x_v(n)$. To implement the developed algorithms in real-time, a host-target software program called *xPC TARGET*[®] was used. A sampling frequency of $f_s=4$ kHz was employed in the real-time experiments.

B. Experimental results

The filtered-x LMS algorithm discussed in Sec. V was implemented in the acoustic duct arrangement to create a moving zone of quiet inside the acoustic duct. Only $I=2$ filter coefficients were used in the experiments presented here, since the duct was excited by a tonal excitation

signal.²² For broadband noise, more filter coefficients need to be used. The aim of the experiments was to estimate and minimize the virtual error signal $e_v(n)$ at a moving virtual location $x_v(n)=x_2+v(n)$ that changed sinusoidally with time, with $v(n)$ the moving virtual distance. The expression governing the desired position of the virtual microphone is given by

$$v(n) = 0.070 + 0.050 \sin\left(\frac{2\pi n}{T_v f_s}\right), \quad (50)$$

where T_v is the period of the sinusoidally time-varying moving virtual distance $v(n)$ in Fig. 1. The virtual microphone is thus moving sinusoidally between a virtual distance bounded by 0.020 and 0.120 m. The performance at the moving virtual distance was measured for two excitation frequencies f of 213 and 249 Hz. These frequencies correspond to the sixth and seventh resonance frequencies of the acoustic duct. For these excitation frequencies, the performance at the moving virtual distance $v(n)$ was measured for three different values of T_v in Eq. (50) given by 10, 5, and 2.5 s. The maximum amplitude of the sinusoidally time-varying velocity of the moving virtual microphone is thus given by $5 \text{ cm} \times 2\pi/2.5 \text{ s} \approx 12.6 \text{ cm/s}$. This was considered to be representative of the likely motion of a head in the intended applications. In total, six experiments were thus conducted in order to measure the performance of the implemented algorithms for various speeds of the moving virtual microphone, and various spatial characteristics of the sound field through which the virtual microphone is moving.

1. Preliminary identification procedure

In a preliminary identification procedure, the optimal microphone weights \mathbf{H}_{u_o} for the secondary field given in Eq. (28) were determined for $\bar{M}_v=16$ spatially fixed virtual locations $\mathbf{x}_v=x_2+\mathbf{v}$, with \mathbf{v} given by

$$\mathbf{v} = [0.000 \ 0.010 \ 0.020 \ \cdots \ 0.150] \text{ m}. \quad (51)$$

These spatially fixed virtual distances \mathbf{v} were evenly positioned throughout the target zone located within a virtual distance range of 0.000–0.150 m. The acoustic duct was excited with bandpass filtered white noise in the frequency range of 50–500 Hz while determining the optimal microphone weights. For each of the spatially fixed virtual distances in Eq. (51), the optimal weights were calculated using Eq. (28) based on 30 s of data obtained from the two physical microphones and the traversing microphone positioned at the spatially fixed virtual distance of interest. The results of this procedure are shown in Fig. 6, where the optimal weights for the secondary sound field have been plotted against \mathbf{v} .

As discussed in Sec. III, it is necessary to verify if the identified optimal weights for the primary and secondary sound fields are equal. Optimal microphone weights \mathbf{H}_{x_o} for the primary sound field were determined in a similar way, and it was observed that the resulting weights for both cases

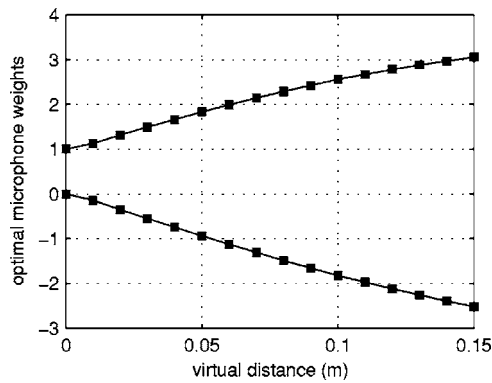


FIG. 6. Optimal physical microphone weights \mathbf{h}_o , plotted against fixed virtual distance v .

were almost identical. The LMS virtual microphone technique as shown in Fig. 2 was therefore implemented in the acoustic duct.

Using the optimal microphone weights shown in Fig. 6, the LMS virtual microphone technique was implemented in the acoustic duct for a number of spatially fixed virtual distances. The tonal attenuations achieved at these virtual distances were measured for excitation frequencies of 213 and 249 Hz. The results are shown in Fig. 7, which confirms previous experimental results presented by other researchers,^{13,20} and indicates that good performance can be achieved at these fixed virtual distances when using the LMS virtual microphone technique.

2. Relative spatial change of primary and secondary sound fields over target zone

In Sec. V, it was discussed that the amount and speed of the tracking needed from the filtered-x LMS algorithm to successfully create a moving zone of quiet is dependent on the spatial rate of change of the relative magnitude and phase between the primary and secondary sound field over the target zone, and the temporal rate of change of the moving virtual location $x_v(n)$. The measured spatial rate of change of the relative magnitude and phase between the primary and secondary sound field inside the acoustic duct over the target zone has been plotted in Fig. 8, for the excitation frequencies of 213 and 249 Hz. In Fig. 8 the relative magnitude and

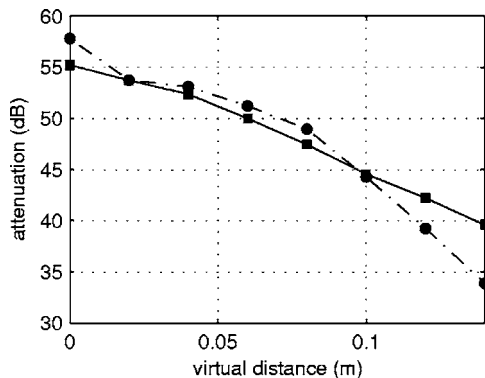


FIG. 7. Average tonal attenuation achieved at a number of fixed virtual distances: (—) $f=213$ Hz; (- - -) $f=249$ Hz.

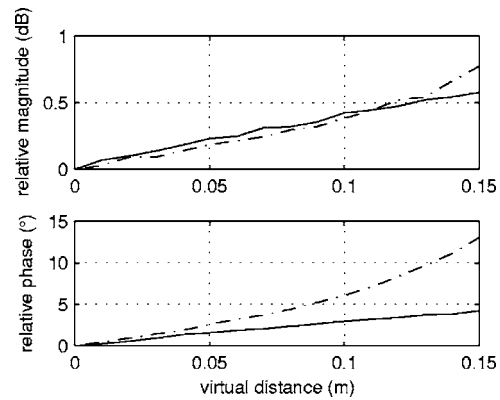


FIG. 8. Spatial change of relative magnitude and phase between primary and secondary sound fields plotted over target zone inside duct. (—) $f=213$ Hz and (- - -) $f=249$ Hz.

phase between the primary and secondary sound fields have been normalized against the relative magnitude and phase at a virtual distance of $v=0.000$ m.

For an excitation frequency of 213 Hz, the relative magnitude and phase change by about 0.58 dB and 4.2° over the target zone, respectively. For an excitation frequency of 249 Hz, the relative magnitude and phase change by about 0.77 dB and 13.0° over the target zone, respectively. Thus, when the virtual location is moving through the target zone located between a virtual distance v bounded by 0.000 and 0.150 m, the filtered-x LMS algorithm has to account for these changes in the relative magnitude and phase between the primary and secondary sound field by adjusting the filter coefficients, such that tracking of the nonstationarities in the estimated virtual primary disturbance is achieved. In the experimental results presented next, the performance at the moving virtual location was measured after convergence of the filtered-x LMS algorithm, such that the tracking capability of the adaptive algorithm was investigated.

3. Performance at moving virtual location

To illustrate the increase in local control performance that can be obtained at a moving virtual location when using the suggested method, the performance at the moving virtual location $x_v(n)=x_2+v(n)$ was also measured for the case of active noise control at a spatially fixed physical microphone located at $v=0.000$ m, and active noise control at a spatially fixed virtual microphone located at $v=0.020$ m. For the spatially fixed physical microphone, the physical error signal was directly measured by the physical microphone located at $x_2=1.475$ m, and was minimized using the standard formulation of the filtered-x LMS algorithm.²² For the spatially fixed virtual microphone, the virtual error signal at $v=0.020$ m was estimated using the LMS virtual microphone technique described in Sec. III. The estimate was then minimized using the filtered-x LMS algorithm. Unlike the currently proposed method, both of these active noise control systems cannot account for the fact that the desired location of the zone of quiet is not spatially fixed.

The performance of the three active noise control systems that each employ a different sensing method was measured at the moving virtual location $x_v(n)=x_2+v(n)$, where

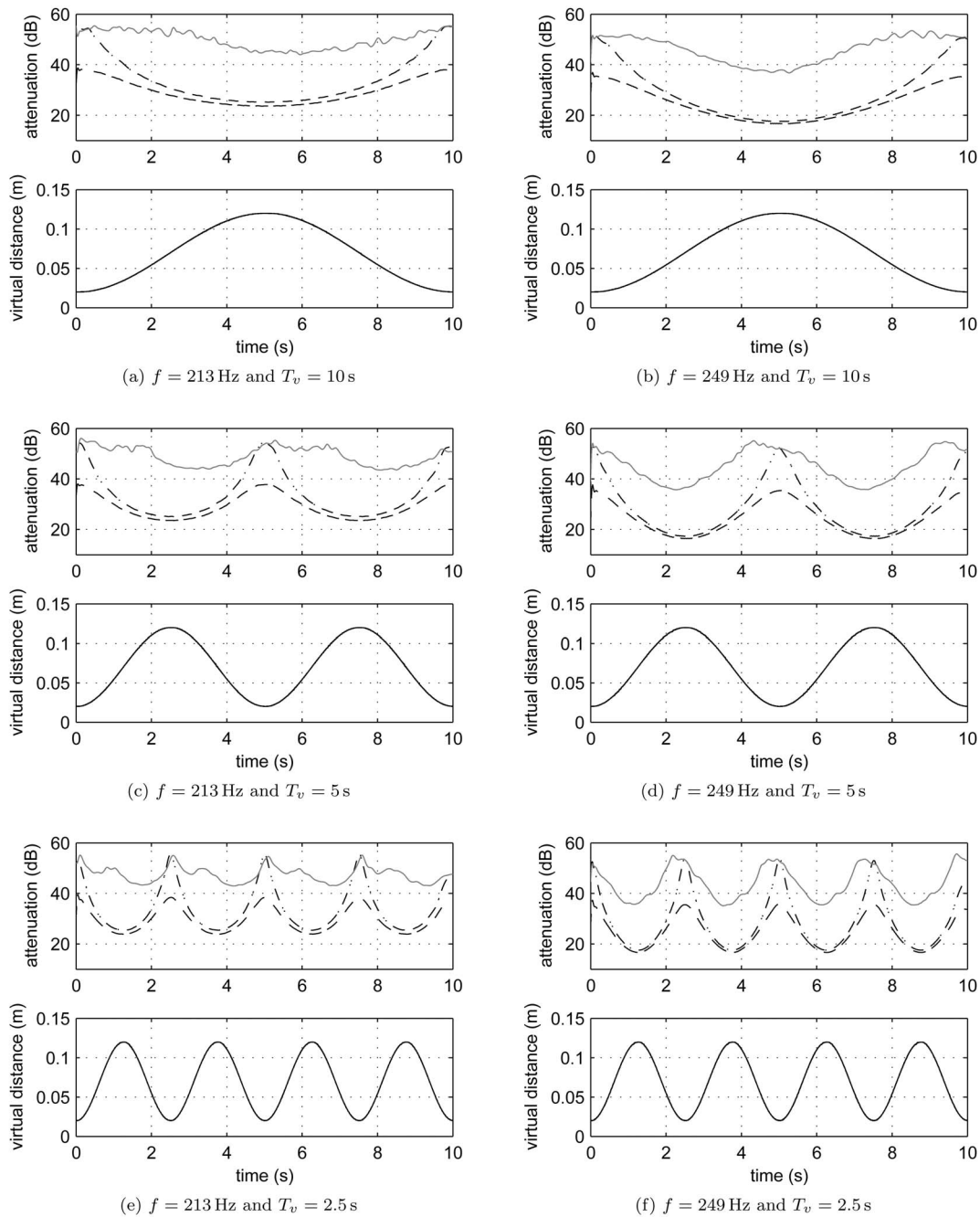


FIG. 9. (Bottom panel) Moving virtual distance $v(n)$ plotted against time. (Top panel) Average tonal attenuation at moving virtual distance plotted against time for active noise control at (- - -) physical microphone spatially fixed at $v=0.000$ m; (- · -) virtual microphone spatially fixed at $v=0.020$ m; and (—) moving virtual at microphone at $v(n)$.

$v(n)$ is defined by Eq. (50) with T_v given by 10, 5, and 2.5 s, respectively. The results of these measurements are illustrated in Fig. 9, where the average tonal attenuation measured at the moving virtual location has been plotted against time for excitation frequencies of 213 and 249 Hz. In Fig. 9, the dashed line is the average tonal attenuation measured using active noise control at the spatially fixed physical microphone located at $v=0.000$ m, the dash-dotted line the average tonal attenuation measured using active noise control at the spatially fixed virtual microphone located at $v=0.020$ m, and the solid gray line the average tonal attenua-

tion measured using active noise control at the moving virtual microphone that tracks the moving virtual distance $v(n)$. Each of these lines was generated by averaging the results of 30 data sets of 10 s measured with the traversing microphone, which was position controlled to track the moving virtual distance $v(n)$. The resulting position of the traversing microphone as measured by the encoder is plotted against time in the bottom parts of the subfigures in Fig. 9. Furthermore, the average tonal attenuations in decibels were low-pass filtered in order to prevent noisy plots in Fig. 9.

Figure 9 indicates that using a spatially fixed virtual microphone gives better performance at the moving virtual location than using a spatially fixed physical microphone. However, the best results are obtained when using a moving virtual microphone that tracks the desired location of the zone of quiet defined by the moving virtual distance $v(n)$. When the moving virtual distance is at $v=0.020$ m in Fig. 9, the system that uses a moving virtual microphone gives similar results to the system that uses a spatially fixed virtual microphone, indicating that the filtered-x LMS algorithm is able to provide sufficient tracking of the nonstationarities in the estimated virtual primary disturbance. When the moving virtual distance is not at $v=0.020$ m, however, Fig. 9 shows that the active noise control system that uses a moving virtual microphone provides the best performance. The results in Fig. 9 indicate that a moving zone of quiet has effectively been created at the moving virtual location in the acoustic duct, for all of the six narrowband experiments that were conducted. When using a moving virtual microphone, the average tonal attenuation at the moving virtual location does not fall below 42 dB for all values of T_v and an excitation frequency of 213 Hz. For the spatially fixed physical and virtual microphones, the average tonal attenuation at this frequency reduces to 24 and 25 dB, respectively, at a virtual distance of $v=0.120$ m. For an excitation frequency of 249 Hz, the average tonal attenuation at the moving virtual location does not fall below 36 dB for all values of T_v when using a moving virtual microphone. For the spatially fixed physical and virtual microphones, the average tonal attenuation at this frequency reduces to 17 and 18 dB, respectively, at a virtual distance of $v=0.120$ m. These results indicate that for the narrowband acoustic duct experiments presented here, the filtered-x LMS algorithm is able to provide sufficient tracking of the nonstationarities in the estimated virtual primary disturbance. As a result, a moving zone of quiet is effectively created inside the acoustic duct for narrowband noise.

VII. CONCLUSIONS

The LMS virtual microphone technique is a virtual sensing method that can be used to estimate the error signal at a virtual microphone that is spatially fixed. In this paper, this technique has been extended to a moving virtual sensing algorithm that is able to estimate the error signal at a moving virtual microphone which is tracking the desired location of the zone of quiet. The proposed algorithm can be combined with the filtered-x LMS algorithm to create a moving zone of quiet. The filtered-x LMS algorithm can provide the tracking needed to account for the nonstationarities in the primary disturbance, which are caused by the movement of the virtual microphone. A practical application of the suggested method is the creation of a moving zone of quiet that tracks a person's head. Here, the developed algorithm was implemented in a one-dimensional acoustic duct, and results of narrowband control experiments were presented to validate the proposed method. These results showed that the proposed algorithm was able to create a moving zone of quiet inside an acoustic duct for narrowband noise. This resulted in an in-

crease in local control performance compared to using a spatially fixed virtual microphone or a spatially fixed physical microphone. The experimental results indicate that the developed algorithm has the potential to improve the scope of successful local active noise control applications. The proposed algorithm can also be used for broadband active noise control, and in three-dimensional sound fields. Ongoing research is currently being conducted to investigate if the filtered-x LMS algorithm can provide the necessary tracking for broadband noise, and to analyze the performance of the suggested method in a three-dimensional sound field.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the University of Adelaide for providing an ASI scholarship, and the Australian Research Council for supporting this research. We also acknowledge the editor, Dr. Kenneth A. Cunefare, for his constructive comments on the original manuscript.

APPENDIX: DERIVATION OF COST FUNCTION

Equations (23) and (27) can be derived as follows. First, the state covariance matrix $\mathbf{\Pi}_u(n)$ is defined as

$$\mathbf{\Pi}_u(n) = E[\mathbf{z}(n)\mathbf{z}(n)^T]. \quad (\text{A1})$$

From Eq. (10), the state covariance matrix satisfies the recursion

$$\begin{aligned} \mathbf{\Pi}_u(n+1) = & \mathbf{A}\mathbf{\Pi}_u(n)\mathbf{A}^T + \mathbf{A}E[\mathbf{z}(n)u(n)^T]\mathbf{B}_u^T \\ & + \mathbf{B}_uE[u(n)\mathbf{z}(n)^T]\mathbf{A}_u^T + \mathbf{B}_u\mathbf{Q}_u\mathbf{B}_u^T. \end{aligned} \quad (\text{A2})$$

It can be shown, using Eqs. (10) and (22), that the current state $\mathbf{z}(n)$ is uncorrelated to the current and past inputs $\{u(k), k=1, \dots, n\}$,²³ such that

$$E[\mathbf{z}(n)u(n)^T] = 0. \quad (\text{A3})$$

This can be seen by deriving the following expression from Eq. (10):

$$\mathbf{z}(n) = \mathbf{A}^n\mathbf{z}(0) + \sum_{m=1}^n \mathbf{A}^{n-m}\mathbf{B}_u u(m-1). \quad (\text{A4})$$

The state $\mathbf{z}(n)$ is thus a linear combination of the initial state $\mathbf{z}(0)$ and the past inputs $\{u(k), k=1, \dots, n-1\}$. From Eq. (22), the input $u(n)$ is uncorrelated to all of these variables, thereby arriving at Eq. (A3). Equation (A2) now reduces to

$$\mathbf{\Pi}_u(n+1) = \mathbf{A}\mathbf{\Pi}_u(n)\mathbf{A}^T + \mathbf{B}_u\mathbf{Q}_u\mathbf{B}_u^T. \quad (\text{A5})$$

When the state $\mathbf{z}(n)$ reaches its mean steady state value, the state covariance matrix $\mathbf{\Pi}_u(n+1) = \mathbf{\Pi}_u(n) = \bar{\mathbf{\Pi}}_u$ in Eq. (A5), and solving the discrete-time Lyapunov equation in Eq. (27) thus gives the steady state solution $\bar{\mathbf{\Pi}}_u$.

The expression for the cost function J_ε given in Eq. (23) is derived next. Using a similar reasoning that was used to derive Eq. (A3), it can be shown that the current state $\mathbf{z}(n)$ is uncorrelated to the current and past measurement noise signals $\{\mathbf{v}_p(k), \bar{\mathbf{v}}_v(k), k=1, \dots, n\}$,²³ such that

$$E[\mathbf{z}(n)\mathbf{v}_p(n)^T] = 0, \quad E[\mathbf{z}(n)\bar{\mathbf{v}}_v(n)^T] = 0. \quad (\text{A6})$$

Using Eqs. (22), (A3), and (A6), the cost function in Eq. (21) can now be written as defined in Eq. (23).

- ¹S. J. Elliott, P. Joseph, A. J. Bullmore, and P. A. Nelson, "Active cancellation at a point in a pure tone diffuse sound field," *J. Sound Vib.* **120**, 183–189 (1988).
- ²S. J. Elliott and A. David, "A virtual microphone arrangement for local active sound control," *Proceedings of the 1st International Conference on Motion and Vibration Control*, Yokohama, 1992, pp. 1027–1031.
- ³J. Garcia-Bonito, S. J. Elliott, and C. C. Boucher, "Generation of zones of quiet using a virtual microphone arrangement," *J. Acoust. Soc. Am.* **101**, 3498–3516 (1997).
- ⁴A. Roure and A. Albarrazin, "The remote microphone technique for active noise control," *Proceedings of Active '99*, 1999, Ft. Lauderdale, Florida, USA, pp. 1233–1244.
- ⁵B. S. Cazzolato, "Sensing systems for active control of sound transmission into cavities," Ph.D. thesis, Department of Mechanical Engineering, The University of Adelaide, Adelaide, Australia, 1999.
- ⁶B. Rafaely, J. Garcia-Bonito, and S. J. Elliott, "Feedback control of sound in headrest," *Proceedings of Active '97*, Budapest, Hungary, 1997, pp. 445–456.
- ⁷B. Rafaely, S. J. Elliott, and J. Garcia-Bonito, "Broadband performance of an active headrest," *J. Acoust. Soc. Am.* **106**, 787–793 (1999).
- ⁸M. Pawelczyk, "Noise control in the active headrest based on estimated residual signals at virtual microphones," *Proceedings of the 10th International Congress on Sound and Vibration*, 2003, Stockholm, Sweden, pp. 251–258.
- ⁹J. L. Nielsen, A. Sæbø, G. Ottesen, T. A. Reinen, and S. Sørdsal, "A local active noise control system for locomotive drivers," *Proceedings of Inter-Noise '00*, Nice, France, 2000, pp. 1–6.
- ¹⁰J. Garcia-Bonito, S. J. Elliott, and C. C. Boucher, "A virtual microphone arrangement in a practical active headrest," *Proceedings of Inter-Noise '96*, Liverpool, 1996, pp. 1115–1120.
- ¹¹S. R. Popovich, "Active acoustic control in remote regions," U.S. Patent No. 5,701,350 (1997).
- ¹²C. D. Kestell, "Active control of sound in a small single engine aircraft cabin with virtual error sensor," Ph.D. thesis, Department of Mechanical Engineering, The University of Adelaide, Adelaide, Australia, 2000.
- ¹³J. M. Munn, "Virtual sensors for active noise control," Ph.D. thesis, Department of Mechanical Engineering, The University of Adelaide, Adelaide, Australia, 2003.
- ¹⁴C. D. Kestell, C. H. Hansen, and B. S. Cazzolato, "Active noise control in a free field with virtual error sensors," *J. Acoust. Soc. Am.* **109**, 232–243 (2000).
- ¹⁵C. D. Kestell, C. H. Hansen, and B. S. Cazzolato, "Active noise control with virtual sensors in a long narrow duct," *Int. J. Acoust. Vib.* **5**, 1–14 (2000).
- ¹⁶J. M. Munn, C. D. Kestell, B. S. Cazzolato, and C. H. Hansen, "Real-time feedforward control using virtual error sensors in a long narrow duct," *Proceedings of the Annual Australian Acoustical Society Conference*, Adelaide, Australia, 2001.
- ¹⁷J. M. Munn, B. S. Cazzolato, C. H. Hansen, and C. D. Kestell, "Higher-order virtual sensing for remote active noise control," *Proceedings of Active '02*, ISVR, Southampton, UK, 2002, pp. 377–386.
- ¹⁸B. S. Cazzolato, "An adaptive LMS virtual microphone," *Proceedings of Active '02*, ISVR, Southampton, UK, 2002, pp. 105–116.
- ¹⁹H. J. Gawron and K. Schaaf, "Interior car noise: Active cancellation of harmonics using virtual microphone," *Proceedings of the Second International Conference on Vehicle Comfort: Ergonomic, Vibrational and Thermal Aspects*, Bologna, Italy, 1992, pp. 739–748.
- ²⁰J. M. Munn, B. S. Cazzolato, and C. H. Hansen, "Virtual sensing: Open loop vs adaptive lms," *Proceedings of the Annual Australian Acoustical Society Conference*, 2002, Adelaide, Australia, pp. 24–33.
- ²¹B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall Signal Processing Series (Prentice-Hall, Upper Saddle River, NJ, 1985).
- ²²S. J. Elliott, *Signal Processing for Active Control*, 1st ed. (Academic, New York, 2001).
- ²³T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation* (Prentice Hall, Upper Saddle River, NJ, 2000).
- ²⁴S. Haykin, *Adaptive Filter Theory*, 4th ed. (Prentice Hall, Upper Saddle River, NJ, 2002).

Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution

Paul Dickens,^{a)} John Smith, and Joe Wolfe

School of Physics, The University of New South Wales, Sydney 2052, Australia

(Received 12 August 2006; revised 4 December 2006; accepted 27 December 2006)

Resonances and/or singularities during measurement and calibration often limit the precision of acoustic impedance spectra. This paper reviews and compares several established techniques, and describes a technique that incorporates three features that considerably improve precision. The first feature is to minimize problems due to resonances by calibrating the instrument using up to three different acoustic reference impedances that do not themselves exhibit resonances. The second involves using multiple pressure transducers to reduce the effects of measurement singularities. The third involves iteratively tailoring the spectrum of the stimulus signal to control the distribution of errors across the particular measured impedance spectrum. Examples are given of the performance of the technique on simple cylindrical waveguides. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2434764]

PACS number(s): 43.58.Bh, 43.75.Pq [AJZ]

Pages: 1471–1481

I. INTRODUCTION

The input impedance of any one-dimensional waveguide is defined as the complex ratio of pressure to volume flow at the input. This quantity is used to describe the linear acoustics of automotive mufflers, air-conditioning ducts, and the passive elements of wind instruments and the vocal tract. For a musical instrument, the input impedance usefully displays important characteristics of the instrument in the absence of a player, and indicates how the instrument will respond when excited at any frequency. In this case, high resolution in magnitude and frequency are particularly important. If pressure and volume flow are measured at different points in a system, their ratio gives a transfer impedance, which is particularly useful in characterizing multiport systems.

In this paper, we review the various approaches to measuring acoustic impedance and calibrating impedance heads and propose a general calibration technique for heads with multiple transducers. We consider the effect of transducer errors on impedance measurements and present a technique for distributing any measurement errors over the frequency range. To demonstrate the technique we use an impedance head with three microphones to measure the input impedance of simple cylindrical waveguides. The effects of calibration and optimization on these measurements are presented and discussed.

II. REVIEW OF MEASUREMENT TECHNIQUES

Many techniques for measuring acoustic impedance have been devised. The major techniques are reviewed by Benade and Ibbi¹ and Dalmont.² Any two transducers with responses that are linear functions of pressure and flow may be used to construct an impedance head; hence many designs are possible. In Table I, several common techniques are il-

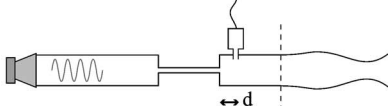
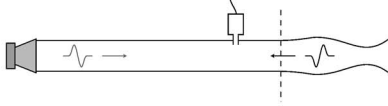
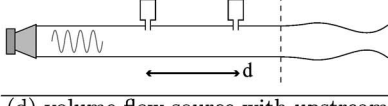
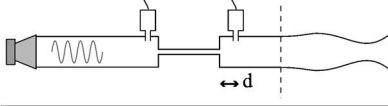
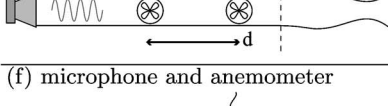
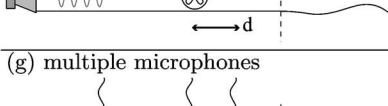
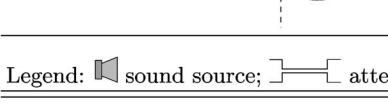




lustrated, along with conditions for singularities. At a singularity, the system of equations governing an impedance head becomes degenerate, and the impedance cannot be determined (see Sec. V E).

In methods (a) and (b) (Table I), a single pressure transducer (microphone) is used. In the volume flow source method (a) the input impedance is proportional to the ratio of the pressure measured with the unknown load to that with a reference load using the same stimulus. An attenuator ensures a source of volume flow, provided the impedance of the attenuator is much greater than that of the unknown load. (The effect of a finite source impedance can be reduced, to first order, by subtracting the attenuator admittance from the measured admittance.) In (b), a method known as pulse reflectometry, a pressure pulse is recorded as it travels toward the load, and again as it returns, yielding the impulse response function. This is mainly used for area reconstruction of musical wind instruments³ and the airway,⁴ however the acoustic impedance can be obtained from the impulse response function after a Fourier transform.

In methods (c) to (e) two similar transducers are used simultaneously. In methods (c) and (e), neither transducer measures pressure or flow at the input to the load alone; these are obtained by computations involving the transfer functions of the duct and the transducer properties. If a linear attenuator is present between the two microphones, as in method (d), the pressure difference between the two microphones is proportional to the flow. Alternatively, a signal approximately proportional to flow may be obtained by measuring the pressure in a fixed cavity at the back of the driver.⁵ Impedance heads have also been devised using a pressure transducer and a flow transducer (f), yielding the impedance with a minimum of computation (provided both transducers are close enough to the reference plane). Because of the difficulties in measuring flow precisely, these have limited dynamic range and signal-to-noise ratio.

^{a)}Electronic mail: pdickens@phys.unsw.edu.au

TABLE I. Several of the more common impedance heads, with conditions for singularities, selected references, and notes. In the expressions for singularities, k is the wave number and $n=1,2,3,\dots$. The reference plane is indicated by a vertical dashed line.

Impedance Head	Singularities	Refs.	Notes
(a) volume flow source 	$kd = (2n - 1)\frac{\pi}{2}$	1, 5-7	<ul style="list-style-type: none"> • computationally simple • requires calibration of the source • prone to errors at high Z
(b) pulse reflectometer 	f -range limited by 3, 4 pulse width		<ul style="list-style-type: none"> • uses same microphone for incident and reflected wave; calibration unnecessary • accuracy limited by length of measurement duct
(c) two microphones 	$kd = (n - 1)\pi$	8-15	<ul style="list-style-type: none"> • fewer simplifying assumptions required • computationally intensive
(d) volume flow source with upstream microphone 	$kd = (2n - 1)\frac{\pi}{2}$	16-18	<ul style="list-style-type: none"> • signal from upstream microphone proportional to flow (for an attenuator with high impedance compared to the unknown load)
(e) two anemometers 	$kd = (n - 1)\pi$	19	<ul style="list-style-type: none"> • computationally similar to (c) • several particle velocity sensors can be made simply with similar characteristics
(f) microphone and anemometer 	$kd = (2n - 1)\frac{\pi}{2}$	20-22	<ul style="list-style-type: none"> • direct measurement of both pressure and flow • correction required to obtain volume flow from particle velocity
(g) multiple microphones 	vary with microphone spacings	23	<ul style="list-style-type: none"> • wide frequency range • increased precision
Legend:  sound source;  attenuator;  microphone;  flow sensor			

The signal-to-noise ratio and frequency range may both be increased if an array of more than two transducers is used. Such a system with three microphones is shown in method (g).

III. THEORY OF ACOUSTIC IMPEDANCE MEASUREMENTS

A general impedance head is shown in Fig. 1. Some source of acoustic energy (shown here as a loudspeaker) excites the air inside a conduit (usually cylindrical) and transducers along the conduit measure some signal proportional to pressure and flow. The impedance is measured at some reference plane, to which load impedances may be attached.

The air inside the conduit is excited at frequencies below the cut-on frequency of the first higher mode, which for a cylindrical duct of radius a occurs at $f=1.84c/2\pi a$, where c is the speed of sound.²⁴ Therefore, all modes except the

plane wave mode are nonpropagating. If higher modes are excited near the transducers and the reference plane, the measured impedance will be something other than the plane-wave impedance. The effect of discontinuities at the transducers may be removed by calibration. If a discontinuity exists at the reference plane, nonpropagating modes are excited, introducing errors into the measured plane wave im-

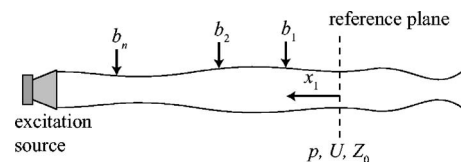


FIG. 1. A generalized impedance head. Impedance is measured at the *reference plane*, where the pressure, volume flow, and characteristic impedance are given by p , U , and Z_0 , respectively. The volume flow is positive flowing *into* the unknown load. n sensors record the signals b_1, b_2, \dots, b_n and are at positions x_1, x_2, \dots, x_n , measured from the reference plane.

pedance. Hence it is preferable that the object under study couple smoothly to the measurement conduit—otherwise multimodal theory can be used to calculate a correction (see Sec. V C).

It should be noted that in some impedance heads, the source and any transducers are located at or very near the reference plane; the sketch in Fig. 1 should be considered a general case only. However, it is often desirable to separate the source from the transducers, to ensure that the acoustic waves are planar at the transducers, subject to the constraints discussed in Sec. V.

For an impedance head with n transducers ($n \geq 2$), the pressure p and flow U at the reference plane are given by the vector \mathbf{x} in a matrix equation of the form $\mathbf{Ax} = \mathbf{b}$:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ \vdots & \vdots \\ A_{n1} & A_{n2} \end{bmatrix} \begin{bmatrix} p \\ Z_0 U \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad (1)$$

where \mathbf{b} is a vector of transducer signals and the elements of each matrix are, in general, functions of frequency. In order for the matrix \mathbf{A} to be dimensionless, any quantities with units of volume flow are parametrized by the characteristic impedance of the head at the input, Z_0 . Thus \mathbf{x} and \mathbf{b} both have units of pressure.

For ideal transducers positioned a known distance from the reference plane and mounted in a cylindrical duct, the elements of \mathbf{A} are given by the transfer matrix for a straight tube. The signal from an ideal pressure transducer with unity gain at position x is given by

$$b_{\text{pressure}}(x) = \cosh(ikx)p + \sinh(ikx)Z_0U \quad (2)$$

and the signal [$b_{\text{flow}}(x) \equiv Z_0U(x)$] from an ideal flow transducer (again at x and with unity gain) is given by

$$b_{\text{flow}}(x) = \sinh(ikx)p + \cosh(ikx)Z_0U, \quad (3)$$

where $k = \omega/v - i\alpha$ where $i = \sqrt{-1}$ and v and α (the phase velocity and attenuation coefficient) are calculated taking into account viscothermal loss (see, e.g., Fletcher and Rossing²⁴). From these, the matrix \mathbf{A} may be built up for any combination of transducers.

Once \mathbf{A} is determined, the pressure and flow (and hence the impedance) for a given measurement \mathbf{b} are obtained by solving Eq. (1). The equation is solved in the normal algebraic sense for $n=2$. For $n > 2$, there are more equations than are algebraically necessary to determine the pressure and flow. In this case Eq. (1) is solved using a least-squares method.

Determining \mathbf{A} from theory does not take into account perturbation of the wave by the transducers or nonidentical transducer responses and also requires an accurate knowledge of the complex wave number k , which depends significantly on temperature, humidity, and surface roughness. For these reasons, one or more calibrations are often used to determine \mathbf{A} .

IV. CALIBRATION OF IMPEDANCE HEADS

A. Review of calibration techniques

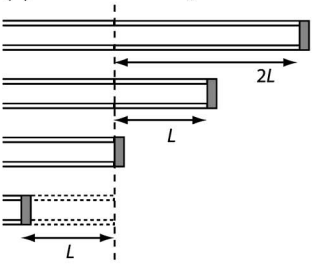
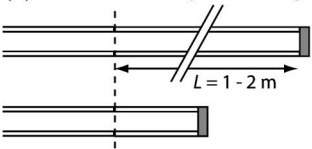
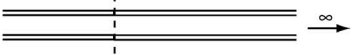

An impedance head can be constructed with two or more transducers that respond linearly to changes in pressure and flow. This includes, for example, microphones with non-negligible compliance and transducers with frequency-dependent gain. These are mounted in a waveguide which need not be cylindrical, and the exact positions of each transducer need not be known. Such a system may be calibrated fully by measuring its response to three test loads of known impedance. For a head with two microphones, three complex parameters must be determined, e.g., the gain ratio of the microphones, the admittance of the microphone closest to the unknown load, and the complex wave number for propagation within the waveguide, although other physical quantities such as the distance to each microphone may be used as parameters. Hence the ratios of microphone signals from three known loads are required to fully calibrate the system. If there are more than two microphones, the calibration parameters are overdetermined with three calibration loads, but two is not enough. If some assumptions can be made about the impedance head, then the number of calibrations required is reduced.

Gibiat and Laloë¹³ describe a complete calibration routine for the two-microphone method [calibration loads (a) in Table II]. They use two stopped pipes of diameter equal to that of the measurement head, and a quasi-infinite impedance (a solid stop at the reference plane) to determine three calibration functions. The lengths of the two stopped pipes must be chosen carefully so that a range of impedances is encountered at each frequency. For example, if a calibration pipe has a resonance at a frequency of interest, its input impedance at that frequency will be very similar to that of the solid stop, and the calibration functions will have large errors at that frequency. For measurements of acoustic impedance over a wide frequency range, several microphone spacings are needed, each with its own set of calibration loads.

The two-microphone-three-calibration (TMTTC) technique described earlier¹³ depends on accurate knowledge of the impedances of the test pipes; this in turn requires accurate knowledge of the complex wave number, a quantity that is strongly dependent on measurement conditions and the surface characteristics of the test pipes. Calibration using resonant pipes explicitly depends on a theory for wall losses. Further, the temperature and humidity must be accurately determined, and the test pipes must be very accurately machined. If an extra calibration is available, then the complex wave number need not be known; this is the approach of van Walstijn *et al.*,¹⁵ who employ the three calibrations of Gibiat and Laloë plus a “negative length” tube, realized by defining the reference plane some distance from the first microphone.

Dalmont²⁵ presents a calibration technique for impedance heads with a volume flow source, which may also be extended to the two-microphone case² [calibration loads (b) in Table II]. The method is based on resonance analysis of a long closed tube. The impedance of the tube is measured using the uncalibrated head and the attenuation coefficient and wave number are derived from the measurement. When

TABLE II. Several techniques for calibration of impedance heads, with selected references and notes.

Calibration Loads	Refs.	Notes
(a) TMTTC technique 	13, 15	<ul style="list-style-type: none"> • ‘complete’ calibration with three known loads • several sets of calibration loads needed to cover wide f-range • complex wavenumber need not be known if a fourth ‘negative length’ load is used
(b) resonance analysis of long tube 	7, 25	<ul style="list-style-type: none"> • after initial measurement of Z for long tube, calibration parameters determined from oscillations in α and k • complex wavenumber need not be known if an extra short tube is used • data obtained only at resonances and antiresonances of long tube—low frequency limit determined by length of tube
(c) semi-infinite pipe 	6	<ul style="list-style-type: none"> • almost purely resistive load—impedance insensitive to complex wavenumber • used for calibration of volume flow sources
(d) resonance-free loads 		<ul style="list-style-type: none"> • complete calibration as in (a) • valid for all frequencies, due to lack of resonances • the flange calibration may be omitted if a model of the impedance head is available

plotted against frequency these will show periodic oscillations with amplitude proportional to any errors in calibration. The three calibration constants are estimated from these oscillations, and the procedure is repeated until the oscillations are tolerably small. The main advantage of this procedure is that it does not depend on exact knowledge of the complex wave number. The main disadvantage is that it only yields calibration data at each resonance frequency of the calibration tube. These may be interpolated, but the lowest frequency that can be measured is limited by the length of the calibration tube (and is around 80 Hz for a 2 m tube—and higher for shorter tubes).

B. General calibration technique using up to three resonant-free loads

Here, we calibrate a three-microphone impedance head in a method similar to that used by Gibiat and Laloë. However, to obviate the need to know the complex wave number precisely, we use three loads without any resonances: a quasi-infinite impedance; an almost purely resistive impedance; and a flange [calibration loads (d) in Table II]. The resistive impedance is in our case a pipe so long that the reflected wave returns reduced in amplitude by 80 dB or more. For lower frequencies, with loss less than 80 dB, it suffices to deliver the signal in pulses of duration T with $T < 2L/c$, where L is the length of the pipe.

The three calibration loads (∞ , Z' , and Z'') are measured and yield the measurement vectors \mathbf{b} , \mathbf{b}' , and \mathbf{b}'' . The measurements are made and the output spectrum optimized (see Sec. VI) using either a theoretical matrix \mathbf{A} derived from Eq. (2) or one derived from a previous calibration on the same measurement head. We then have the three calibration equations

$$p\mathbf{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \mathbf{b}, \quad (4)$$

$$p'\mathbf{A} \begin{bmatrix} 1 \\ 1/\bar{Z}' \end{bmatrix} = \mathbf{b}', \quad (5)$$

$$p''\mathbf{A} \begin{bmatrix} 1 \\ 1/\bar{Z}'' \end{bmatrix} = \mathbf{b}'', \quad (6)$$

where the bar represents a reduced impedance ($\bar{Z} \equiv Z/Z_0$) and p , p' , and p'' are the pressures at the reference plane during measurement of each of the three calibration loads. Dividing each subsequent row in Eq. (4) by the first row yields the first column of \mathbf{A} ,

$$A_{j1} = A_{11}b_j/b_1, \quad (7)$$

in terms of A_{11} . A_{11} can be given any value without affecting impedance measurements; it is usually set equal to

$\cosh(ikx_1)$ (equivalent to assuming the first microphone has unity gain and that the measurement duct is cylindrical).

Taking pairs of rows from Eq. (5), we may eliminate p' and obtain linear equations in the unknowns A_{j2} . For example, rows 1 and 2 combine to give $b'_2(A_{11}+A_{12}/\bar{Z}') = b'_1(A_{21}+A_{22}/\bar{Z}')$. The elements A_{j2} are determined by eliminating the pressure (p' or p'') from each pair of rows in Eqs. (5) and (6) and solving the resulting system. Note that for all $n > 2$ the system is overdetermined (in the algebraic, noise-free sense) and for $n=2$ is algebraically equivalent to the TMTTC technique of Gibiat and Laloë.

So, e.g., for a head with two microphones, calibrated with three known loads,

$$\begin{bmatrix} b'_2/\bar{Z}' & -b'_1/\bar{Z}' \\ b''_2/\bar{Z}'' & -b''_1/\bar{Z}'' \end{bmatrix} \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} = \begin{bmatrix} b'_1A_{21} - b'_2A_{11} \\ b''_1A_{21} - b''_2A_{11} \end{bmatrix}. \quad (8)$$

The above-outlined calibration technique assumes very little about the geometry of the impedance head and the characteristics of the transducers. If the calibration is complete, then wall losses within the impedance head do not need to be taken into account explicitly. In the multiple microphone technique with cylindrical waveguide and microphones attached at known distances from the reference plane, the calibration parameters may be recast in a more instructive form. If each microphone has an admittance of y_j/Z_0 , then the pressure and upstream flow at microphone j are related to those at microphone $j-1$ according to

$$\begin{bmatrix} p_j \\ Z_0 U_j^+ \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -y_j & 1 \end{bmatrix} \mathbf{T} \begin{bmatrix} p_{j-1} \\ Z_0 U_{j-1}^+ \end{bmatrix}, \quad (9a)$$

where \mathbf{T} is the transfer matrix for a cylindrical pipe

$$\mathbf{T} = \begin{bmatrix} \cosh(ikd_j) & \sinh(ikd_j) \\ \sinh(ikd_j) & \cosh(ikd_j) \end{bmatrix} \quad (9b)$$

and $d_j = x_j - x_{j-1}$. For $j=1$, p_{j-1} and U_{j-1}^+ are the pressure and flow at the reference plane and $d_1 = x_1$. The microphone signals are equal to $\kappa_j p_j$, where κ_j is the gain of microphone j . Taking Eq. (9) and a calibrated matrix \mathbf{A} , one can determine k , κ_j for $j=1, \dots, n$ and y_j for $j=1, \dots, n-1$ (y_n , the dimensionless admittance of the microphone closest to the source, cannot be determined). For a measurement setup with a combination of pressure and flow transducers, or all flow transducers, the calibration proceeds in much the same way and an equation similar to Eq. (9) can be constructed.

Sometimes a third calibration is unnecessary or impracticable. In these cases one may precalculate the complex wave number k , using a theory that accounts for viscothermal losses within the waveguide. For a given impedance head, a single set of three calibrations can determine the degree of confidence one may take in this assumption, and the errors involved in making it. The remaining elements of \mathbf{A} are then found from Eqs. (4) and (5) as described.

If one were confident in making further assumptions about the impedance head, then a single calibration may be used to determine either the microphone gains or admittances. For example, one might assume that $y_j=0$ for j

$=1, \dots, n-1$ (a reasonable assumption for small microphones coupled closely to a large waveguide). Then the matrix elements A_{j1} would be found from a measurement of the quasi-infinite impedance load [Eq. (7)] and the elements A_{j2} are given by

$$A_{j2} = A_{j1} \tanh(ikx_j). \quad (10)$$

Otherwise, we may decide to assume that $\kappa_j=1$ for all j and determine the microphone admittances from the quasi-infinite impedance calibration using Eq. (9).

C. Choice of calibration loads

The TMTTC technique works well for small frequency ranges but depends critically on the theory used to account for wall losses. Larger frequency ranges can be covered by using several microphone spacings and calibration tubes. By using the general technique presented earlier, where the signals from two or more microphones are processed simultaneously, a wide frequency range can be covered without measuring piecewise. The three resonant-free calibration loads used here are sufficient to determine the calibration parameters over the entire frequency range, although the impedance for the flange calibration must be derived from theory. For an impedance head with cylindrical waveguide, and nearly ideal microphones at known distances from the reference plane, one or more of the calibration loads may be omitted. Thus calibrating with the quasi-infinite impedance alone may be sufficient for many applications. If another calibration is required and a resistive impedance load is not available, one or more closed tubes of different length may be used instead. Then \bar{Z}' in Eq. (5) is the reduced impedance of the closed tube and several such equations should be solved simultaneously if several closed tubes are used (to choose resonant tube lengths see Gibiat and Laloë¹³).

V. ERRORS

A. Inadequate spectral resolution

Large resonances are usually present in any duct system used to measure acoustic impedance. As discussed by Bodén and Åbom⁹ in the context of the two-microphone method, the pressure spectrum at each microphone varies periodically with frequency, due both to changes in the standing wave pattern in the duct as frequency is varied, and to resonances of the total duct system. When this pressure spectrum is estimated with a frequency resolution Δf , the period of any variation in the spectrum must be large compared to Δf in order to avoid errors associated with resonances in the head and impedance system. For this reason, the microphones should be positioned as close as possible to the impedance to be measured, and the duct length should be kept small. Some acoustic damping between the loudspeaker and the transducers may be used to reduce the amplitude of these duct resonances.

B. Nonlinear transducer responses

Microphone and loudspeaker distortion can both produce errors in a measured impedance. In methods where the

impedance is calculated solely from the signals of two or more transducers obtained simultaneously, a distorting loudspeaker will not affect the measurement, as the distortion is present in each transducer signal and is canceled out. In methods using a single transducer and attenuator, loudspeaker distortion will only affect the measurement if it changes with the load (possible at impedance extrema). Distortion in the transducers will produce errors in the measured impedance. Microphone distortion is always present but is reduced at lower pressures, and the excitation signal may be adjusted to achieve a compromise between random noise and distortion. In this study, the compromise between these two effects was made during calibration on the quasi-infinite impedance by adjusting the output signal level to give microphone signal ratios with minimal contamination from random noise or distortion (as measured by the deviation of the signal from a smooth curve on a small frequency scale).

C. Diameter mismatch at the reference plane

At any bore discontinuity, nonpropagating modes are evoked. Automotive mufflers usually have several such discontinuities. Most woodwind instruments have such discontinuities (for example at tone holes), but provided they are located a sufficient distance from the input, they do not influence the measurement of the plane-wave impedance. If, however, a discontinuity is present at the reference plane, the measured impedance will not be the plane-wave impedance of the object but some combination of the elements of the generalized impedance matrix described by Pagneux *et al.*²⁶

The difference between the measured impedance and the “true” plane-wave impedance may be determined by multimodal theory and expressed as an error term. Alternatively, if an impedance head is calibrated on pipes with entry diameter equal to the entry diameter of the object under study, higher modes evoked at the reference plane are automatically taken into account.¹³ In practice this approach requires many sets of calibration pipes, and it is often easier to apply one of the following corrections.

Van Walstijn *et al.*¹⁵ discuss the effect of a duct discontinuity in the context of the two-microphone technique, but the results are applicable to any system where the transducers are a sufficient distance from the reference plane to measure only plane waves. They derive an expression for a suitable correction term, providing that all higher modes excited at the reference plane are evanescent and do not couple to any higher modes in other parts of the object.

If a volume flow source is used [see method (a) in Table I], the attenuator output and microphone will be located at, or very close to, the reference plane. Brass and Locke²⁷ and Fletcher *et al.*²⁸ have derived suitable correction terms for this situation.

An impedance determined by applying the above-mentioned corrections will usually be less accurate than one measured with a matching impedance head, as turbulent losses are not taken into account in the multimodal model.

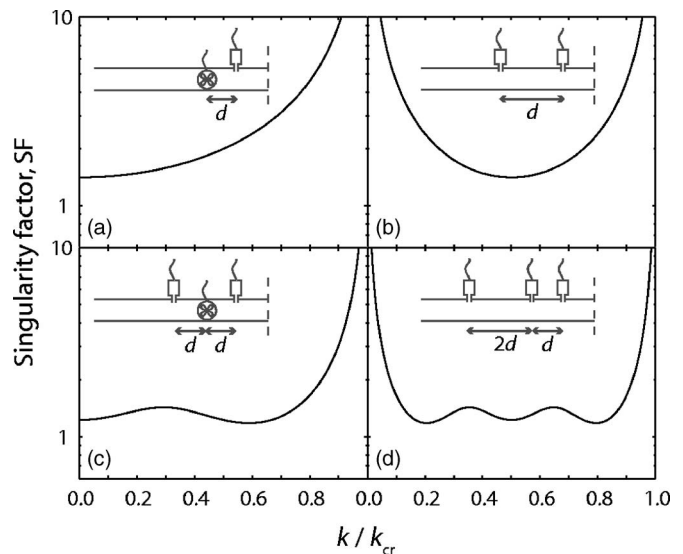


FIG. 2. Singularity factors plotted from $k=0$ to the first singularity for several impedance heads (shown schematically in inset). The heads comprise (a) a microphone and anemometer, (b) two microphones, (c) two microphones and an anemometer, and (d) three microphones. For (a) and (c) $k_{cr}d = \pi/2$ while for (b) and (d) $k_{cr}d = \pi$.

D. Random noise (acoustical or electrical)

Each transducer signal is contaminated by random noise. Whether this is of acoustical or electrical origin is usually unimportant. This noise often has an overall $1/f^n$ dependence, where $0 < n < 1$ and the quality of measurements may be improved by increasing the power in the lower frequencies.

E. The “singularity factor”

The sensitivity of any impedance head to errors in the input quantities varies over frequency. In the two-microphone method, for example, the head becomes “singular” when the microphone spacing is an integral multiple of $\lambda/2$ and in this vicinity large errors in impedance result from small measurement errors. Conversely, for a microphone spacing of $\lambda/4$, the head is least sensitive to errors in the measured quantities. This effect is conveniently represented by the function SF (for singularity factor), defined by Jang and Ih,²³ their Eq. (16), and derived from the singular value decomposition of the matrix \mathbf{A} . (Jang and Ih are primarily interested in reflection coefficients, and their matrix \mathbf{A} relates the incident and reflected wave forms to the measured pressures. Their expression for SF remains valid when our modified matrix \mathbf{A} is used.) The singularity factor is useful to compare different impedance heads according to their sensitivity to measurement errors. SF is plotted for four systems in Fig. 2. Here attenuation is neglected and ideal transducers are assumed. The head in (a) utilizes a flow transducer and a microphone coupled to the duct at different positions. SF for this setup is smallest at low wave number and has a singularity when the distance between the transducers is equal to a quarter wavelength. In order to reduce error, the transducers should be as close as practicable to each other. However they are often separated by some distance, and the effect of this on the error in Z should not be neglected. Also, if the trans-

ducers deviate at all from ideality, they may be considered as ideal transducers separated by a certain effective “distance” (which may be complex and frequency dependent),²⁵ and two such transducers may exhibit behavior as depicted in Fig. 2(a) even if physically located at the same position.

In Figs. 2(b) and 2(d) SF is shown for heads comprising two and three microphones, respectively. These functions are singular at $k=0$ and when the smallest microphone separation is equal to $\lambda/2$. Addition of the third microphone reduces SF over the entire frequency range and widens the range over which the method may be used. In Fig. 2(c) SF is shown for two microphones and a flow transducer; not surprisingly the function is lower than that for either Fig. 2(a) or 2(b) in isolation.

F. Calculating the error in Z

Following Jang and Ih,²³ we may write the measured signals $\tilde{\mathbf{b}}$ as the sum of the (hypothetical) real values, \mathbf{b} , and the measurement errors \mathbf{m} :

$$\tilde{\mathbf{b}} = \mathbf{b} + \mathbf{m}. \quad (11)$$

The errors in \mathbf{x} may then be written in terms of the measurement errors:

$$\tilde{\mathbf{x}} - \mathbf{x} = \mathbf{A}^+ \mathbf{m}, \quad (12)$$

where \mathbf{A}^+ is the generalized (Moore-Penrose) inverse of \mathbf{A} .

If we right-multiply the error in \mathbf{x} by its complex conjugate transpose (denoted here by a superscript H), the expectation value of the resulting matrix is given by

$$\begin{aligned} E[(\tilde{\mathbf{x}} - \mathbf{x})(\tilde{\mathbf{x}} - \mathbf{x})^H] &= E[(\mathbf{A}^+ \mathbf{m})(\mathbf{A}^+ \mathbf{m})^H] \\ &= E[\mathbf{A}^+ \mathbf{m} \mathbf{m}^H (\mathbf{A}^+)^H] = \mathbf{A}^+ \mathbf{V} (\mathbf{A}^+)^H, \end{aligned} \quad (13)$$

where $\mathbf{V} = E[\mathbf{m} \mathbf{m}^H]$ is the covariance matrix. If each transducer signal is contaminated by uncorrelated noise of variance σ_j^2 , where σ_j can in general be a function of frequency, then \mathbf{V} is a diagonal matrix with the σ_j^2 s on the diagonal. The errors in p and U , Δp and ΔU are given by the diagonal elements of the matrix in Eq. (13):

$$|\Delta p|^2 = [\mathbf{A}^+ \mathbf{V} (\mathbf{A}^+)^H]_{11}, \quad (14)$$

$$|Z_0 \Delta U|^2 = [\mathbf{A}^+ \mathbf{V} (\mathbf{A}^+)^H]_{22} \quad (15)$$

and the error ΔZ in the impedance Z is obtained by propagation of errors:

$$\frac{\Delta Z}{Z} = \sqrt{\left| \frac{\Delta p}{p} \right|^2 + \left| \frac{\Delta U}{U} \right|^2}. \quad (16)$$

The Pythagorean sum used in Eq. (16) is only strictly correct if the errors in p and U are independent. While this is not true in general, the equation will be approximately correct at impedance maxima and minima, where either $|\Delta p/p| \gg |\Delta U/U|$ or $|\Delta U/U| \gg |\Delta p/p|$. At intermediate values of impedance, where $|\Delta p/p| \approx |\Delta U/U|$, Eq. (16) will overestimate the error. Since in this work we are mostly interested in impedance maxima and minima, use of Eq. (16) to estimate

the error will lead only to a more conservative redistribution of power.

The variance of the measurement errors may be estimated from experiment or simply assumed equal for each microphone, with possibly an $f^{-0.5}$ dependence if the noise is of mostly acoustical origin. In this study the sound source was excited with repeated cycles of a periodic signal (see Sec. VI) and the microphone signals were synchronously acquired in blocks corresponding to one cycle of the excitation signal. Spectra were computed for each block and averaged to reduce noise. The standard deviation in these spectra was used to fit σ_j for each microphone as an exponential function of frequency. The σ_j s thus obtained were used to calculate the covariance matrix and the error in Z [Eqs. (14)–(16)].

VI. OPTIMIZATION OF THE OUTPUT SIGNAL

To measure an impedance spectrum, an output signal covering all frequencies in the range is required. Some authors use a swept-sinusoid² as the output signal. However, this takes much longer than using a signal with all of the frequency components present, such as white noise or chirps. In the present work, a signal is generated as a sum of components of all sampled frequencies. This signal is applied to the loudspeaker and the impedance and error are calculated.

Initially, the wave is synthesized numerically from components of equal amplitude. To improve the signal-to-noise ratio, the relative phases are adjusted so as to reduce the ratio of the maximum of the sum of sinusoids to the amplitude of each sinusoid, as described by Smith.²⁹ A wave with a flat power spectrum at the computer does not result in the acoustical wave produced at the reference plane having a flat spectrum, however, because the amplifiers, loudspeaker, and connecting conduit all have frequency-dependent responses. These responses could be removed by calculating the power function and multiplying the output spectrum by its inverse (such an approach is used by Wolfe *et al.*⁶). A flat acoustical spectrum does not, however, produce a flat noise response, because a given head has greater sensitivity to noise at some frequencies than at others. This can be compensated for by multiplying the output spectrum by the correction factor

$$C_1(f) = \frac{\Delta Z}{Z^w} \quad (17)$$

(where w is a weighting factor that may be varied to give preference to impedance maxima or minima) and using the resulting wave form in a second impedance measurement. If there are significant nonlinearities in the loudspeaker system, this procedure may be repeated until $C_1(f)$ is tolerably flat, but this is usually unnecessary.

VII. MATERIALS AND METHODS

A. The impedance spectrometer

For the experiments described in this paper the impedance spectrometer is configured as shown in Fig. 3. The signal is synthesized on a computer (Macintosh G4) and output via a nominal 24 bit DAC (MOTU 828) to a power amplifier and midrange speaker. A truncated cone helps match the

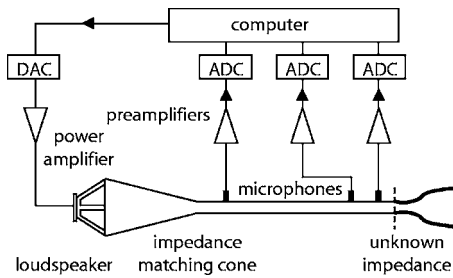


FIG. 3. The experimental setup. Calibration used two or three resonant-free calibration loads [see calibration loads (d) in Table II].

speaker to the measurement head. The impedance spectra are measured between 120 Hz and 4 kHz (a range that encompasses the fundamental frequency of all of the notes of many wind instruments, such as the clarinet and flute, and includes their cutoff frequency). It is easy to extend the frequency range of this technique to other instruments, e.g., the didjeridu.³⁰

A brass measurement head was used with an inner diameter of 15.0 mm and with a 6 mm wall thickness. The brass construction and relatively massive design serve to lessen mechanical conduction of sound and reduce any temperature fluctuations as the head is handled during an experiment.

Three 1/4-in. condenser microphones (Brüel & Kjær 4944 A) are mounted in the impedance head, perpendicular to the cylindrical axis. A 1 mm hole couples each microphone to the waveguide. The compliance of these microphones (equivalent to an air volume of 0.25 mm³ at 250 Hz) is negligible compared to that of the volume of air between the microphone and coupling hole, which together with the mass of air in the coupling hole forms a Helmholtz resonator with a natural frequency of 6.8 kHz. The coupling hole should be small enough so that the pressure wave is sampled over a distance small compared to its wavelength, and (for measurements using fewer than three calibrations) so that the impedance head is cylindrical, to a good approximation. On the other hand, it should be large enough so that the Helmholtz frequency of the microphone and coupling is much higher than the highest measured frequency. The chosen size of 1 mm represents a compromise between these competing considerations. The three microphones are mounted at 10, 50, and 250 mm from the reference plane. With the microphones positioned thus, a singularity occurs at $\lambda=80$ mm, which for $c=345$ m s⁻¹ corresponds to a frequency of 4.3 kHz (outside the frequency range of interest).

The signals from each microphone are preamplified and adjusted for calibrated gain by a Brüel & Kjær Nexus conditioning amplifier (2693-0S4) and digitized and recorded by the MOTU interface and the AUDIODESK software package. Wave forms are sampled at 44.1 kHz throughout and the output wave form is synthesized at 2¹⁴ points (giving a frequency resolution of 2.7 Hz). To improve the signal-to-noise ratio the output is cycled repeatedly (100 cycles are typical, resulting in a total measurement time of 37 s) and the recorded signals are averaged. Fourier transforms are performed on the averaged data using the built-in functions in MATLAB.

B. Calibration loads

The calibration loads (d) in Table II were used—a quasi-infinite impedance (brass plate), an almost purely resistive impedance (very long pipe), and a quasi-infinite flange.

As pointed out by Dalmont,²⁵ in most applications it is sufficient to assume that the admittance of a rigid wall is equal to zero (infinite impedance). Dalmont gives $Z_0 Y_{\text{rigid}} = 9.6 \times 10^{-6} \sqrt{f}(1+i)$ for the reduced admittance of a rigid wall at 20 °C. For a tube of radius 10 mm at 100 Hz, the imaginary part of this admittance corresponds to a length correction of 0.05 mm and the real part to viscothermal dissipation on a length of 3 mm.

The almost purely resistive impedance is a straight PVC pipe of length 97 m and 15 mm internal diameter. The pipe is capped at its far end and filled with a small length (approximately 100 mm) of acoustically absorbing wool. If one assumes a fully reflective termination, i.e., neglecting the low reflection termination, at 120 Hz the reflected wave returns with a loss of at least 76 dB.²⁴ The actual loss at 120 Hz will be greater than this lower-bound due to absorption by the acoustic wool. The loss will also increase at higher frequencies due to viscothermal effects. Thus if there were equal power at each of the approximately 1400 frequencies, each reflected component would lie below the effective resolution of the ADC (~105 dB).

The quasi-infinite flange is a square perspex plate of side 600 mm in the center of which is a hole for mounting on the measurement head (for the end effect of a square flange see Dalmont *et al.*³¹). Over the frequency range of interest the impedance of a flange is lower than that of the resistive impedance load and so these three loads give complementary information. Preliminary experiments have shown that the inclusion of the third load changes the calibration very little, thereby justifying the assumptions made about propagation in the waveguide, and so it was omitted for most measurements.

VIII. RESULTS AND DISCUSSION

A. Effect of optimizing the output signal

As discussed earlier, equal distribution of energy among frequencies does not result in equal distribution of errors in the measurements, and the size of the largest error can be substantially reduced by adjusting the output spectrum for particular circumstances and loads.

Figure 4 shows the magnitude of and (absolute) fractional error in the measured impedance for three sequential impedance measurements. The error was calculated using Eqs. (14)–(16). In Fig. 4(a), the impedance Z of the resistive impedance load is measured with a nonideal source. The measured impedance spectrum is not completely flat since this measurement was performed before calibration. The error function $\Delta Z/Z$ in Fig. 4(a) has broad features corresponding primarily to loudspeaker and conduit resonances. The output spectrum is multiplied by the error function in Fig. 4(a) and used to measure an open pipe [Fig. 4(b)]. The output spectrum is again multiplied by the error function and the pipe is measured a second time. The error in the resulting measurement [Fig. 4(c)] is uniformly distributed over fre-

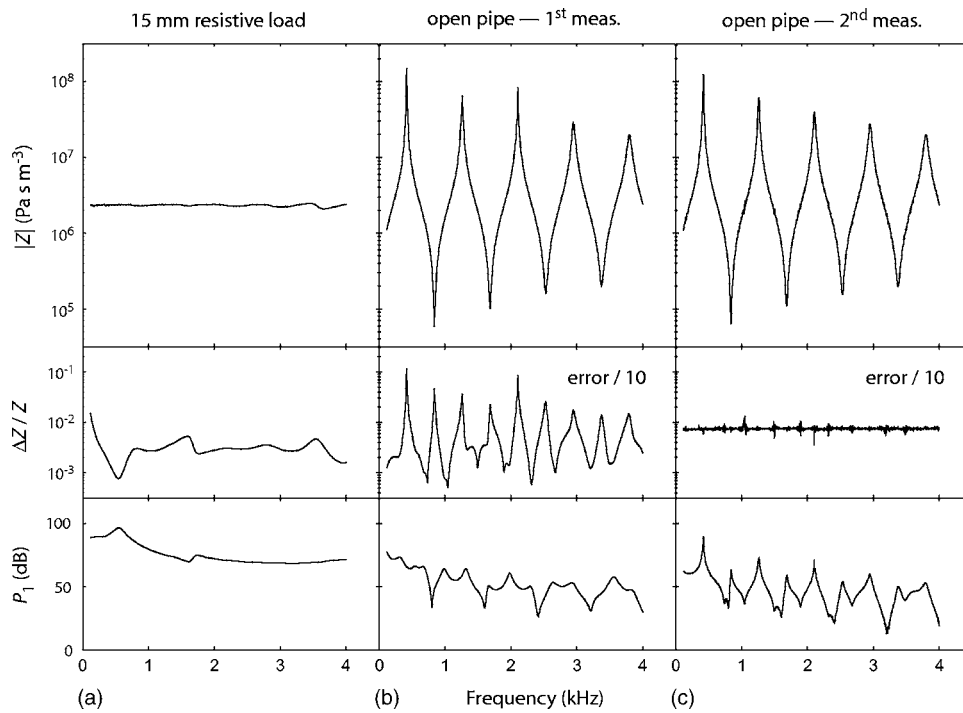


FIG. 4. Optimizing the output signal for measurements on 15 mm pipes. Impedance magnitude (upper panel) and fractional error (middle panel) are shown for (a) the 15 mm resistive impedance load measured with a non-ideal source and for an open pipe of length 200 mm measured with a deliberately very low signal (b) before and (c) after optimization. The error spectra in (b) and (c) have been scaled for comparison with (a). Also shown for each measurement is the sound pressure in dB re 20 μ Pa (lower panel) as measured by the microphone nearest to the reference plane.

quency. The errors shown in Figs. 4(b) and 4(c) have been divided by a factor of 10 (the signal amplitude was deliberately kept smaller than optimum to increase the apparent errors).

It is worth noting that in the uncorrected measurement (b) very high or very low impedances are measured with a greater error than are impedances close to Z_0 . This is unfortunate in the case of musical instruments where we are particularly interested in impedance maxima and/or minima. On subsequent iterations, more power is put into these frequencies and the error is thereby reduced. As can be seen in Fig. 4(c), the error increases for impedances close to Z_0 (the error is also more apparent here since the slope of the curve is not as steep as at impedance extrema). However, the maximum error over the entire spectrum is reduced by a factor of approximately 10 through this optimization procedure.

Also shown in Fig. 4 is the sound pressure spectrum measured by the microphone closest to the reference plane for each measurement. This is relatively uniform over frequency for the measurement of the resistive impedance and has minima corresponding to nodes of the standing wave for measurements of the 200 mm closed pipe.

The choice of the exponent w in $C_1(f)$ [Eq. (17)] determines which part of the spectrum will be measured with the greatest precision. For $w=1$, the fractional error will be constant for all impedances. For $w>1$, the impedance minima will be determined with greater precision, and for $w<1$ the impedance maxima will be given preference. The case where $w>1$ is particularly useful for measuring the impedance of instruments in the flute family, which play near impedance minima, whereas $w<1$ is useful for reed and lip valve instruments such as the clarinet which play near impedance maxima.

In some cases (such as when measuring calibration loads) it is not desirable to use the function $C_1(f)$ to modify

the output spectrum, but we may still like to compensate for the system responses and the singularity factor of the head. In these cases, we modify the output spectrum to ensure that the acoustic energy density at the reference plane is the same as it would be during a (hypothetical) measurement of Z_0 with $\Delta Z/Z=K$ where K is independent of frequency. The acoustic energy density at the reference plane during a measurement of an impedance Z is proportional to $\epsilon=(|p|^2+|Z_0U|^2)/2$. For a measurement of Z_0 ,

$$\epsilon_0=|Z_0U|^2=\frac{|\Delta p|^2+|Z_0\Delta U|^2}{K^2}, \quad (18)$$

where Eq. (16) was used for K with the substitution $p=Z_0U$. The correction factor $C_2(f)$ used to modify the output spectrum will be proportional to the square root of the energy ratio ϵ_0/ϵ . We use

$$C_2(f)=K\sqrt{\frac{\epsilon_0}{\epsilon}}=\sqrt{2}\sqrt{\frac{|\Delta p|^2+|Z_0\Delta U|^2}{|p|^2+|Z_0U|^2}}, \quad (19)$$

where the factor K has no effect on the output wave form (being independent of frequency) but is used to ensure that $C_2(f)=\Delta Z/Z$ when $Z=Z_0$.

B. Effect of calibration

Figure 5 illustrates the effects of calibrating with varying numbers of known calibration loads. Measured impedance spectra are shown for a closed 15 mm pipe, 200 mm long. To simulate the effect of using unmatched microphones, in Figs. 5(a) and 5(b) the signals from the second and third microphones were scaled by 1.2 and 0.8, respectively. The measured impedance spectrum before calibration (i.e., assuming unity gain for each microphone) is shown in Fig. 5(a), while in Fig. 5(b) the impedance was calculated

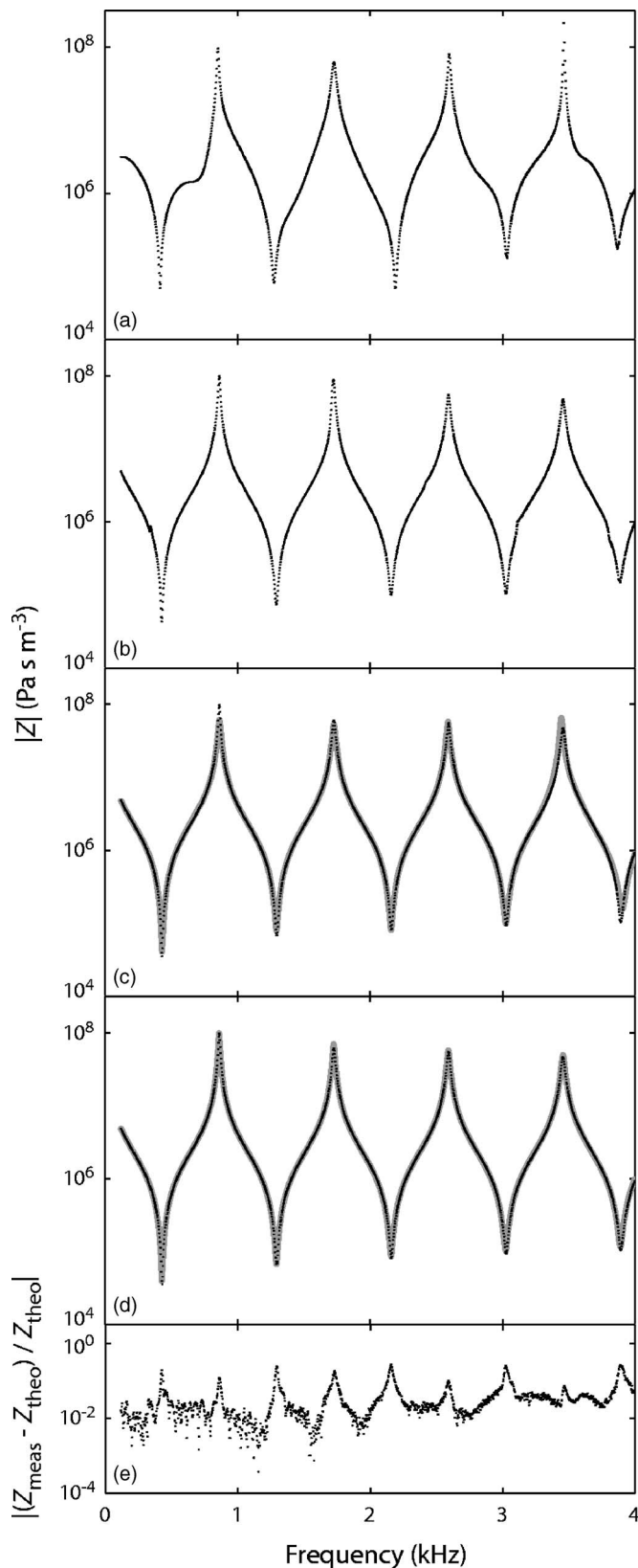


FIG. 5. The input impedance for a closed 15 mm pipe, 200 mm long. In (a) and (b) the signals from microphones 2 and 3 were multiplied by 1.2 and 0.8, respectively, to deliberately introduce errors. The impedance was measured before calibration (a) and after calibrating using one known load (b). In (c) the impedance is given for matched microphones before (grey, solid line) and after (black, dotted line) calibration with two known loads, and in (d) the calibrated measurement from (c) (black, dotted line) is compared with the theoretical impedance for an ideal tube (grey, solid line). The fractional difference between this theory and the measurement is shown in (e).

after calibrating on the quasi-infinite impedance load [using Eqs. (7) and (10)]. The precalibration measurement deviates significantly from the theoretical impedance, while calibrating with only one load (b) improves the accuracy significantly. In Fig. 5(c) the measured impedance spectrum using the matched microphone signals is shown before and after calibration with two known loads. While the two spectra are similar in many respects, the size and frequency of extrema (particularly maxima) are significantly different. For this measurement, adding the flange calibration made very little difference. The measurement made without using the two calibration loads shows similar features to those of the precise measurement over most of the range, but has noticeable errors in the magnitude of extrema (particularly maxima) and smaller errors in the frequencies at which they occur. In Fig. 5(d) the measurement on the 200 mm closed pipe using two calibration loads is compared with the theoretical impedance and the fractional difference between these two measurements is shown in Fig. 5(e). The peaks in this difference function at impedance extrema are likely due to slightly greater attenuation in the measured impedance than is predicted by the theory for viscothermal losses at walls that are ideally smooth.

C. Conclusions and practical considerations for measurement

The combination of three microphones, three nonresonant calibrations, and spectral shaping can give precise measurements over a wide range of frequencies and impedance. It has the added advantage that no assumptions need be made about the microphone characteristics or about the exact geometry of the impedance head. Furthermore, there is no need to invoke a theoretical model for waveguide losses.

In many practical measurement situations, however, the reduced performance of a simpler combination might still be appropriate.

Thus two microphones, rather than three, would often be sufficient for a smaller frequency range (perhaps 2–3 octaves, depending upon the precision required).

In many practical cases, fewer than three nonresonant calibrations might be sufficient, albeit with some loss in performance. Thus if a well-defined cylindrical impedance head is used, and if the perturbations of the cylinder by the microphones are sufficiently small, it is possible to remove one calibration (this is a consequence of the good theoretical model available for a cylindrical waveguide). It is also possible to remove one calibration if the characteristics of the microphones are already known with a high degree of precision.

The redistribution of power in the source function (spectral shaping) can improve the signal-to-noise ratio at extrema by a factor of 10 or more. Whether or not this feature is needed depends on the size of errors that may be tolerated in the measurements. Further, this feature would be less important if the unknown impedance has no strong resonances.

ACKNOWLEDGMENTS

Technical assistance and advice was provided by John Tann. P.D. was supported by an Australian Postgraduate Award (Industry) jointly funded by the Australian Research Council, the Powerhouse Museum, and Terry McGee (flute-maker).

- ¹A. H. Benade and M. I. Ibsi, "Survey of impedance methods and a new piezo-disk-driven impedance head for air columns," *J. Acoust. Soc. Am.* **81**, 1152–1167 (1987).
- ²J.-P. Dalmont, "Acoustic impedance measurement. I. A review," *J. Sound Vib.* **243**, 427–439 (2001).
- ³A. P. Watson and J. M. Bowsher, "Impulse measurements on brass musical instruments," *Acustica* **66**, 170–174 (1988).
- ⁴J. J. Fredberg, M. E. B. Wohl, G. M. Glass, and H. L. Dorkin, "Airway area by acoustic reflections measured at the mouth," *J. Appl. Physiol.: Respir., Environ. Exercise Physiol.* **48**, 749–758 (1980).
- ⁵R. Singh and M. Schary, "Acoustic impedance measurement using sine sweep excitation and known volume velocity technique," *J. Acoust. Soc. Am.* **64**, 995–1003 (1978).
- ⁶J. Wolfe, J. Smith, J. Tann, and N. H. Fletcher, "Acoustic impedance spectra of classical and modern flutes," *J. Sound Vib.* **243**, 127–144 (2001).
- ⁷A. M. Bruneau, "An acoustic impedance sensor with two reciprocal transducers," *J. Acoust. Soc. Am.* **81**, 1168–1178 (1987).
- ⁸M. Åbom and H. Bodén, "Error analysis of two-microphone measurements in ducts with flow," *J. Acoust. Soc. Am.* **83**, 2429–2438 (1988).
- ⁹H. Bodén and M. Åbom, "Influence of errors on the two-microphone method for measuring acoustic properties in ducts," *J. Acoust. Soc. Am.* **79**, 541–549 (1986).
- ¹⁰J. Y. Chung and D. A. Blaser, "Transfer function method of measuring acoustic intensity in a duct system with flow," *J. Acoust. Soc. Am.* **68**, 1570–1577 (1980).
- ¹¹J. 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).
- ¹²J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. II. Experiment," *J. Acoust. Soc. Am.* **68**, 914–921 (1980).
- ¹³V. Gibiat and F. Laloë, "Acoustical impedance measurements by the two-microphone-three-calibration (TMTC) method," *J. Acoust. Soc. Am.* **88**, 2533–2545 (1990).
- ¹⁴A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," *J. Acoust. Soc. Am.* **61**, 1362–1370 (1977).
- ¹⁵M. van Walstijn, M. Campbell, J. Kemp, and D. Sharp, "Wideband measurement of the acoustic impedance of tubular objects," *Acust. Acta Acust.* **91**, 590–604 (2005).
- ¹⁶J. Backus, "Input impedance curves for the reed woodwind instruments," *J. Acoust. Soc. Am.* **56**, 1266–1279 (1974).
- ¹⁷J. Backus, "Input impedance curves for the brass instruments," *J. Acoust. Soc. Am.* **60**, 470–480 (1976).
- ¹⁸R. Caussé, J. Kergomard, and X. Lurton, "Input impedance of brass musical instruments—Comparison between experiment and numerical models," *J. Acoust. Soc. Am.* **75**, 241–254 (1984).
- ¹⁹F. J. M. van der Eerden, H.-E. de Bree, and H. Tijdeman, "Experiments with a new acoustic particle velocity sensor in an impedance tube," *Sens. Actuators, A* **A69**, 126–133 (1998).
- ²⁰S. Elliott, J. Bowsher, and P. Watkinson, "Input and transfer response of brass wind instruments," *J. Acoust. Soc. Am.* **72**, 1747–1760 (1982).
- ²¹M. Kob and C. Neuschaefer-Rube, "A method for measurement of the vocal tract impedance at the mouth," *Med. Eng. Phys.* **24**, 467–471 (2002).
- ²²R. L. Pratt, S. J. Elliott, and J. M. Bowsher, "The measurement of the acoustic impedance of brass instruments," *Acustica* **38**, 236–246 (1977).
- ²³S.-H. Jang and J.-G. Ih, "On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow," *J. Acoust. Soc. Am.* **103**, 1520–1526 (1998).
- ²⁴N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998).
- ²⁵J.-P. Dalmont, "Acoustic impedance measurement. II. A new calibration method," *J. Sound Vib.* **243**, 441–459 (2001).
- ²⁶V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. I. Theory and validation," *J. Acoust. Soc. Am.* **100**, 2034–2048 (1996).
- ²⁷D. Brass and A. Locke, "The effect of the evanescent wave upon acoustic measurements in the human ear canal," *J. Acoust. Soc. Am.* **101**, 2164–2175 (1997).
- ²⁸N. H. Fletcher, J. Smith, A. Z. Tarnopolsky, and J. Wolfe, "Acoustic impedance measurements—Correction for probe geometry mismatch," *J. Acoust. Soc. Am.* **117**, 2889–2895 (2005).
- ²⁹J. R. Smith, "Phasing of harmonic components to optimize measured signal-to-noise ratios of transfer functions," *Meas. Sci. Technol.* **6**, 1343–1348 (1995).
- ³⁰J. Smith, G. Rey, P. Dickens, N. Fletcher, L. Hollenberg, and J. Wolfe, "Vocal tract resonances and the sound of the Australian didgeridu (yidaki). III. Determinants of playing quality," *J. Acoust. Soc. Am.* **121**, 547–558 (2007).
- ³¹J.-P. Dalmont, C. J. Nederveen, and N. Joly, "Radiation impedance of tubes with different flanges: Numerical and experimental investigations," *J. Sound Vib.* **244**, 505–534 (2001).

Calibration sphere for low-frequency parametric sonars

Kenneth G. Foote^{a)}

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

David T. I. Francis and Philip R. Atkins

Department of Electronic, Electrical and Computer Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

(Received 11 February 2006; revised 16 October 2006; accepted 17 December 2006)

The problem of calibrating parametric sonar systems at low difference frequencies used in backscattering applications is addressed. A particular parametric sonar is considered: the Simrad TOPAS PS18 Parametric Sub-bottom Profiler. This generates difference-frequency signals in the band 0.5–6 kHz. A standard target is specified according to optimization conditions based on maximizing the target strength consistent with the target strength being independent of orientation and the target being physically manageable. The second condition is expressed as the target having an immersion weight less than 200 N. The result is a 280-mm-diam sphere of aluminum. Its target strength varies from -43.4 dB at 0.5 kHz to -20.2 dB at 6 kHz. Maximum excursions in target strength over the frequency band due to uncertainty in material properties of the sphere are of order ± 0.1 dB. Maximum excursions in target strength due to variations in mass density and sound speed of the immersion medium are larger, but can be eliminated by attention to the hydrographic conditions. The results are also applicable to the standard-target calibration of conventional sonars operating at low-kilohertz frequencies. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2434244]

PACS number(s): 43.58.Vb, 43.30.Xm, 43.25.Zx, 43.30.Pc [AJZ]

Pages: 1482–1490

I. INTRODUCTION

Parametric sonars are fundamentally different from conventional sonars, whose transmit signals propagate linearly, without interacting. At sufficiently high intensities, transmit signals propagating collinearly interact through the intrinsic nonlinearity of the medium, generating waves at the sum and difference frequencies.^{1,2} If the transmit signals consist of two waves at relatively high and similar primary frequencies, then the difference frequency will be relatively low, hence capable of propagating to relatively large ranges. Remarkably, the same difference-frequency wave possesses a directionality resembling that of the beams at the primary frequencies, but without sidelobes. In addition, modest changes in the signal frequency during transmission can generate a broad bandwidth at the difference frequency.

These properties of exceptional directionality and bandwidth at low frequencies have been exploited in a wide range of applications. One class of applications exploits the parametric sonar as a directional and/or broadband sound source in transmission over a one-way path. These include applications to shallow-water communications,³ and fish swimbladder resonance,^{4,5} among others. Another, larger class of applications exploits the properties of parametric sonar in backscattering, hence over a two-way path. These include acoustic scattering by the sea surface,^{6,7} water column,⁸ bottom as in the determination of geoacoustic properties^{6,9–11} and seafloor characterization,^{12–14} and seabed vegetation.¹⁵ Some other backscattering applications include sub-bottom

profiling,¹⁶ detection of objects on and in the seabed,^{17–21} and marine archaeology.^{22,23} The parametric sonar has also been configured as a sidescan sonar for bottom and sub-bottom surveying.²⁴ Rather recently it has been combined with synthetic aperture processing for detection, imaging, and classification of buried waste such as dumped ordinance in the Baltic Sea,²⁵ and buried mines.²⁶

Calibration is useful for most of these applications, but is seldom mentioned, much less explicitly addressed. In some of the very few cases where it is addressed, measurements have been made of the difference-frequency source level,^{4,5} and range and/or angular dependence of the transmit field.^{11,27} These measurements have been made by a hydrophone, which itself requires calibration. At other ranges and angles where knowledge of the difference-frequency field is required, models may be exercised, for example, those of Moffett and Mellon²⁸ or by means of the Bergen Code.²⁹

In backscattering applications, the receiver is integral to sonar performance. Given the number of quantitative backscattering applications, as well as interests in detecting and imaging objects on and beneath the seafloor, it may be wondered whether another calibration method might determine the overall response of the transmit and receive functions of the parametric sonar at low difference frequencies. In particular, might the standard-target method, which has a solid grounding in theory and in practice with high-frequency scientific echo sounders and multibeam sonars, be extended to difference-frequencies in the low-kilohertz range? If so, this would enable parametric and other low-frequency sonars to quantify scattering effects due to excitation of the powerful breathing-mode resonance of fish swimbladders^{30–34} in both the forward^{4,5} and other directions.

^{a)}Electronic mail: kfoote@whoi.edu

Application of the standard-target method would enable the sonar to be calibrated as an integral system in which the combined transmit and receive response function is measured against a primary standard, with typical uncertainty less than ± 0.1 dB over a broad frequency range, from low-kilohertz to ultrasonic frequencies.^{35,36} In contrast, in hydrophone-based methods the transmit and receive parts of the sonar are measured separately, and with less accuracy because of their reference to other hydrophones or standards, whose uncertainty even at national laboratories is of order ± 0.5 dB.³⁷ When the overall system response function is derived from such piecewise measurements, errors also compound. Over time, the acoustic properties of robust standard targets retain their basic values; their stability over any measurement period is ensured. Hydrophones may perform stably over a period of time, but their stability over a period of a single measurement is of paramount importance.^{37,38} Environmental sensitivity may be an issue for both standard targets and hydrophones, but compensation is routine for standard targets; it is of more significant concern for hydrophones.

In the following, the standard-target method is briefly reviewed for high-frequency sonars in both wideband and narrowband modes. Issues peculiar to parametric sonar are identified. An optimization algorithm is then defined for specifying a robust spherical target to serve as the standard target at the difference frequencies of a particular parametric sonar, with bandwidth 0.5–6 kHz. Results of preliminary computations of target strength³⁵ are elaborated for a single calibration target. Effects of uncertainty in knowledge of the material properties and effects due to variations in temperature and salinity of the immersion medium are described in detail. Both the magnitude and stability of target strength of the standard target are discussed.

II. STANDARD-TARGET METHOD

A. Conventional sonar

For conventional sonars operating linearly, the standard-target method of calibration typically involves measurement of a chosen target at a known position in the transducer beam.^{39,40} This is usually on the axis and at a relatively accessible range in the transducer farfield, hence where the field quantities change inversely with range. At other ranges than that of the calibration measurement, the effect of receiver processing can be determined by a purely electronic measurement, as with a device that plays a signal of known magnitude and time delay into the transducer leads.⁴¹

One aim of such a calibration measurement is determination of the overall response H of the transmit-receive system as a function of frequency ν . This relates the received echo spectrum S_R to the transmit signal spectrum S_T by means of the backscattering form function F of the standard target:

$$S_R = S_T F H P, \quad (1)$$

where P expresses the dependence of the echo on the propagation path, including range and absorption if not fully com-

pensated in the receiver, i.e., the two-way acoustic-path loss.⁴²

In applications to narrowband sonar or wideband sonar in which the output is a single measure of echo strength, the constituent quantities in Eq. (1) are typically squared and integrated. The applicable operational measure of backscattering cross section of the standard target in this case is⁴³

$$\sigma_1 = 4\pi \int |SFH|^2 d\nu / \int |SH|^2 d\nu, \quad (2)$$

where the integration is performed over the entire frequency domain. When the sonar is used in a wideband mode, with the intent of determining the frequency dependence of the echo strength, the idealized single-frequency backscattering cross section of the standard target is

$$\sigma_2 = 4\pi |F|^2. \quad (3)$$

This is also the limiting form of Eq. (2) when the transmit signal at ν_0 is essentially monochromatic, with $S \propto \delta(\nu - \nu_0)$, or the receiver frequency response function centered at ν_0 is ideally sharp, with $H \propto \delta(\nu - \nu_0)$, where δ is the Dirac delta function.

The standard-target method of calibrating sonars is widely used for narrowband scientific echo sounders operating at ultrasonic frequencies, e.g., over the range 18–710 kHz.^{44,45} It is also used for wideband scientific echo sounders, e.g., the Broadband Acoustic Scattering Signatures System, with seven octave-bandwidth transducers spanning the total frequency range 25 kHz–3.2 MHz.^{46,47} More recently, protocols for calibrating high-frequency multibeam sonars by the standard-target method have been worked out and reduced to practice for multibeam sonars operating at 90, 200, and 240 kHz.⁴⁸

Standard targets designed and used for these systems have included precision spheres made of copper and tungsten carbide with 6% cobalt binder. Diameters of copper spheres vary from 64 mm for use at 18 kHz⁴⁵ to 60 mm at 38 kHz⁴³ to 23 mm at 120 kHz. Diameters of tungsten carbide spheres have varied from 38.1 mm for use at 38, 120, and 200 kHz,^{36,44} to 10 mm at low megahertz frequencies.⁴⁹ Miyano-hana *et al.*⁵⁰ have advised against using high-carbon steel spheres and have specified eight tungsten carbide spheres with diameters from 35 to 54.1 mm to span the frequency range 10–200 kHz. Corresponding target strengths of the various targets are in the approximate range from –50 to –30 dB.

B. Parametric sonar

The nonlinear interaction of two collinear waves that gives rise to the difference-frequency wave occurs in the immersion medium. This is significant for the particular realization of the standard-target calibration method. Nonlinear interactions also occur in the target material, but these are entirely negligible in the particular case, as argued in the Appendix.

To discuss details of the standard-target calibration method for parametric sonars, it is useful to distinguish the operating regimes of the array. As summarized by Moffett

and Konrad,² there are three ordinary regimes. These are differentiated by two quantities: (i) the scaled source level,

$$SL_{0,sc} = SL_0 + 20 \log \nu_0, \quad (4)$$

where SL_0 is the source level of a primary frequency and ν_0 is the mean primary frequency, and (ii) the absorption. This second quantity is expressed as the product of the absorption coefficient α_0 at ν_0 and the Rayleigh length

$$R_0 = A/\lambda, \quad (5)$$

where A is the area of a primary-frequency projector and λ is the wavelength at the primary frequency. Since R_0 is a measure of the collimation length of the transmitter, $\alpha_0 R_0$ measures the absorption loss in the primary-frequency nearfield.

The three regimes are the following. (i) When $SL_{0,sc} > 280$ dB re 1 μ Pa at 1 m and 1 kHz, harmonic distortion is significant, and the difference-frequency wave is fully formed in the array nearfield, within the range R_0 . (ii) When $SL_{0,sc} < 280$ dB re 1 μ Pa at 1 m and 1 kHz, and $\alpha_0 R_0 > 1$ dB, signal levels are relatively small and most of the absorption, hence difference-frequency generation, occurs within the range R_0 . (iii) When $SL_{0,sc} < 280$ dB re 1 μ Pa at 1 m and 1 kHz and $\alpha_0 R_0 \ll 1$ dB, most of the primary-frequency absorption, hence difference-frequency generation, occurs in the array farfield.

In the first two cases, the array length is effectively limited. If a standard target is placed beyond the effective array length R_0 , then calibration may proceed as with a conventional sonar, for the difference-frequency field is formed and varies inversely with range r for $r > R_0$. This presumes that the particular parametric sonar is to be used for ordinary measurements outside of the interaction zone, which is a reasonable assumption for the two cases.

In the third case, the array length is generally so extended that measurements must be made within the interaction zone. In this general case, which also subsumes the first two cases when measurements are to be performed in the interaction zone, it is necessary to find a means of extrapolating to other ranges. This may be done at least partly through additional calibration measurements, but ultimately requires a computational model, for example, one established by Moffett and Mellen,²⁸ among others, with inclusion of finite-aperture and spherical-spreading effects,^{51,52} as in the Bergen Code.²⁹ Effects due to dispersion, as in the presence of air bubbles,⁵³ may have to be included.

A key issue in the standard-target calibration of parametric sonars at low difference frequencies is finding an appropriate target. If the difference frequency is relatively high, say of order 10 kHz or higher, then there are precedents. If the difference frequency is relatively low, say of order 1–10 kHz, then finding a suitable target presents challenges. These are addressed in the following after describing the particular parametric sonar.

III. SPECIFICATION OF A PARAMETRIC SONAR

For definiteness, an operational parametric sonar with low difference frequencies is chosen. This is the Simrad topographic parametric sonar TOPAS PS18 Parametric Sub-

bottom Profiler.¹⁷ It is based on two primary frequencies in the range 15–21 kHz. Difference frequencies are generated over the band 0.5–6 kHz by each of several excitations, e.g., a continuous wave or a chirp, or linear frequency sweep, of the primary signals. At 4 kHz the secondary beam source level is at least 204 dB re 1 μ Pa at 1 m. It can be steered over an 80° sector.

The operating regime of the TOPAS parametric sonar can be inferred. The primary beamwidth is about 3.5°. If the array is assumed to be circular, then its radius can be determined from the formula relating the product of wave number k and radius a , namely

$$ka = 1.615/\sin(\Delta\theta), \quad (6)$$

where $\Delta\theta$ is the half-beamwidth, namely 1.75° or roughly $\pi/103=0.0305$ rad. The product ka is thus about 53, and the directivity index at 18 kHz is $20 \log ka=34$ dB to an excellent approximation.⁵⁴ The output power P is specified as being at least 32 kW. Assuming, conservatively, that this is also the acoustic power, then

$$SL = 10 \log P + DI + 171, \quad (7)$$

or about 250 dB re 1 μ Pa at 1 m. The scaled source level is thus $SL+20 \log 18=276$ dB re 1 μ Pa at 1 m and 1 kHz.

Secondary, difference-frequency source levels have been both calculated and measured for the mentioned TOPAS PS 18 sonar by Dybedal.¹⁷ These measurements were made for each of four frequencies, 0.5, 1, 2, and 4 kHz, at each of three ranges: 40, 60, and 120 m. The calculations indicate a gradual approach to a maximum level, or saturation state, which is still not reached at 120 m, whereas the data suggest attainment of saturation already at 60 m for the extreme frequencies of investigation, 0.5 and 4 kHz, but not so for the intermediate frequencies, 1 and 2 kHz.

Since $ka=53$, and $k=2\pi/\lambda$, where λ is the wavelength at the mean primary frequency, 18 kHz, namely 8.2 cm, a is about 70 cm. The nearfield distance from Eq. (5) is $R_0 = \pi a^2/\lambda$, hence about 18.3 m, which is consistent with the minimum depth of operation, namely 20 m, as given in the manufacturer's specifications. This distance will vary with the hydrographic conditions through the sound speed, hence λ . For a temperature change from 0 to 20 °C in seawater of salinity 35 ppt, λ will change from 8.05 to 8.45 cm and R_0 will vary from 19.1 to 18.2 m.

The absorption coefficient at the mean primary frequency of 18 kHz for water of temperature 10 °C and salinity 35 ppt is about 0.0027 dB/m.⁵⁵ Thus, the absorption at 18.3 m, namely $\alpha_0 R_0$, is about 0.05 dB. Changes in hydrographic conditions will generally change α_0 , but without significantly changing the magnitude of $\alpha_0 R_0$ relative to unity.

The particular parametric sonar therefore conforms to the third of the Moffett and Konrad regimes, without nearfield saturation due to harmonic distortion and with difference-frequency generation in the farfield of the primary, source arrays. As noted in Sec. II B, this corresponds to the more general case. Thus, calibration measurements with a standard target would have to be made at substantial ranges compared to those of more conventional calibrations, e.g., of high-frequency scientific echo sounders, where target

ranges are generally less than about 25 m. Target deployment and control at large ranges are technical challenges. Water depth, surprisingly perhaps, may be less of an issue because of the special directionality characteristics of parametric sonar, namely the narrowness of the transmit beam and absence of sidelobes. Depending on the mounting of the particular TOPAS PS18 unit, calibration could be performed in the horizontal plane. Finding a relatively distant standard target, similarly, may not present particular problems since the TOPAS PS18 beam can be steered over an 80° sector. For present purposes, however, the critical problem is finding a suitable target for use over the difference-frequency range 0.5–6 kHz.

IV. TARGET OPTIMIZATION

The problem of target optimization was originally addressed for high-frequency scientific echo sounders,⁴³ with reference to the theory of acoustic scattering by homogeneous, solid elastic spheres^{56,57} or as a limiting case of scattering by homogeneous elastic shells.⁵⁸ Electrolytic-grade copper was identified as being particularly suitable because of its hardness, other elastic properties, corrosion resistance, commercial availability in a high state of purity due to its worldwide use by the electrical industry, and specification of its physical properties. Variations in the backscattering form function with respect to wave number k due to changes in sound speed in the immersion medium were exploited to achieve a substantially higher backscattering cross section than that of the geometric cross section. The optimization condition imposed at 38 kHz was that the backscattering cross section be an extremum with respect to temperature and salinity over the range [0,30]°C in seawater of salinity 35 ppt, while the backscattering cross section was of order 40 cm². The result was a copper sphere of diameter 60 mm. Subsequent optimization at this and other ultrasonic frequencies has demonstrated the suitability of tungsten carbide with 6% cobalt binder. Materials such as aluminum, brass, bronze, and steel for calibration spheres are always *prima facie* attractive, but knowing their exact composition and elastic properties can be problematic.

For the problem of finding a suitable target at the rather low frequencies of the TOPAS PS18 parametric sonar, several optimization conditions were considered in addition to the condition that the backscattering cross section be independent of orientation, i.e., a sphere. Briefly, the condition was imposed that the target be as acoustically powerful as possible consistent with being physically manageable. Initially, a mass limit of 10 kg was imposed. This condition was then relaxed in the case of aluminum to allow an immersion weight of 200 N, or an effective mass under immersion of about 20 kg.

Three different materials were considered, based on experience at sea with sonar targets made of each of these, although at higher, ultrasonic frequencies. These materials are aluminum alloys, copper, and tungsten carbide with 6% cobalt binder. Nominal physical properties of each, sufficient for scattering computations, are presented in Table I. The property values for aluminum alloys were derived from list-

TABLE I. Properties of three materials: mass density ρ_1 , longitudinal-wave sound speed c_1 , and transverse-wave sound speed c_2 .

Material	ρ_1 (kg/m ³)	c_1 (m/s)	c_2 (m/s)
Aluminum	2731	6334	3117
Copper	8947	4760	2288.5
Tungsten carbide	14,900	6853	4171

ings of mass density and elasticity moduli for each of 68 alloys, including both wrought and cast aluminum.⁵⁹ The values for copper are those of the essentially pure, electrolytic grade copper of earlier experience.⁴³ The values for tungsten carbide with 6% cobalt binder are those determined by laboratory measurement.⁶⁰

The backscattering cross section was computed from the farfield backscattering form function according to Faran's theory,^{56,57} but with the corrections noted in Ref. 43. Since this is to be used over a very wide bandwidth but without detailed specification of receiver processing operations, the backscattering cross section is expressed according to Eq. (3). The target strength is derived from this according to

$$TS = 10 \log \left[\frac{\sigma}{4\pi r_0^2} \right], \quad (8)$$

where r_0 is the reference distance of 1 m.

V. RESULTS AND DISCUSSION

Initial computations for the 100-mm-diam spheres indicated quite low target strengths at the lowest frequencies in the band 0.5–6 kHz, illustrated in Table II. Larger spheres of aluminum and copper were then considered consistent with the 10 kg mass limit. Again, the results, shown in Table II, indicated weak target strengths at the lowest frequencies.

The mass constraint was relaxed further to allow the immersion weight of an aluminum sphere to be 200 N. The immersed weight and effective mass of a 280-mm-diam sphere of aluminum are 192 N and 19.6 kg, respectively. In this case, the nominal TS values at 0.5 and 1 kHz are -43.4 and -32.9 dB, respectively. The dependence of TS on frequency is shown in Fig. 1. This assumes the nominal material property values indicated in Table I and assumed seawater properties of mass density 1027 kg/m³ and sound speed 1490 m/s, corresponding to salinity 35 ppt and temperature 10 °C.

In order to assess the stability of acoustic properties of this target with respect to potential uncertainty in knowledge

TABLE II. Sphere diameter $2a$, mass m , immersed effective mass m_{im} , immersion weight w , and target strength values at 0.5 and 1.0 kHz.

Material	$2a$ (mm)	m (kg)	m_{im} (kg)	w (N)	TS(0.5)	TS(1)
Aluminum	100	1.430	0.892	8.8	-69.8	-58.0
Copper	100	4.685	4.147	40.7	-67.7	-55.8
Tungsten carbide	100	7.802	7.264	71.3	-67.3	-55.4
Copper	125	9.150	8.099	79.5	-61.9	-50.1
Aluminum	190	9.808	6.120	60.0	-53.3	-41.9
Aluminum	280	31.390	19.586	192.1	-43.4	-32.9

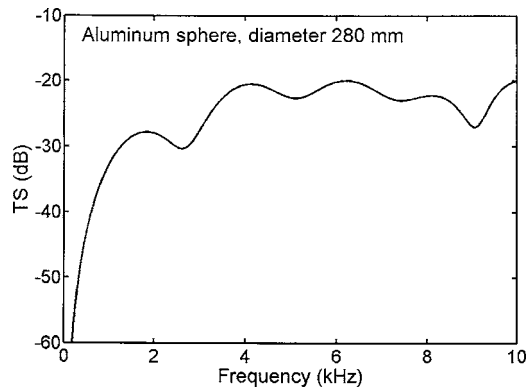


FIG. 1. Theoretical target strength spectrum of an immersed 280-mm-diam sphere of aluminum, with material properties given in Table I and assuming a medium sound speed 1490 m/s and mass density 1027 kg/m³.

of the physical properties of aluminum, $\pm 2\%$ variations in each of the properties were considered. This is believed to be a reasonable range, based on the observed variations in mass density and elastic moduli according to the mentioned listings.⁵⁹ In the case of the transverse-wave sound speed, whose influence on the acoustic scattering properties is recognized to be most sensitive,^{60–63} the total range of variation is 3029–3206 m/s, that is, $\pm 3\%$, but extreme values can be excluded by avoiding the particular alloys associated with these. The nominal reference against which the comparisons were made is that of Fig. 1. Effects of uncertainty in aluminum mass density, longitudinal-wave sound speed, and transverse-wave sound speed are shown in Figs. 2–4, respectively.

Variations in target strength with respect to the aluminum mass density, Fig. 2, are seen to lie within ± 0.1 dB over nearly the entire frequency band of interest, 0.5–6 kHz, with maximum difference of about ± 0.12 dB near 2.3 kHz. Variations in target strength with respect to each of the longitudinal- and transverse-wave sound speeds, Figs. 3 and 4, both lie within about ± 0.1 dB over the same band of interest.

Larger excursions in target strength due to potential uncertainty in the sound speeds are observed over the frequency band 6–10 kHz, with that of transverse-wave sound speed dominating that of longitudinal-wave sound speed. This is not unexpected, given earlier observations of the sen-

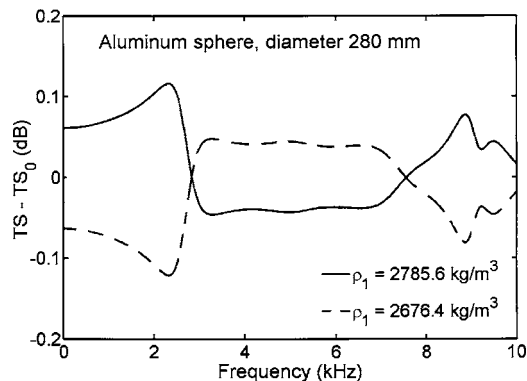


FIG. 2. Spectrum of differences in theoretical target strength of an immersed 280-mm-diam Al sphere for $\pm 2\%$ changes in mass density of aluminum.

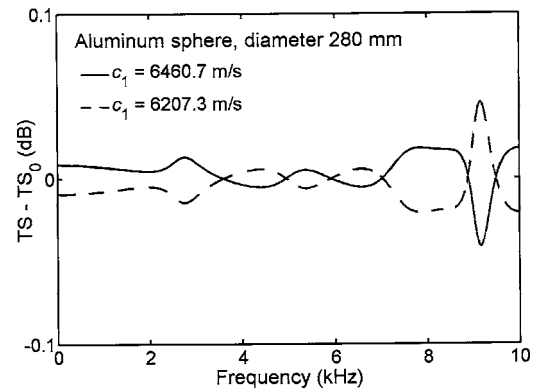


FIG. 3. Spectrum of differences in theoretical target strength of an immersed 280-mm-diam Al sphere for $\pm 2\%$ changes in longitudinal-wave sound speed of aluminum.

sitivity of acoustic scattering to the lesser, transverse-wave sound speed in the resonance regime, as in Refs. 60–63. It might be anticipated, because free vibrations of elastic spheres, hence scattering by the same,⁵⁶ are sensitive to the transverse-wave sound speed and insensitive to the longitudinal-wave sound speed.^{57,61}

The effects of variations in properties of the immersion medium are shown in Figs. 5 and 6. In the first, the effect of variations is shown for variations in mass density over a range encompassing the cold freshwater–warm seawater range, namely 998–1032 kg/m³. The greatest excursions lie within ± 0.2 dB of the nominal number.

The effect of variations in medium sound speed is shown in Fig. 6 for a $\pm 2\%$ range of variation. The variations are as large as ± 0.7 dB over the frequency band 0.5–6 kHz; specifically, 0.69 dB at 3.02 kHz and -0.68 dB at 3.08 kHz. These are the largest excursions in target strength due to the variations of any property over the frequency band of interest. However, the sound speed is generally known to a very high accuracy⁶⁴ during conduct of a calibration exercise, when the temperature and salinity are routinely measured, allowing compensation in the assignment of target strength. Variations in the water mass density⁶⁵ are similarly known to a very high accuracy for the measured hydrography.

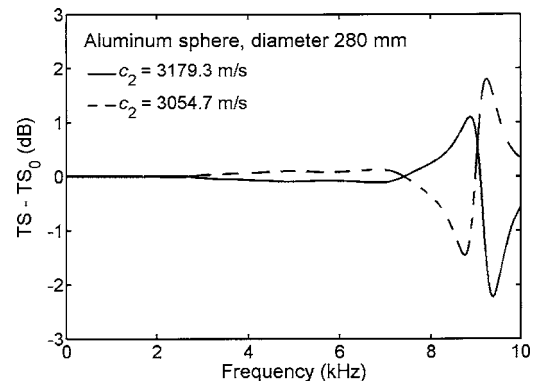


FIG. 4. Spectrum of differences in theoretical target strength of an immersed 280-mm-diam Al sphere for $\pm 2\%$ changes in transverse-wave sound speed of aluminum.

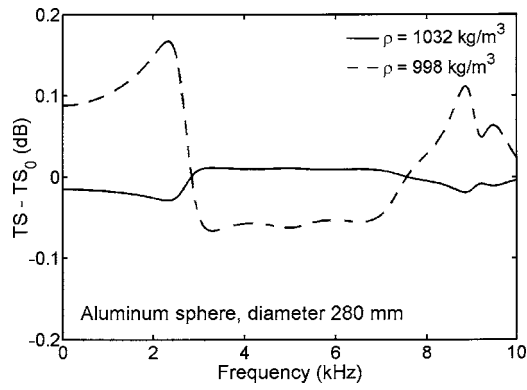


FIG. 5. Spectrum of differences in theoretical target strength of an immersed 280-mm-diam Al sphere for changes in medium mass density spanning those from cold freshwater to warm seawater, namely 998–1032 kg/m³.

VI. CONCLUSIONS

The problem of calibrating parametric sonars at low difference frequencies by the standard-target method has been considered. For a particular parametric sonar, with difference frequencies in the band 0.5–6 kHz, a standard target has been specified that satisfies the conditions of having a target strength independent of target orientation and a weight in water less than 200 N. The result is a 280-mm-diam sphere of aluminum, with mass 31.4 kg, immersed weight 192 N, and effective equivalent mass in water 19.6 kg. The total range of variation in target strength over the frequency band is [−43.4, −20.1] dB. The greatest source of variation in target strength is that of the sound speed in the immersion medium, about ±0.7 dB over the frequency band; however, this can be eliminated by careful attention to the hydrographic conditions. The second greatest source of variation in target strength is that of the mass density of the immersion medium, of order ±0.2 dB, which similarly can be eliminated by attention to the hydrography.

While a standard target has been specified for the performance parameters of a particular parametric sonar, the target could also be used to calibrate other sonars over a similar frequency band. Thus, conventional active sonars operating at low-kilohertz frequencies could be calibrated using the derived 280-mm-diam aluminum sphere.

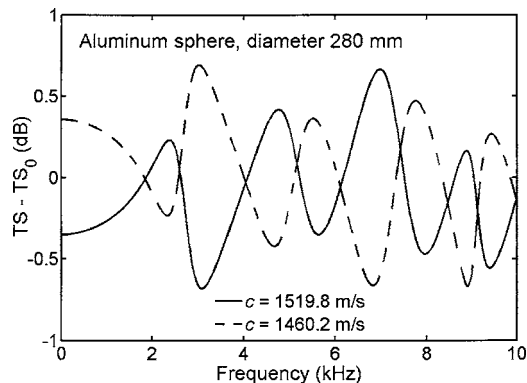


FIG. 6. Spectrum of differences in theoretical target strength of an immersed 280-mm-diam Al sphere for ±2% changes in medium sound speed.

The circumstances of calibrating a particular sonar may require that the calibration be performed in the transducer nearfield when the sonar is used in the farfield or vice versa. In either case, attention must be given to the so-called nearfield problem to address extrapolation from the nearfield to the farfield and/or from the farfield to the nearfield.

ACKNOWLEDGMENTS

M. Parmenter and G. Feijoo are thanked for assistance in preparing the figures. S. Barkley is thanked for advising on electronic document compatibility.

APPENDIX

Nonlinear interactions of waves occur in solids as well as in fluids.^{66–68} An exact treatment of this subject would require significant effort, but this is unnecessary since it is possible to bound the effect. Initially, it is shown that second-harmonic generation is negligible for the particular target, namely the 280-mm-diam aluminum sphere, at the primary frequencies, which suggests even weaker difference-frequency generation when two harmonic waves are present.

The one-dimensional case of finite-amplitude waves in an isotropic elastic medium has been summarized succinctly by Breazeale:⁶⁸

$$\rho_0 \frac{\partial^2 \xi}{\partial t^2} = K_2 \frac{\partial^2 \xi}{\partial x^2} + (3K_2 + K_3) \frac{\partial^2 \xi}{\partial x^2} \frac{\partial \xi}{\partial x}, \quad (\text{A1})$$

where ρ_0 is the unperturbed solid density, ξ is the particle displacement, and x is the distance in the direction of propagation. For an isotropic solid, the coefficients K_2 and K_3 are identical to C_{11} and C_{111} , respectively, where the coefficients C_{ij} are the ordinary second-order elastic constants, and C_{ijk} are the third-order elastic constants. The second harmonic can be derived by a perturbation solution to Eq. (A1):^{69,70}

$$\xi = A_1 \sin(kx - \omega t) + A_2 \cos 2(kx - \omega t) + \dots \quad (\text{A2})$$

Solving, $A_2 = \beta A_1^2 k^2 x / 4$, with the nonlinearity parameter $\beta = -(3 + K_3 / K_2) / 2$. Given the pressure amplitude p_1 , the value for A_1 can be derived from the first-order wave equation, hence $A_1 = p_1 / (\rho_0 c_0 \omega)$, where c_0 is the small-amplitude longitudinal-wave sound speed, i.e., the ordinary longitudinal-wave sound speed.

For the mean primary frequency 18 kHz, $\omega = 1.13 \times 10^5$ rad/s. The second-order elastic constant $K_2 = C_{11}$ is assumed to be 108 GPa.⁷¹ The third-order elastic constant $K_3 = C_{111}$ is derived by averaging the five values given for aluminum alloys in Ref. 72, hence −1942 GPa. Assuming the value for mass density given in Table I, $\rho_0 = 2731$ kg/m³, the longitudinal-wave sound speed is $c_0 = 6289$ m/s. The source level is assumed to be 250 dB re 1 μ Pa at 1 m, as in Sec. III. If the nearest measurement to the transmitter is made at 20 m, which is one-half the distance of the nearest measurement made by Dybedal,¹⁷ the incident pressure level is 234 dB re 1 μ Pa. The corresponding pressure level is $10^{5.7}$ Pa = 0.5 MPa in the immersion medium and, using the series solution for the form factor,⁵⁸ the displacement amplitude on the surface of the sphere facing the incident wave is

found to be $A_1=2.68 \times 10^{-6}$ m. The displacement amplitude decreases toward the center, so the displacement is actually greatest at the surface. If the distance of propagation in the target is taken to be the diameter, 280 mm, then since $\beta=7.5$, $A_2=1.22 \times 10^{-9}$ m. That is, the displacement amplitude of the second harmonic is about 4.5×10^{-4} that of the fundamental. This is in effect an upper limit since inspection of the solution for the displacement amplitude shows that it is greatest at the point selected.

An attempt has been made to observe the nonlinear generation of second harmonics in an available aluminum sphere of diameter 60 mm. The sphere was suspended in the beam of a RESON broadband transducer, model TC2116, spanning the frequency band 25–150 kHz. The target range was 3.5 m; that of the wall of the indoor tank, 7.1 m. Amplitude-weighted sinusoidal pulses were transmitted at 100-Hz intervals, with pulse duration 0.5 ms, across the entire band with an insonification level of approximately 173 dB re 1 μ Pa at the target. Echoes were recorded with nominal 72-dB dynamic range. The echo signals were processed for both the fundamental, second, and third harmonics. The noise floor, as defined by the energy in the second and third harmonics, was 40–50 dB less than that of the fundamental echo signal due to ordinary linear backscattering by the target. Several radio signals raised this noise floor in places. Harmonic analysis of both the target and wall echoes gave similar results for the second and third harmonics. Barrett and Matsinger,⁷³ with a similar if cleaner experimental configuration, were able to observe both sum- and difference-frequency waves due to nonlinear interaction in a crystal, but with primaries of order 0.2–1 GPa and 7–50 MHz.

The measurements were repeated with the target sphere placed close to the projector, with an insonification level of 237 dB re 1 μ Pa. Analog filtering was used to reduce the second and third harmonic levels of the drive signal such that the limit of harmonic measurement was –60 dB with respect to the carrier, that is, the fundamental frequency. No detectable harmonics were introduced by the insertion of the target sphere. It is believed that the present results for the second harmonic, as well as the third harmonic, represent noise and contain no detectable signal due to the nonlinear behavior of the target sphere.

A theoretical three-dimensional approach to the nonlinear interaction of two harmonic waves in a semi-infinite solid is also available.^{74–77} Let u_1 and u_2 denote the displacement amplitudes of the two primary longitudinal waves incident on and just within the surface of the solid, and k_1 and k_2 denote the respective wave numbers in the solid. To a rather rough approximation, based on collinearity of all longitudinal waves in the solid, notwithstanding the angle condition required for resonance,^{74,76} and assumption of interaction in a spherical volume of radius a , the difference-frequency transverse-wave displacement amplitude is

$$u^{(2)} \approx \frac{Du_1u_2k_1k_2(k_2 - k_1)a^2}{3\rho_0c_0^2}, \quad (\text{A3})$$

where $D=626$ GPa is the result of combining five elastic constants: two independent second-order constants and three independent third-order constants, required to de-

scribe the elasticity of an isotropic solid, with numerical values derived from Hearmon.^{71,72}

For an incident wave of pressure amplitude 0.5 MPa, the displacements at the primary frequencies, 15 and 21 kHz, are found from the series solution⁵⁸ to be $u_1=2.00 \times 10^{-6}$ m and $u_2=1.76 \times 10^{-6}$ m, respectively. Evaluation of Eq. (A3) for $a=0.14$ m yields the estimate $u^{(2)} \approx 2.1 \times 10^{-10}$ m, which compares with that of the second-harmonic displacement amplitude A_2 computed earlier. This is suggestive of a symmetry in the difference- and sum-frequency components, as with the parametric acoustic array in fluids, although inexact owing to different wave numbers at the respective frequencies. Thus the displacement amplitude of the difference-frequency wave is about 10^{-4} times that of the fundamental. A more meticulous treatment of the problem would undoubtedly change the estimate of $u^{(2)}$, but it is believed to within an order of magnitude. However, according to the basic theory, this result still applies to spherical waves interacting in a semi-infinite solid rather than to plane-wave-induced pressure waves interacting inside an immersed, finite, solid sphere.

¹P. J. Westervelt, "Parametric acoustic array," *J. Acoust. Soc. Am.* **35**, 535–537 (1963).

²M. B. Moffett and W. L. Konrad, *Nonlinear Sources and Receivers*, in *Encyclopedia of Acoustics* Vol. **1**, edited by M. J. Crocker (Wiley, New York, 1997), pp. 607–617.

³R. F. W. Coates, M. Zheng, and L. Wang, "BASS 300 PARACOM: A model underwater parametric communication system," *IEEE J. Ocean. Eng.* **21**, 225–232 (1996).

⁴O. Diachok, "Effects of absorptivity due to fish on transmission loss in shallow water," *J. Acoust. Soc. Am.* **105**, 2107–2128 (1999).

⁵O. Diachok, "Absorption spectroscopy: A new approach to estimation of biomass," *Fish. Res.* **47**, 231–244 (2000).

⁶W. L. Konrad, W. I. Roderick, and L. F. Carlton, "Parametric sonar as applied to high resolution boundary scattering measurement," *J. Acoust. Soc. Am.* **72**, S41 (1982).

⁷B. Nuetzel, H. Herwig, J. M. Monti, and P. D. Koenigs, "A further investigation of acoustic scattering from the sea surface," *J. Acoust. Soc. Am.* **79**, S68 (1986).

⁸J. M. Syck and C. L. Brown, "Volume scattering measurements using parametric sonar," *J. Acoust. Soc. Am.* **83**, S48 (1988).

⁹W. I. Roderick, R. K. Dullea, and J. M. Syck, "High resolution bottom backscatter measurements," *J. Acoust. Soc. Am.* **75**, S31 (1984).

¹⁰O. Bergem, E. Pouliquen, G. Canepa, and N. G. Pace, "Time-evolution modeling of seafloor scatter. II. Numerical and experimental evaluation," *J. Acoust. Soc. Am.* **105**, 3142–3150 (1999).

¹¹P. C. Hines, J. C. Osler, and D. J. MacDougald, "Acoustic backscatter measurements from littoral seabeds at shallow grazing angles at 4 and 8 kHz," *J. Acoust. Soc. Am.* **117**, 3504–3516 (2005).

¹²A. Caiti, O. Bergem, E. Pouliquen, A. Biancheri, and S. Scaricciottoli, "Seabed characterization by inversion of parametric sonar data: Selection of the cost function," in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Pouliquen, O. Bergem, and A. P. Lyons, SACLANTCEN Conference Proceedings Series No. CP-45 (1997), pp. 83–90.

¹³A. Caiti, O. Bergem, E. Pouliquen, and E. Michelozzi, "Model-based seafloor identification with parametric sonar data: Experimental results," *J. Acoust. Soc. Am.* **105**, 1207 (1999).

¹⁴A. Caiti, J. M. Hovem, and B. Berntsen, "Quantitative characterization of the uppermost seabed stratum: Some results from the ISACS project," *J. Acoust. Soc. Am.* **109**, 2393 (2001).

¹⁵A. P. Lyons and E. Pouliquen, "Measurements of high-frequency acoustic scattering from seabed vegetation," *J. Acoust. Soc. Am.* **103**, 2934 (1998).

¹⁶C. C. Fox and O. L. Akervold, "Parametric acoustic arrays," *J. Acoust. Soc. Am.* **53**, 382 (1973).

¹⁷J. Dybedal, "TOPAS: Parametric end-fire array used in offshore applications," in *Advances in Nonlinear Acoustics*, edited by H. Hoback (World Scientific, Singapore, 1993), pp. 264–275.

- ¹⁸S. Fioravanti, A. Maguer, and A. Løvik, "A synthetic aperture parametric sonar for buried object detection," *J. Acoust. Soc. Am.* **100**, 2636 (1996).
- ¹⁹W. L. J. Fox and A. Maguer, "Experimental results for detection of buried objects at low grazing angles," *J. Acoust. Soc. Am.* **104**, 1784 (1998).
- ²⁰S. Guyonic, "Buried objects detection and classification with a 3-D parametric sonar," *J. Acoust. Soc. Am.* **108**, 2649 (2000).
- ²¹K. W. Commander, J. L. Lopes, and R. Lim, "Variation of sea-bed backscattering strength due to Bragg scattering," *J. Acoust. Soc. Am.* **113**, 2300 (2003).
- ²²T. G. Muir and R. S. Adair, "Potential use of parametric sonar in marine archaeology," *J. Acoust. Soc. Am.* **52**, 122 (1972).
- ²³W. L. Konrad and W. L. Clay, "Discovery of a sunken vessel by parametric sonar," *J. Acoust. Soc. Am.* **67**, S67 (1980).
- ²⁴W. T. Wood, J. F. Gettrust, M. K. Sen, and J. G. Kosalas, "Bottom/subbottom surveying using a new, parametric, sidescan sonar," in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Poulquien, O. Bergem, and A. P. Lyons, SACLANTCEN Conference Proceedings Series No. CP-45 (1997), pp. 607–612.
- ²⁵M. E. Zakharia, S. Hibral, C. Pollet, and E. Rigaud, "Parametric synthetic interferometric sonar for bottom and subbottom imaging and mapping," *J. Acoust. Soc. Am.* **116**, 2577 (2004).
- ²⁶M. Amate, A. Hétet, S. Guyonic, M. Legris, R. Bellec, F. Maussang, J. Chanussot, P. Cervenka, and J. Marchal, "Buried mines detection and classification: Advanced technologies and signal processing," *Proceedings IEEE Oceans 2005 Europe*, Vol. **1**, pp. 153–159.
- ²⁷O. Bergem and N. G. Pace, "Installation and calibration of a parametric array for shallow water backscatter measurements," *Proceedings Oceans 1996 MTS/IEEE Conference*, pp. 773–778.
- ²⁸M. B. Moffett and R. H. Mellen, "Model for parametric acoustic sources," *J. Acoust. Soc. Am.* **61**, 325–337 (1977).
- ²⁹J. Berntsen, "Numerical calculations of finite amplitude sound beams," in *Frontiers of Nonlinear Acoustics*, *Proceedings of the 12th International Symposium on Nonlinear Acoustics* (Elsevier, London, 1990), pp. 191–196.
- ³⁰D. E. Weston, *Sound Propagation in the Presence of Bladder Fish*, in *Underwater Acoustics Vol. 2*, edited by V. M. Albers (Plenum, New York, 1967), pp. 55–88.
- ³¹B. S. McCartney and A. R. Stubbs, "Measurements of target strength of fish in dorsal aspect, including swimbladder resonance," *J. Sound Vib.* **15**, 397–420 (1971).
- ³²R. H. Love, "Resonant acoustic scattering by swimbladder-bearing fish," *J. Acoust. Soc. Am.* **64**, 571–580 (1978).
- ³³A. Løvik and J. M. Hovem, "An experimental investigation of swimbladder resonance in fishes," *J. Acoust. Soc. Am.* **66**, 850–854 (1979).
- ³⁴C. Feuillade and R. W. Nero, "A viscous-elastic swimbladder model for describing enhanced-frequency resonance scattering from fish," *J. Acoust. Soc. Am.* **103**, 3245–3255 (1998).
- ³⁵K. G. Foote and D. T. I. Francis, "Scheme for parametric sonar calibration by standard target," in *Proceedings Oceans 2005 MTS/IEEE Conference*, 6 pp.
- ³⁶K. G. Foote and D. N. MacLennan, "Comparison of copper and tungsten carbide calibration spheres," *J. Acoust. Soc. Am.* **75**, 612–616 (1984).
- ³⁷S. P. Robinson, P. M. Harris, J. Ablitt, G. Hayman, A. Thompson, A. L. van Buren, J. F. Zalesak, R. M. Drake, A. E. Isaev, A. M. Enyakov, C. Purcell, Z. Houqing, W. Yuebing, Z. Yue, P. Botha, and D. Krüger, "An international key comparison of free-field hydrophone calibrations in the frequency range 1 to 500 kHz," *J. Acoust. Soc. Am.* **120**, 1366–1373 (2006).
- ³⁸R. J. Bobber, *Underwater electroacoustic measurements* (Naval Research Laboratory, Washington, DC, 1970).
- ³⁹K. G. Foote, "Maintaining precision calibrations with optimal copper spheres," *J. Acoust. Soc. Am.* **73**, 1054–1063 (1983).
- ⁴⁰K. G. Foote, H. P. Knudsen, G. Vestnes, D. N. MacLennan, and E. J. Simmonds, "Calibration of acoustic instruments for fish density estimation: A practical guide," ICES Cooperative Research Rep. No. 144, 1987, 69 pp.
- ⁴¹H. P. Knudsen, "T-A-F: Time-amplitude-frequency: A special electronic unit for measuring the TVG function in research echo-sounders," *Fiske-ridir. Skr., Ser. Havunders.* **17**, 529–541 (1985).
- ⁴²K. G. Foote, "Standard-target calibration of broadband sonars," *J. Acoust. Soc. Am.* **108**, 2484 (2000).
- ⁴³K. G. Foote, "Optimizing copper spheres for precision calibration of hydroacoustic equipment," *J. Acoust. Soc. Am.* **71**, 742–747 (1982).
- ⁴⁴K. G. Foote, "Spheres for calibrating an eleven-frequency acoustic measurement system," *J. Cons., Cons. Int. Explor. Mer* **46**, 284–286 (1990).
- ⁴⁵K. G. Foote, "Calibrating a narrowband 18-kHz sonar," *Proceedings of the MTS/IEEE Oceans 2001 Conference*, pp. 2503–2505.
- ⁴⁶K. G. Foote, T. Knutsen, P. R. Atkins, C. Bongiovanni, D. T. I. Francis, P. K. Eriksen, and T. Mortensen, "A seven-octave-bandwidth echo sounding system for application to fish and zooplankton," *Acust. Acta Acust.* **85**, S68 (1999); *J. Acoust. Soc. Am.* **105**, 994 (1999).
- ⁴⁷K. G. Foote, P. R. Atkins, D. T. I. Francis, and T. Knutsen, "Measuring echo spectra of marine organisms over a wide bandwidth," in *Proceedings of the International Conference on Underwater Acoustic Measurements: Technologies and Results*, edited by J. S. Papadakis and L. Bjørnør, Heraklion, Crete, Greece, 28 June–1 July 2005, pp. 501–508.
- ⁴⁸K. G. Foote, D. Chu, T. R. Hammar, K. C. Baldwin, L. A. Mayer, L. C. Hufnagle, Jr., and J. M. Jech, "Protocols for calibrating multibeam sonar," *J. Acoust. Soc. Am.* **117**, 2013–2027 (2005).
- ⁴⁹K. G. Foote, D. T. I. Francis, H. Furset, and H. Hobæk, "Spheres for calibrating high-frequency broadband echo sounders," *J. Acoust. Soc. Am.* **110**, 1112–1113 (1999).
- ⁵⁰Y. Miyanoana, K. Ishii, and M. Furusawa, "Spheres to calibrate echo sounders at any frequency," *Nippon Suisan Gakkaishi* **59**, 933–942 (1993).
- ⁵¹J. N. Tjøtta and S. Tjøtta, "Effects of finite aperture in a parametric acoustic array," *J. Acoust. Soc. Am.* **68**, 970–972 (1980).
- ⁵²J. N. Tjøtta and S. Tjøtta, "Nonlinear interaction of two collinear, spherically spreading sound beams," *J. Acoust. Soc. Am.* **67**, 484–490 (1980).
- ⁵³M. F. Hamilton, "Parametric acoustic array formation in dispersive fluids," *J. Acoust. Soc. Am.* **76**, 1474–1492 (1984).
- ⁵⁴R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
- ⁵⁵R. E. Francois and G. R. Garrison, "Sound absorption based on ocean measurements. II. Boric acid contribution and equation for total absorption," *J. Acoust. Soc. Am.* **72**, 1879–1890 (1982).
- ⁵⁶J. J. Faran, Jr., "Sound scattering by solid cylinders and spheres," *J. Acoust. Soc. Am.* **23**, 405–418 (1951).
- ⁵⁷R. Hickling, "Analysis of echoes from a solid elastic sphere in water," *J. Acoust. Soc. Am.* **34**, 1582–1592 (1962).
- ⁵⁸R. R. Goodman and R. Stern, "Reflection and transmission of sound by elastic spherical shells," *J. Acoust. Soc. Am.* **34**, 338–344 (1962).
- ⁵⁹ASM Committee on Aluminum and Aluminum Alloys, *Metals handbook*, 9th ed. *Properties and Selection: Nonferrous Alloys and Pure Metals* (American Society for Metals, Metals Park, OH, 1979), Vol. **2**.
- ⁶⁰D. N. MacLennan and J. R. Dunn, "Estimation of sound velocities from resonance measurements on tungsten carbide calibration spheres," *J. Sound Vib.* **97**, 321–331 (1984).
- ⁶¹W. G. Neubauer, R. H. Vogt, and L. R. Dragonette, "Acoustic reflection from elastic spheres. I. Steady-state signals," *J. Acoust. Soc. Am.* **55**, 1123–1129 (1974).
- ⁶²K. G. Foote, "Refined determination of the rigidity modulus by echo sounder measurement," in *1981 Ultrasonics Symposium Proceedings* (IEEE, New York, 1981), pp. 900–903.
- ⁶³K. G. Foote, "Acoustic resonance determination of the rigidity modulus," *J. Acoust. Soc. Am.* **110**, 2628 (2001).
- ⁶⁴K. V. Mackenzie, "Nine-term equation for sound speed in the oceans," *J. Acoust. Soc. Am.* **70**, 807–812 (1981).
- ⁶⁵G. Dietrich, *Ozeanographie: Physikalische Eigenschaften des Meerwassers* ("Oceanography: Physical Properties of Sea Water"), in *Landolt-Börnstein Zahlenwerte und Funktionen* (Numerical Values and Functional Relationships), 6th ed., Vol. **3** (Springer, Berlin, 1952), pp. 426–440.
- ⁶⁶R. T. Beyer, "Nonlinear Acoustics," in *Physical Acoustics: Principles and Methods*, edited by W. P. Mason (Academic, New York, 1965), Vol. **II**, pt. B.
- ⁶⁷R. T. Beyer, *Nonlinear Acoustics* (Naval Ship Systems Command, Washington, DC, 1974).
- ⁶⁸M. A. Breazeale, "Finite-amplitude waves in solids," in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), Vol. **1**, pp. 227–235.
- ⁶⁹J. Melngailis, A. A. Maradudin, and A. Seeger, "Diffraction of light by ultrasound in anharmonic crystals," *Phys. Rev.* **131**, 1972–1975 (1963).
- ⁷⁰M. A. Breazeale and J. Ford, "Ultrasonic studies of the nonlinear behavior of solids," *J. Appl. Phys.* **36**, 3486–3490 (1965).
- ⁷¹R. F. S. Hearmon, "The elastic constants of crystals and other anisotropic materials," in *Landolt-Börnstein Zahlenwerte und Funktionen aus Naturwissenschaften und Technik* (Numerical Values and Functional Relationships in Science and Technology), *New Series, Group III* (Springer, Berlin,

1979), Vol. **11**, pp. 1–244.

⁷²R. F. S. Hearmon, “The third- and higher-order elastic constants,” in *Landolt-Börnstein Zahlenwerte und Funktionen aus Naturwissenschaften und Technik (Numerical Values and Functional Relationships in Science and Technology), New Series, Group III* (Springer, Berlin, 1979), Vol. **11**, pp. 245–286.

⁷³H. H. Barrett and J. H. Matsinger, “Interaction of almost-collinear longitudinal phonons,” *Phys. Rev.* **154**, 877–886 (1967).

⁷⁴G. L. Jones and D. R. Kobett, “Interaction of elastic waves in an isotropic solid,” *J. Acoust. Soc. Am.* **35**, 5–10 (1963).

⁷⁵P. Waldow and F. R. Rollins, Jr., “Remarks on a paper by Jones and Kobett,” *J. Acoust. Soc. Am.* **35**, 1060–1061 (1963).

⁷⁶J. D. Childress and C. G. Hambrick, “Interactions between elastic waves in an isotropic solid,” *Phys. Rev.* **136**, A411–A418 (1964).

⁷⁷I. A. Beresnev, “Interaction of two spherical elastic waves in a nonlinear five-constant medium,” *J. Acoust. Soc. Am.* **94**, 3400–3404 (1993).

Imaging with unfocused regions of focused ultrasound beams

Roger Zemp

Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130

Michael F. Insana^{a)}

Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 7 April 2006; revised 18 December 2006; accepted 19 December 2006)

This article gives an analytical, computational, and experimental treatment of the spatial resolution encoded in unfocused regions of focused ultrasound beams. This topic is important in diagnostic ultrasound since ultrasound array systems are limited to a single transmit focal point per acoustic transmission, hence there is a loss of spatial resolution away from the transmit focus, even with the use of dynamic receive focusing. It is demonstrated that the spatial bandwidth of a Gaussian-apodized beam is approximately constant with depth, which means that there is just as much encoded spatial resolution away from the transmit focus as there is in the focal region. The practical application of this principle is discussed, an algorithm for retrospectively focusing signals from unfocused regions of fixed-focus beams is presented, and a quantitative comparison between the authors' methods and dynamic-receive beamforming is provided. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2434247]

PACS number(s): 43.60.Fg [TDM]

Pages: 1491–1498

I. INTRODUCTION

Current ultrasound arrays systems typically transmit acoustic pulses at a fixed focal depth, then dynamically adjust element phase delays so that the receive focus is steered along the receive scan line. Because current arrays systems do not have the ability to dynamically focus on transmission, spatial resolution degrades away from the transmit focus. In this article we discuss the acoustics of unfocused beams in the context of ultrasonic imaging. The framework provides fundamental insights and offers practical applications.

Originally, this work was motivated by a recent detection performance theory developed by our group.^{1,2} A Bayesian ideal observer for the task of detecting low contrast lesions was found to have a log-likelihood test statistic which first whitened data by Wiener spatiotemporal deconvolution, then used the filtered image to make a decision about whether a lesion was present or absent. From this perspective, spatiotemporal Wiener filtering is the strategy of the ideal observer, the observer with maximum possible detection performance given full knowledge of the signal likelihoods. Wiener filtering reduces to matched filtering in low signal-to-noise ratio (SNR) conditions or with significant regularization. Spatiotemporal deconvolution methods have been well studied in the literature.^{3–7}

Matched rather than Wiener filtering is discussed in this article for simplicity. Spatiotemporal matched filtering involves time-reversal of the point-spread function followed by convolution with the rf image data. Spatiotemporal matched filtering has been investigated, for example, by Jensen and Gori,⁸ who proposed that focusing can be accomplished by spatial matched filtering, however, their experimental data

were acquired using a weakly focused mechanically scanned transducer, offering little improvement over standard imaging. They suggested using a more highly focused probe to see an image quality advantage. We use an array transducer with electronic focusing and investigate larger numerical aperture scanning. The time-reversal procedure in matched filtering also lends a connection to time-reversal literature.^{9,10}

Freeman *et al.*¹¹ proposed retrospective dynamic transmit focusing by deconvolving out-of-focus transmit regions with a scan angle-independent but depth-dependent filter. They applied their filter to dynamic-receive beamformed data. This approach was modified by Jeng and Huang¹² to account for depth-dependent SNR. While their work focused on correction of dynamic receive focused data, we concentrate on fixed focused beams. We build on the work of Li and Li,¹³ who showed a one-dimensional lateral filter for filtering fixed focus wave fronts to improve point-spread function compactness. They showed that filtering techniques with fixed receive focusing can achieve an image quality similar to that of dynamic receive focusing with filtering, a potential advantage for developing low complexity systems.

A number of authors have developed synthetic aperture approaches to accomplish transmit focusing. Nikolov¹⁴ presented an echo SNR-improving technique for synthetic transmit-receive focusing that used a virtual source “behind” an array. This technique allows a greater subaperture to be used for transmission, thus improving transmitted signal power. Additionally, Passman and Ermert¹⁵ and Frazier and O’Brien¹⁶ use a synthetic aperture method for single element transducers, treating the focal region as a virtual source. Our article contains a dynamic focusing extension of their work adapted for array transducers.

The novel contributions in this article include the following: (1) development of an analytic framework for understanding spatial bandwidth in unfocused regions of focused

^{a)}Electronic mail: mfi@uiuc.edu

beams. The theory predicts that the spatial bandwidth is approximately conserved throughout the nearfield, farfield, and focal zone of a fixed focus transducer. (2) We provide a useful approximation of the spatial bandwidth of a beam when the transmit and receive foci differ. (3) We describe a delay-and-sum algorithm for *dynamically* refocusing signals from unfocused regions of fixed-focus ultrasound beams. Our algorithm is an array-based shift-variant extension of the virtual source-detector synthetic aperture method^{15,16} for single element transducers. Some advantages of the technique compared to dynamic receive focusing are discussed. (4) We show that by focusing in the nearfield of a fixed unfocused aperture, the beam can be focused retrospectively as if it were transmitted from a low f -number transducer. One practical motivation behind fixed focusing rather than dynamic-receive focusing is that front-end ultrasound system complexity is greatly reduced when no dynamic beamforming circuit is required. One possible ultrasound array system architecture is discussed that would require simply fixed-delay analog delay lines and a single analog-to-digital converter. Another application is for improving pre- and postfocal-zone image quality in systems using mechanically scanned single element transducers.

II. THEORY

A. Unfocused regions of focused beams

We first concentrate on writing the equations of curved wave fronts from focused transducers. Rather than consider the details of individual elements in an array, we model the array as a continuous aperture with defined focusing and apodization properties. Our goal is to derive expressions for the lateral bandwidth of a transducer at axial depths at and away from the focus.

1. Linear systems model

For a single A-scan line, the spatiotemporal impulse response can be written as a series of temporal convolutions,¹⁷

$$h(\mathbf{x}, t) = -\frac{1}{c^2} \frac{\partial^2}{\partial t^2} \times \left\{ h_y(t) * h_y(t) * v(t) * \frac{\partial h_{Tx}(\mathbf{x}, t)}{\partial t} * h_{Rx}(\mathbf{x}, t) \right\}, \quad (1)$$

where h_y is the electromechanical impulse response of the transducer, $v(t)$ is the excitation voltage, and h_a (where “a” represents the transmit “Tx” or receive “Rx” aperture) is the acoustic impulse response of the transducer given by the Rayleigh integral

$$h_a(\mathbf{x}, t) = \frac{1}{2\pi} \int_S dS \xi(r) \frac{\delta(t - |\mathbf{r} - \mathbf{x}|/c)}{|\mathbf{r} - \mathbf{x}|}. \quad (2)$$

Here ξ is the transducer apodization function, and the vector \mathbf{r} defines points on the surface of the transducer S .

2. Spatial frequency domain

As we are interested in the spatial and temporal bandwidth characteristics with various focusing configurations, we need to calculate the Fourier magnitude of the point-spread function as a function of \mathbf{u} (the spatial frequency vector conjugate to \mathbf{x}) and f (the temporal frequency conjugate to time t). To obtain the two-dimensional Fourier magnitude $|H(u_2, f)|$ of $\text{psf}(x_2, t) = h(x_2, t|x_1, x_3)$ we need the Fourier transform of $\partial h_{Tx}(\mathbf{x}, t) / \partial t * h_{Rx}(\mathbf{x}, t)$, given as

$$i2\pi f [H_{Tx}(u_2, f|x_1, x_3) * H_{Rx}(u_2, f|x_1, x_3)] \quad (3)$$

where f is the temporal frequency, $H_a(u_2, f|x_1, x_3) = \mathcal{F}_{x_2, t}\{h_a(\mathbf{x}, f)\}$, and all other quantities are temporal Fourier transforms of the quantities in Eq. (1). The convolution is strictly over spatial frequencies u_2 .

3. Fresnel approximation

We begin our computation of $H_a(u_2, f|x_1, x_3)$ for the transmit (or receive) aperture by using the Fresnel approximation¹⁸ to compute

$$h_a(\mathbf{x}, f) \cong \frac{e^{ikx_1}}{i\lambda x_1} e^{i(k/2x_1)(x_2^2 + x_3^2)} \int dS \xi(\mathbf{r}) e^{i(k/2x_1)|\mathbf{r}|^2} e^{-i(k/x_1)\mathbf{x} \cdot \mathbf{r}} \quad (4)$$

The Fresnel approximation is applicable for quasiplanar apertures and for paraxial points far from the aperture with respect to the aperture dimensions.

4. Lateral bandwidth of Gaussian apodized fixed focus transducer

In this section we consider Gaussian apodized transducers having identical transmit and receive focus. Without the assumption of Gaussian apodization the analysis is less transparent. Let us consider a separable complex apodization function $\xi(r_2, r_3) = \xi_2(r_2)\xi_3(r_3)$ for quasiplanar transducers, where the azimuthal apodization functions can be considered a product of a real Gaussian apodization and a complex phase term representing focusing:

$$\xi_2(r_2) = e^{-r_2^2/2\sigma_2^2} e^{-jk[\sqrt{F^2 + r_2^2} - F]} \quad (5)$$

where F is the focal length. For analytical convenience we do not impose any finite aperture—we simply assume that the Gaussian apodization is not severely truncated, i.e., that the aperture width L is significantly greater than σ_2 . Additionally we find it advantageous to assume parabolic focusing by expanding the complex argument of Eq. (5) in a second-order Taylor series expansion in r_2 about 0 so that $\sqrt{F^2 + r_2^2} - F \cong r_2^2/(2F)$. With these approximations, Eq. (4) can be integrated by completing the square to become

$$h_a(\mathbf{x}, f) \cong \frac{k}{j2\pi x_1} e^{jkx_1} \sqrt{2\pi\sigma_3^2} \sqrt{\frac{\pi}{a_2}} e^{-\Psi x_2^2} \quad (6)$$

where

$$\frac{1}{a_2} = \frac{\frac{1}{2\sigma_2^2} - j\frac{k}{2}\left(\frac{1}{F} - \frac{1}{x_1}\right)}{\left(\frac{1}{2\sigma_2^2}\right)^2 + \left(\frac{k}{2}\right)^2\left(\frac{1}{F} - \frac{1}{x_1}\right)^2}. \quad (7)$$

Ψ is a complex quantity $\Psi = \Psi_r + j\Psi_i$, with real and imaginary parts given as

$$\Psi_r = \left(\frac{k}{2x_1}\right)^2 \frac{1/2\sigma_2^2}{\left(\frac{1}{2\sigma_2^2}\right)^2 + \left(\frac{k}{2}\right)^2\left(\frac{1}{F} - \frac{1}{x_1}\right)^2} \cong \frac{1}{2\sigma_2^2} \frac{F^2}{(F - x_1)^2} \quad (8)$$

and

$$\begin{aligned} \Psi_i &= \left(\frac{k}{2x_1}\right)^2 \frac{\left(\frac{k}{2}\right)\left(\frac{1}{F} - \frac{1}{x_1}\right)}{\left(\frac{1}{2\sigma_2^2}\right)^2 + \left(\frac{k}{2}\right)^2\left(\frac{1}{F} - \frac{1}{x_1}\right)^2} + \frac{k}{2x_1} \\ &\cong \left(\frac{k}{2}\right)\left(\frac{1}{x_1 - F}\right), \end{aligned} \quad (9)$$

respectively. The approximation is true for points x_1 away from the focus when $k|x_1 - F| \gg (F/\sigma_2)(x_1/\sigma_2)$. For F -numbers greater than 1 and for distances x_1 within a couple of focal lengths from the origin, this approximation means that the distance from the focus should be more than a few wavelengths.

The real part defines a lateral Gaussian envelope $e^{-\Psi_r x_2^2}$ for h_a , and the imaginary part defines a linear spatial frequency phase modulation $e^{-i(\Psi_i x_2) x_2}$, i.e., a baseband chirp, where Ψ_i determines the linear chirp rate and increases in magnitude as one approaches the focal region. The extent of chirping is regulated by Ψ_r as this determines the spatial extent of the curved wave fronts.

We need to transform our result to the spatial frequency domain to compute the lateral bandwidth (BW):

$$H_a(u_2, f) \propto \mathcal{F}\{e^{-\Psi_r x_2^2}\} \propto e^{-\Sigma u_2^2} \quad (10)$$

where

$$\text{Re}\{\Sigma\} = \pi^2 \frac{\Psi_r}{\Psi_r^2 + \Psi_i^2} \cong \pi^2 \frac{2F^2}{\sigma_2^2 k^2}. \quad (11)$$

The middle term is valid for the full expressions of Ψ_r and Ψ_i . The approximation is valid at the focus when $2\pi(\sigma_2/\lambda)(\sigma_2/F) \gg 1$ (which is practically always the case), and for points x_1 away from the focus when $k|x_1 - F| \gg (F/\sigma_2)(x_1/\sigma_2)$. Equation (10) can be substituted into Eq. (3) to find

$$|H(u_2, f; x_1, x_3)| \propto e^{-2 \text{Re}\{\Sigma\} u_2^2} \quad (12)$$

which means that the lateral Gaussian BW for all axial locations x_1 (that satisfy the approximations of the model) can be written as

$$\text{BW}_{\text{lat}} = \frac{2}{\sqrt{\text{Re}\{\Sigma\}}} = \frac{\sigma_2 k}{F}. \quad (13)$$

Note that the result is constant for nearfield, farfield, and focal axial depths and equal to the reciprocal of the expected focal resolution! This is a rather remarkable result which is validated for arrays using FIELD II simulations^{19,20} in Sec. III B.

5. Extension of the concept of time-BW product

The product of the time-duration of a coded wave form with its bandwidth is termed the time-bandwidth product (TBP). The TBP is a unitless quantity that is representative of potential information, and is one for typical pulses and greater than one for coded wave forms. It is appropriate to extend the TBP concept to spatial coding. Here we define a quantity which we shall call the lateral space-BW product (SBP_{lat}) which is given as the product of the lateral spatial extent of the psf times the lateral BW of the psf. For our Gaussian apodized transducer

$$\text{SBP}_{\text{lat}} = \text{BW}_{\text{lat}} \sigma_{\text{lat}} \cong 1 + \left(\frac{k\sigma_2^2 |F - x_1|}{F^2}\right)^2. \quad (14)$$

Note that σ_2 is the Gaussian aperture apodization width, whereas σ_{lat} is defined as the -6 dB lateral width of the psf. As might be expected, the SBP_{lat} is one at the focus (no wave front curvature thus no lateral coding). It is greater than 1 away from the focus and is greater for distances far from the focus, the expression holding as long as the Fresnel approximation is obeyed.

6. Differing transmit and receive foci

It can be readily shown that the lateral psf bandwidth due to transmit and receive foci F_{Tx} and F_{Rx} is given as

$$\text{BW}_{\text{lat}} = \frac{\sqrt{2}k}{\sqrt{\left(\frac{F_{\text{Tx}}}{\sigma_{\text{Tx}}}\right)^2 + \left(\frac{F_{\text{Rx}}}{\sigma_{\text{Rx}}}\right)^2}}, \quad (15)$$

where σ_{Tx} and σ_{Rx} are the transmit and receive aperture Gaussian apodization parameters, hence $F_{\text{Tx}}/\sigma_{\text{Tx}}$ and $F_{\text{Rx}}/\sigma_{\text{Rx}}$ are the transmit and receive f -numbers, respectively. Again this expression is approximately true even in the pre- and postfocal regions.

B. Time domain

We are interested in

$$\tilde{h}_i(\mathbf{x}, t) \equiv \frac{\partial h_a(\mathbf{x}, t)}{\partial t} *_i h_a(\mathbf{x}, t). \quad (16)$$

To compute this, consider the temporal frequency domain expression

$$\tilde{h}_i(\mathbf{x}, f) = jkc \times h_a^2(\mathbf{x}, f) \cong -\frac{jkc}{2} \left(\frac{k}{x_1}\right)^2 e^{j2kx_1} \sigma_3^2 \frac{1}{a_2} e^{-2\psi x_2^2}, \quad (17)$$

where σ_3 is the elevational Gaussian apodization parameter. Before taking the inverse temporal Fourier transform of this, note that the real part of Eq. (7) is a Lorentzian function of k , and thus has an inverse temporal Fourier transform of the form $e^{-a|\tau|}$. The imaginary part also looks like a Lorentzian but has an additional factor of jk in the numerator corresponding to a time derivative in the temporal domain.

When the rightmost term in the denominator of Eq. (7) dominates, the approximation of neglecting $(1/2\sigma_2^2)^2$ is useful because the k^2 in the denominator cancels with a k^2 in the numerator of Eq. (17)—simplifying the analysis. This can be written as

$$\tilde{h}_i(\mathbf{x}, f) \cong -\frac{\sigma_3^2}{2} \left(\frac{1}{x_1}\right)^2 \frac{1}{\left(\frac{1}{F} - \frac{1}{x_1}\right)^2} \exp\left[-\frac{1}{\sigma_2^2} \frac{F^2}{(F-x_1)^2} x_2^2\right] \times j2\pi f e^{j2\pi f \tau_F} \left[\frac{1}{\sigma_2^2} - j\frac{2\pi f}{c} \left(\frac{1}{F} - \frac{1}{x_1}\right)\right] \quad (18)$$

where

$$\tau_F = \frac{2x_1}{c} - \frac{1}{x_1 - F} \frac{x_2^2}{c}. \quad (19)$$

Now proceeding with the inverse temporal Fourier transform, we have

$$h_i(\mathbf{x}, t) \cong -\frac{\sigma_3^2}{2} \left(\frac{1}{x_1}\right)^2 \frac{1}{\left(\frac{1}{F} - \frac{1}{x_1}\right)^2} \exp\left[-\frac{1}{\sigma_2^2} \frac{F^2}{(F-x_1)^2} x_2^2\right] \times \left[\frac{1}{\sigma_2^2} - \frac{1}{c} \left(\frac{1}{F} - \frac{1}{x_1}\right) \frac{d}{dt}\right] \frac{d}{dt} \delta(t - \tau_F). \quad (20)$$

We may apply the temporal derivatives to the excitation or electromechanical coupling responses $h_{\text{pulse}}(t) \equiv h_y(t) * h_v(t) * v(t)$. In this way, the time delay for the system impulse response is τ_F :

$$h(\mathbf{x}, t) = p(\mathbf{x}, t) * \delta(t - \tau_F(\mathbf{x})) \quad (21)$$

where $*$ is a temporal convolution and

$$p(\mathbf{x}, t) = -\frac{\sigma_3^2}{2} \left(\frac{1}{x_1}\right)^2 \frac{1}{\left(\frac{1}{F} - \frac{1}{x_1}\right)^2} \exp\left[-\frac{1}{\sigma_2^2} \frac{F^2}{(F-x_1)^2} x_2^2\right] \times \left[\frac{1}{\sigma_2^2} - \frac{1}{c} \left(\frac{1}{F} - \frac{1}{x_1}\right) \frac{d}{dt}\right] \frac{d}{dt} h_{\text{pulse}}(t). \quad (22)$$

This time-delay factor can help us reduce the spatial matched filtering operation for image reconstruction to a delay and sum procedure.

C. Spatiotemporal filtering to recover spatial resolution

While spatial bandwidth is a measure of the spatial resolution encoded in a point-spread function, spatiotemporal filtering is required to recover this resolution in unfocused regions of the beam. By time-reversing the point-spread function at a given depth and convolving it with the fixed focus beamformed rf data it is possible to improve lateral spatial resolution, as discussed in previous work, and illustrated in the experimental section of this article. A similar effect can be produced by delay and sum postprocessing.

D. Retrospective delay-and-sum dynamic focusing

Noting that the impulse response of a Gaussian apodized focused aperture can be written as Eq. (21) the spatiotemporal matched filtering procedure involves as its principle operation, convolution with the delta function $\delta(t - \tau_F(x))$, which motivates the delay-and-sum reconstruction

$$y(x_2, t|x_1) = \int g(x_2 - x_2', t - \tau_F(x_1, x_2 - x_2')|x_1) dx_2'. \quad (23)$$

The above-presented delay-and-sum procedure can be extended to discrete scan lines $g_n(t)$ and shift-varying dynamic focusing by considering that $x_1 = ct/2$ in expression (19) for τ_F , and ignoring the linear propagation term $2x_1/c$. The m th reconstructed scan line as a function of time is then given as

$$y_m(t) = \sum_n w_n(t) g_{m-n} \left(t + \frac{1}{ct/2 - F} \frac{\Delta x_n^2}{c} \right), \quad (24)$$

where Δx_n is the distance from the center of the walking subaperture to the n th array element, and $w_n(t)$ is a time-dependent aperture weighting function. The aperture should shrink the closer one gets to the focal zone, especially when $k|x_1 - F|$ is not much larger than $(F/\sigma_2)(x_1/\sigma_2)$, as discussed earlier.

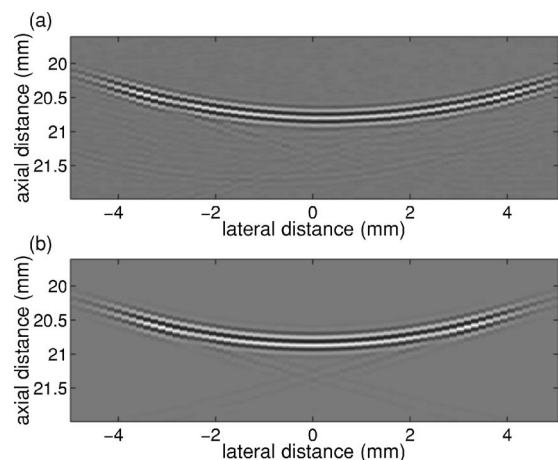


FIG. 1. (a) Measured nearfield psf and (b) simulated psf. We used a VF10-5 array transducer with the following parameters: F -number of 2.1, transmit focus of 4 cm, receive focal distance at 3.9 cm, elevation focus approximately 2 cm, and 6.67 MHz excitation frequency. The array had 192 elements of dimension 0.2×5 mm with 0.02 mm kerf.

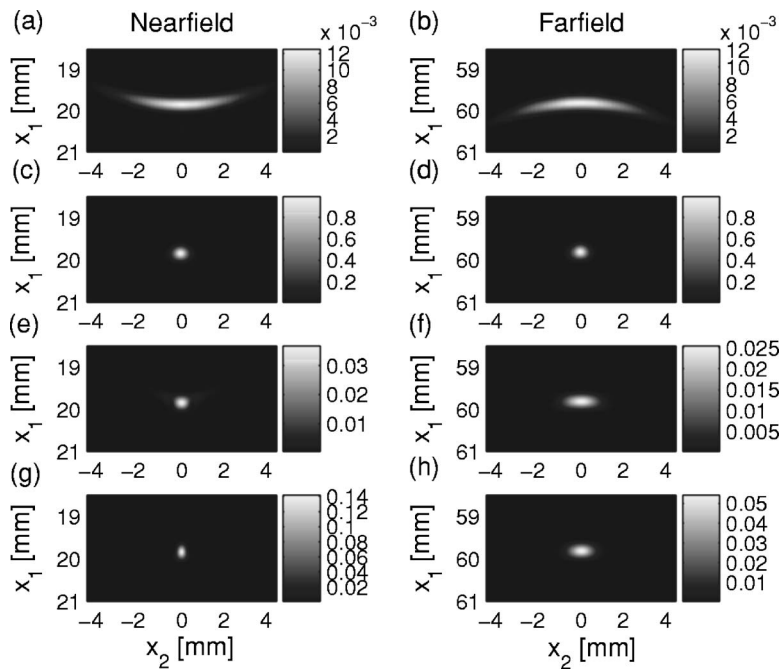


FIG. 2. Comparison of psfs: (a) and (b) Near- and farfield psf due to a Gaussian-apodized ($\sigma=6.7$ mm) subaperture truncated at width 2 cm, and with a 4 cm transmit-receive focus. (c) and (d) psf due to retrospective dynamic focusing of the psfs in (a) and (b). (e) and (f) psfs due to dynamic receive focusing with 4 cm transmit focus and same transmit aperture. (g) Nearfield psf due to focusing at 20 mm depth on transmit and receive using the above-mentioned aperture. (h) Farfield psfs due to focusing at 6 cm on transmit and receive using the above-mentioned aperture.

E. The virtual source, virtual detector interpretation

The delay function can be derived from simple geometric considerations by understanding that the transmit focal points act as an array of virtual sources. Similarly, the receive focal points act as arrays of virtual detectors. In the fixed focus paradigm, the virtual sources and virtual detectors are spatially identical. Time delays $2x_1/c - (1/x_1 - F) \times (\Delta x_n^2/c)$ represent a Taylor expansion in Δx_n to the hypotenuse $2\sqrt{(\Delta x_n - F)^2 - \Delta x_n^2/c}$ of the triangle whose vertices are the walking subaperture center, field point, and n th virtual element a distance d_n from the subaperture center. Equation (24) thus may be interpreted as a process of applying dynamic time delays to the pulse-echo signals of the virtual array elements to dynamically focus along the desired scan lines. This interpretation helps us consider more general scanning and beamsteering geometries.

III. SIMULATIONS AND EXPERIMENTS

We use simulations and experiments to test some of our ideas. For experiments, we used a programmable Siemens Sonoline Antares ultrasound scanner. This scanner possesses a commercially available ultrasound research interface (URI)

that allows us to control acquisition parameters not accessible in clinical mode, and to save beamformed rf to files for offline analysis. A library of MATLAB functions (offline processing tool or OPT) for reading and processing the data was available to us to assist in our analysis.

A. Nearfield point-spread functions: Experimental validation of simulations

We use FIELD II to simulate psfs to compare with measured psfs from the Antares. This gives us confidence that our simulations are realistic, and allows us the flexibility to try more settings than are allowable with the current URI. To measure the nearfield psf of a fixed-focus beam on the Antares, we used the URI to turn off dynamic receive, aperture growth, and receive apodization functions. We set the receive F -number to 2.1, the transmit focus at 4 cm, and the receive focus at 3.9 cm (the URI allows only several discrete choices for these parameters). To image psfs we simply acquired rf data from sparse dust particles in de-gassed water. The measured and corresponding simulated psf are shown in Fig. 1. The curved wave front of the simulated psf is similar to the measurement.

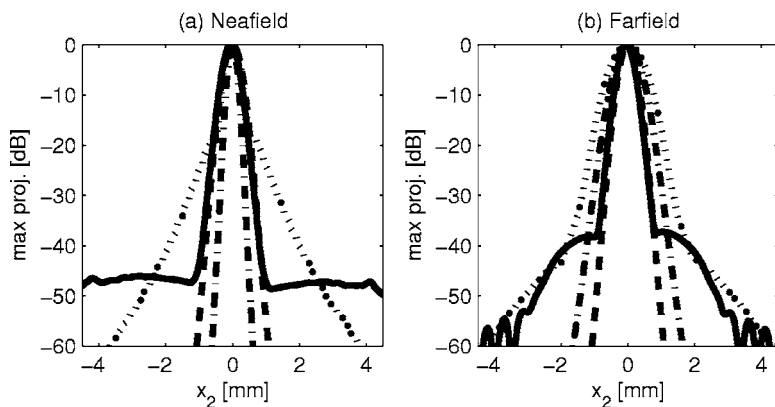


FIG. 3. Normalized cross-range maximum amplitude projections of near (left) and farfield (right) psfs. Solid line: retrospective dynamic focusing; dotted line: the projection of the dynamic receive focusing psf in Figs. 2(e) and 2(f); dashed line: projection of the psf due to a scatterer placed at the 4 cm transmit-receive focus; dot-dashed line: projection of the focused psfs (g) and (h).

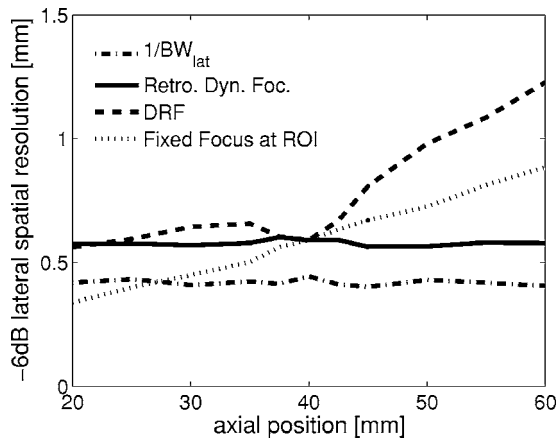


FIG. 4. Full width at half maximum of Gaussian-apodized psfs as a function of axial distance. Solid line: Retrospective dynamic focusing of psfs due to scatterers placed at various depths, using a 2 cm Gaussian-apodized aperture ($\sigma=6.7$ mm) focused at 4 cm (transmit and receive); dashed line: resolution curve due to dynamic receive focusing and 4 cm transmit focus; dot-dashed line: the reciprocal of the lateral bandwidth of fixed 4 cm transmit-receive focus psfs as measured by the full width at half maximum of the two-dimensional power spectral density at the axial center frequency; dotted line: the -6 dB beamwidth of a transducer with fixed focus at the plotted axial depth (i.e., focused in the region of interest).

B. Gaussian-apodized psfs

Having established the accuracy of the simulations in modeling our ultrasound system, we now simulate Gaussian-apodized beams to validate some of the theoretical predictions, and to test the performance of the reconstruction algorithm. Figure 2 shows Gaussian-apodized near- and farfield point-spread functions, including retrospectively focused psf results in Figs. 2(c) and 2(d). The retrospectively filtered psfs have approximately the same lateral spatial resolution as the focal region psf, as shown in Fig. 3 and predicted by Eq. (13). This is also seen in Fig. 4, which shows that the lateral bandwidth is approximately constant through the near- and

farfield, and that retrospective dynamic focusing is able to sustain focal-zone lateral spatial resolution through the near- and farfield regions. In the farfield, the spatial resolution attainable with retrospective focusing is finer than that attainable with dynamic receive focusing and even finer than that attainable when one focuses (transmit and receive) at the region of interest. This means that for a given aperture size, by focusing before rather than at the farfield region, spatial resolution a few aperture lengths away from the transducer can be retrospectively focused with a spatial resolution equivalent to a low $F\#$ psf. This conclusion may have important implications for applications with limited aperture. Although sidelobes due to retrospective dynamic focusing are nonideal, they are lower than dynamic receive focusing within a couple of millimeters of the mainlobe, and sufficiently low beyond this. The retrospective focusing method further offers enhanced signal-to-noise compared to dynamic receive focusing as evidenced by the grayscale magnitudes in Fig. 2, where each column of psfs is normalized by the maximum of the retrospective focusing image.

C. Comparison with dynamic receive focusing: Unapodized apertures

Here we consider unapodized apertures. We use computer simulations since analytic tractability is more challenging for this case. Figures 5 and 6 show that in the nearfield, the retrospective dynamic focusing method shows comparable spatial resolution to focusing at the region of interest (for a fixed aperture size). This is slightly counterintuitive since Eq. (13) predicted that the spatial bandwidth for a 2 cm transmit focus should be greater than that for a 4 cm transmit focus for a fixed aperture. Computations in Fig. 7 show that the prediction of constant lateral bandwidth with axial depth must be rethought for non-Gaussian apodizations. In fact, it appears that the lateral bandwidth at the 4 cm focus is mini-

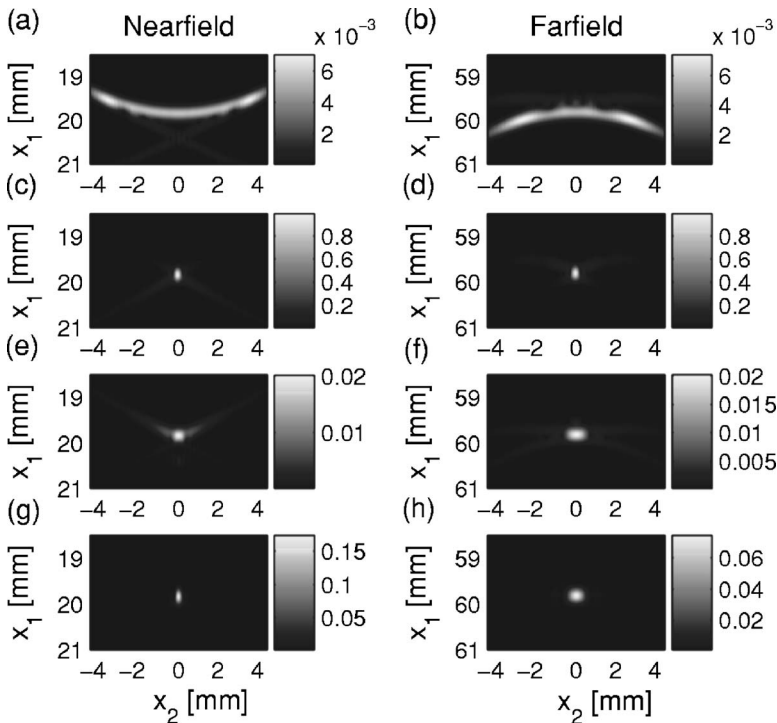


FIG. 5. Comparison of psfs: (a) and (b) Near- and farfield psf due to an unapodized 2 cm walking subaperture, and with a 4 cm transmit-receive focus. (c) and (d) psf due to retrospective dynamic focusing of the psfs in (a) and (b). (e) and (f) psfs due to dynamic receive focusing with 4 cm transmit focus and same transmit aperture. (g) Nearfield psf due to focusing at 20 mm depth on transmit and receive using the above-mentioned aperture. (h) Farfield psfs due to focusing at 6 cm on transmit and receive using the above-mentioned aperture.

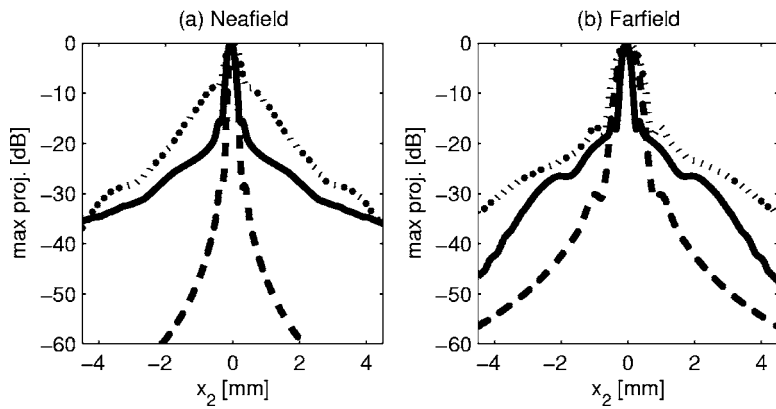


FIG. 6. Normalized cross-range maximum amplitude projections of near (left) and farfield (right) psfs. Solid line: retrospective dynamic focusing; dotted line: the projection of the dynamic receive focusing psf in Figs. 5(e) and 5(f); dashed line: projection of the psf due to a scatterer placed at the 4 cm transmit-receive focus; dot-dashed line: projection of the focused psfs (g) and (h).

num across the axial range. This roughly explains the retrospective dynamic focusing curve in Fig. 7. In the farfield regions, retrospective dynamic focusing is seen to offer substantial spatial resolution and sidelobe improvements over dynamic receive focusing. Again, similar to the Gaussian apodization case, we see that by focusing before rather than at the farfield region, spatial resolution a few aperture lengths away from the transducer can be retrospectively focused with a spatial resolution equivalent to a low $F\#$ psf. The retrospective focusing method again offers enhanced signal-to-noise compared to dynamic receive focusing (and even focused imaging).

D. Phantom experiments

The Siemens Antares was used for phantom experiments. Figure 8 shows a retrospectively focused image of anechoic inclusions in a scattering phantom compared to the image obtained using dynamic receive focusing in the nearfield of an $F/2.1$ linear array transmitting 6.67 MHz broadband pulses. Improvements in resolution and SNR are

apparent visually from the image and from contrast (C) and contrast-to-noise (CNR) estimates in Table I. Here contrast is defined as

$$C = \frac{\mu_i - \mu_b}{\mu_b}, \quad (25)$$

where μ_i and μ_b are the mean envelope-detected signal levels in the inclusion and in the background, respectively. Contrast-to-noise is defined as

$$CNR = \frac{|\mu_i - \mu_b|}{\sqrt{\frac{1}{2}(\sigma_i^2 + \sigma_b^2)}}, \quad (26)$$

where σ_i and σ_b are the standard deviations of the envelope-detected signal in the inclusion and in the background, respectively. Both lesions were imaged in the nearfield at a depth of 2 cm, and in both cases the transmit focus was set at 4 cm. The F -number in (a) was 2.1. The experimental results are consistent with predictions of improved spatial resolution discussed earlier.

IV. DISCUSSION

One practical motivation behind fixed focusing rather than dynamic-receive focusing is that front-end ultrasound system complexity is greatly reduced when no dynamic beamforming circuit is required.

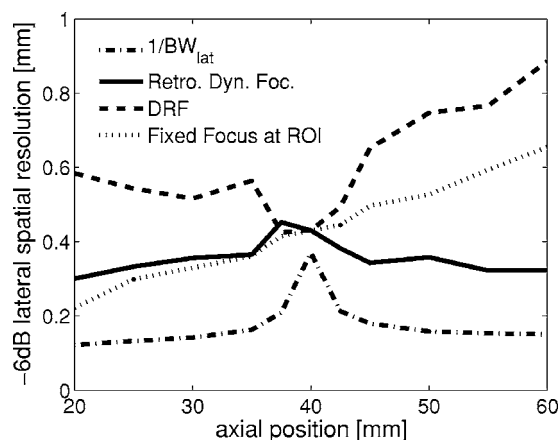


FIG. 7. Full width at half maximum of unapodized psfs as a function of axial distance. Solid line: Retrospective dynamic focusing of psfs due to scatterers placed at various depths, using a 2 cm unapodized aperture focused at 4 cm (transmit and receive); dashed line: resolution curve due to dynamic receive focusing and 4 cm transmit focus; dot-dashed line: the reciprocal of the lateral bandwidth of fixed 4 cm transmit-receive focus psfs as measured by the full width at half maximum of the two-dimensional power spectral density at the axial center frequency; Dotted line: the -6 dB beamwidth of a transducer with fixed focus at the plotted axial depth (i.e., focused in the region of interest).

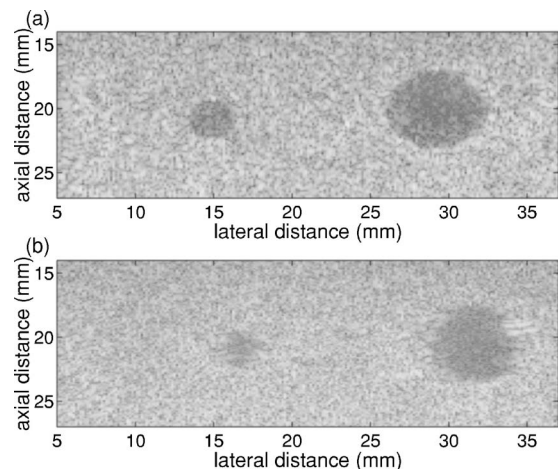


FIG. 8. Anechoic lesion phantom: (a) Nearfield retrospectively focused image, $F\#=2.1$. (b) Dynamic receive focusing image. In both images the transmit focus is at 4 cm.

TABLE I. Anechoic lesions: Comparison of retrospectively focused nearfield imaging with dynamic receive focusing.

Method	C	CNR
	Large lesion	
Retro. Foc.	-0.75	1.5
DRF	-0.61	1.0
	Small Lesion	
Retro. Foc.	-0.71	1.4
DRF	-0.45	0.8

One possible architecture could use an analog switch array to route incoming channel data through fixed-delay analog delay lines followed by channel summation. This would eliminate tapped delay lines in analog dynamic-receive beamformers, which have stringent demands on tap intervals that are difficult to obtain over long delays. It also eliminates multiple analog-to-digital converters and fast first-in-first-out (FIFO) memory prevalent in digital systems, thus reducing cost and complexity. The option of switching between a few focal zones or using multiple focal zones in a scan could further enhance flexibility for image quality.

Another important application may be in high-frequency ultrasound systems using mechanically scanned high-numerical aperture single element transducers. The developments here may prove important for improving image quality away from the focal zone.

Potential artifacts and promising solutions. A known issue with spatiotemporal filtering is that sidelobe levels are higher than desired. This article's purpose is to provide a fundamental perspective imaging with unfocused regions of focused beams, hence we do not explicitly deal with sidelobe reduction strategies. Apodization methods may prove advantageous as may filtering channel data before beamforming, as discussed by Kim *et al.*²¹ This may be a topic of future research.

Motion within the scan duration could produce significant artifacts using unfocused imaging techniques. Some groups, however, have successfully implemented related synthetic aperture techniques *in vivo* at very high frame rates.²² If scan times are fast enough, motion artifacts may be minimal.

Phase aberrations could be another source of artifacts. Aberrations due to differences in the speed of sound in tissue or due to refractive tissue interfaces could shift energy away from the expected wave fronts. This, however, is also a problem for conventional focused imaging using dynamic receive focusing (DRF) methods. Moreover, deaberration strategies or filtering techniques may prove useful. Understanding these concepts better and applying them to *in vivo* imaging should be a topic of future work. A rigorous investigation is not within the scope of this article.

V. CONCLUSIONS

By studying the spatial bandwidth in Gaussian apodized point-spread functions, we conclude that spatial resolution encoded in curved wave fronts remains approximately constant throughout the nearfield, focal, and farfield zones. A

retrospective dynamic focusing approach successfully recovers this resolution, and may provide a solution to the problem of loss of spatial resolution away from the transmit focus in present state-of-the art dynamic receive focusing array systems.

ACKNOWLEDGMENT

We acknowledge partial support from NIH R01 CA 082497.

- ¹R. J. Zemp, M. D. Parry, C. K. Abbey, and M. F. Insana, "Detection performance theory for ultrasound imaging systems," *IEEE Trans. Med. Imaging* **24**, 300–310 (2005).
- ²C. K. Abbey, R. J. Zemp, and M. F. Insana, "Pre-envelope deconvolution for increased lesion detection efficiency in ultrasonic imaging," *Proc. SPIE* **5034**, 280–288 (2003).
- ³T. S. F. Lingvall and T. Olofsson, "Synthetic aperture imaging using sources with finite aperture-deconvolution of the spatial impulse response," *J. Acoust. Soc. Am.* **114**, 225–234 (2003).
- ⁴K. Ranganathan and W. F. Walker, "A novel beamformer design method for medical ultrasound. I. Theory," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50**, 15–24 (2003).
- ⁵J. Jensen, J. Mathorne, T. Gravesen, and B. Stage, "Deconvolution of *in-vivo* ultrasound B-mode images," *Ultrason. Imaging* **15**, 122–133 (1993).
- ⁶C. K. Abbey, R. J. Zemp, J. Liu, K. K. Lindfors, and M. F. Insana, "Observer efficiency in discrimination tasks simulating malignant and benign breast lesions with ultrasound," *IEEE Trans. Med. Imaging* **25**, 198–209 (2006).
- ⁷B. Haider, P. Lewin, and K. E. Thomenius, "Pulse elongation and deconvolution filtering for medical ultrasonic imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45**, 98–113 (1998).
- ⁸J. A. Jensen and P. Gori, "Spatial filters for focusing ultrasound images," *Proceedings of the IEEE Ultrasonics Symposium*, Atlanta, GA, 2001, pp. 1507–1511.
- ⁹M. Fink, "Time reversal of ultrasonic fields. I. Basic principles," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **39**, 555–566 (1992).
- ¹⁰G. S. Kino, *Acoustic Waves: Devices, Imaging and Analog Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1987).
- ¹¹S. Freeman, P. C. Li, and M. O'Donnell, "Retrospective dynamic transmit focusing," *Ultrason. Imaging* **17**, 173–196 (1995).
- ¹²G.-S. Jeng and S.-W. Huang, "SNR-dependent filter design for improving depth of field using modified least squares," *Proceedings of the IEEE Ultrasonics Symposium*, Honolulu, Hawaii, 2003, pp. 1235–1238.
- ¹³M. L. Li and P. C. Li, "Filter based synthetic transmit and receive focusing," *Ultrason. Imaging* **23**, 73–89 (2001).
- ¹⁴S. Nikolov and J. A. Jensen, "Virtual ultrasound sources in high-resolution ultrasound imaging," *Proc. SPIE* **3**, 395–405 (2002).
- ¹⁵C. Passman and H. Ermert, "A 100 MHz ultrasound imaging system for dermatologic and ophthalmologic diagnostics," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **43**, 545–552 (1996).
- ¹⁶C. Frazier and W. O'Brien, "Synthetic aperture techniques with a virtual source element," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45**, 196–207 (1998).
- ¹⁷P. R. Stepanishen, "Acoustic transients from planar axisymmetric vibrators using the impulse response approach," *J. Acoust. Soc. Am.* **70**, 1176–1181 (1981).
- ¹⁸J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York) (1996).
- ¹⁹R. Zemp, "Detection theory in ultrasonic imaging," Dissertation, University of California, Davis, (2004).
- ²⁰R. J. Zemp and M. F. Insana, "Spatial coding with curved wavefronts," in *Proceedings of the IEEE Ultrasonics Symposium*, Montreal, Quebec, Canada, 2004.
- ²¹K. S. Kim, J. Liu, and M. F. Insana, "Efficient array beam forming by spatial filtering for ultrasound B-mode imaging," *J. Acoust. Soc. Am.* **120**, 852–861 (2006).
- ²²J. A. Jensen *et al.*, "Ultrasound research scanner for real-time synthetic aperture data acquisition," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **52**, 881–891 (2005).

Spectrum analysis of seismic surface waves and its applications in seismic landmine detection

Mubashir Alam,^{a)} James H. McClellan,^{b)} and Waymond R. Scott, Jr.^{c)}

School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive, NW, Atlanta, Georgia 30332

(Received 19 January 2006; revised 24 July 2006; accepted 8 August 2006)

In geophysics, spectrum analysis of surface waves (SASW) refers to a noninvasive method for soil characterization. However, the term spectrum analysis can be used in a wider sense to mean a method for determining and identifying various modes of seismic surface waves and their properties such as velocity, polarization, etc. Surface waves travel along the free boundary of a medium and can be easily detected with a transducer placed on the free surface of the boundary. A new method based on vector processing of space-time data obtained from an array of triaxial sensors is proposed to produce high-resolution, multimodal spectra from surface waves. Then individual modes can be identified in the spectrum and reconstructed in the space-time domain; also, reflected waves can be separated easily from forward waves in the spectrum domain. This new SASW method can be used for detecting and locating landmines by analyzing the reflected waves for resonance. Processing examples are presented for numerically generated data, experimental data collected in a laboratory setting, and field data. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2346130]

PACS number(s): 43.60.Gk, 43.60.Fg, 43.60.Ac, 43.20.Gp [EJS]

Pages: 1499–1509

I. INTRODUCTION

Waves that propagate in a medium can be roughly divided into two main categories: body waves and surface waves. Surface waves are generated at a boundary and can be essentially of two types: Love waves and Rayleigh waves. Rayleigh waves are always generated when a free surface exists in a continuous body. In a vertically heterogeneous medium the phase velocity of the Rayleigh wave is a function of frequency and this dependence is strictly related to the mechanical parameters of the medium.^{1,2} Conversely, if we can determine the dispersion curve (i.e., phase velocity versus frequency), it is theoretically possible to calculate the mechanical parameters of the medium. This inversion technique is the basis of spectrum analysis of surface waves (SASW) methods. Traditional methods use data collected at two receivers from which the phase of the average cross-power spectrum is used to calculate the phase velocity.^{1,2} One crucial step in this process is unwrapping the cross power spectrum phase, because additive noise can produce fictitious phase jumps. For data from multiple receivers, array techniques have been developed based on two-dimensional (2D) frequency-wave-number analysis^{1,2} and slant-stack analysis followed by a Fourier-Bessel transform,³ but these use the Fourier transform so they have less spatial resolution than the method described in Sec. II.

The technique proposed here is based on the combination of a temporal Fourier transform across t followed by a pole-zero model across the spatial domain x . Using the complex amplitude and root estimates from pole-zero modeling, it is possible not only to construct dispersion curves, but also

to obtain insight into several other important parameters, e.g., attenuation and polarization. We have extended our algorithm to the two-channel case so that we can also estimate polarization. When the array data is collected by means of triaxial sensors, we use two channels that provide the horizontal and vertical particle motion (acceleration is actually measured) to derive a vector pole-zero model that estimates the complex amplitudes of the measurements in these two channels from which polarization ellipses are estimated. A surface wave consists of particle motion along a specific path, e.g., a Rayleigh wave involves particle motion along a retrograde elliptical path.^{1,4,5} We have successfully used polarization to identify these waves in field data as well a laboratory data.

Detection of buried landmines and subsurface structures has been investigated at Georgia Tech (Atlanta, GA) and elsewhere in recent years using seismic waves.^{5–8} A seismic wave is launched from a source, and it travels through the soil and interacts with targets. This interaction is recorded by means of a sensor array placed on the surface of the earth. The goal is to use this collected space-time data to localize the buried targets and, in some cases, to isolate the resonance of the target for better imaging and clutter discrimination. The location of the buried targets can be estimated from the reflected wave, if the reflected wave can be identified and separated from the forward probing wave. In this new SASW technique, it is possible not only to separate the forward wave from the reflected wave, but also to resynthesize both of them. Based on this wave separation, a simple imaging algorithm will be presented to locate buried landmines.

The next sections will describe the details of the parametric pole-zero model and the 2D spectrum analysis technique. The algorithm will be used to extract individual waves and their modes, to estimate polarization, and to derive a

^{a)}Electronic mail: ma@ece.gatech.edu

^{b)}Electronic mail: jim.mcclellan@ece.gatech.edu

^{c)}Electronic mail: waymond.scott@ece.gatech.edu

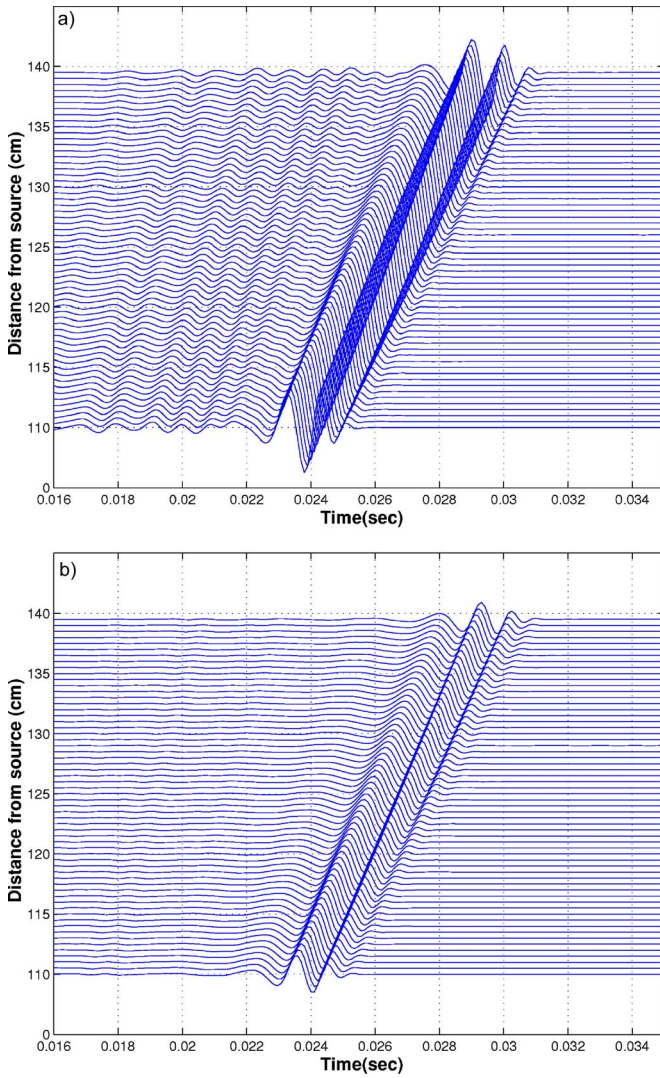


FIG. 1. (Color online) Space-time data for (a) horizontal channel $s_x(x,t)$ and (b) vertical channel $s_z(x,t)$.

new imaging method. The performance of algorithm will be demonstrated on numerical data and field data with landmines and clutter.

II. PARAMETRIC MODEL FOR SURFACE WAVES-VECTOR SENSOR APPROACH

The parametric model is based on a technique developed for borehole sonic logging applications.⁹ For the single channel case, the collected data $s(x,t)$ is a function of space and time, see Fig. 1. We can represent the data with its 2D Fourier transform

$$s(x,t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{S}(k,\omega) e^{i(kx-\omega t)} dk d\omega, \quad (1)$$

where x is the spatial position, k is the spatial wave number, and ω is the temporal frequency. If we only take a temporal Fourier transform across t , we have

$$S(x,\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{S}(k,\omega) e^{j(kx)} dk. \quad (2)$$

At each temporal frequency ω , we can imagine that we approximate the integral on the right-hand side of Eq. (2) with a sum of exponentials. This is carried out with pole-zero modeling across the spatial dimension to get a model consisting of a sum of exponentials that represents propagating waves

$$S(x,\omega) \approx \sum_{p=1}^P a_p(\omega) e^{jk_p(\omega)x}, \quad (3)$$

where P is the model order.

In the multichannel case, the collected data $s(x,t)$ is a vector. For the two-channel case, we have

$$\mathbf{s}(x,t) = \begin{bmatrix} s_x(x,t) \\ s_z(x,t) \end{bmatrix}, \quad (4)$$

where $s_x(x,t)$ is the horizontal displacement channel (along the direction of propagation), and $s_z(x,t)$ is the vertical displacement channel as in Fig. 1. If we do pole-zero modeling *separately for each channel*, then we obtain two independent P th-order exponential models like Eq. (3)

$$S_x(x,\omega) \approx \sum_{p=1}^P A_{xp}(\omega) e^{jk_{xp}(\omega)x}, \quad (5)$$

$$S_z(x,\omega) \approx \sum_{p=1}^P A_{zp}(\omega) e^{jk_{zp}(\omega)x}. \quad (6)$$

However, there is no guarantee that the wave-number information for the two channels, $k_{xp}(\omega)$ and $k_{zp}(\omega)$, will match in the $k-\omega$ domain. If not, it will be very hard to align the vertical and horizontal complex amplitudes and calculate polarization.

A. Vector IQML

A better approach is to determine the two models simultaneously so that they share the same poles.¹⁰ The pole-zero modeling technique used in this paper is based on the IQML (iterative quadratic maximum likelihood) algorithm which is also called the Steiglitz-McBride extension of Prony's method.¹¹ We have reformulated the IQML algorithm for the multichannel case.

For a plane wave impinging on m two-channel sensors, we can take temporal Fourier transforms of the collected data (4) to obtain

$$\underline{\mathbf{S}}(\omega) = [\underline{\mathbf{S}}_1(\omega), \underline{\mathbf{S}}_2(\omega), \dots, \underline{\mathbf{S}}_m(\omega)], \quad (7)$$

where

$$\underline{\mathbf{S}}_i(\omega) = \begin{bmatrix} S_x(x_i,\omega) \\ S_z(x_i,\omega) \end{bmatrix}. \quad (8)$$

The transform vector in Eq. (7) consists of the complex amplitudes from both channels at a specific frequency ω . The IQML technique determines P estimates for the poles which are the same for both channels, and then computes the best

complex amplitudes A_{xp} and A_{zp} , which are different for the two channels. The resulting exponential model from IQML is

$$\underline{\mathbf{S}}(\omega) \approx \begin{bmatrix} \sum_{p=1}^P A_{xp}(\omega) e^{j\kappa_p(\omega)x} \\ \sum_{p=1}^P A_{zp}(\omega) e^{j\kappa_p(\omega)x} \end{bmatrix}. \quad (9)$$

The poles determine the exponent $\kappa_p(\omega)$ whose real part is the wave-number $k_p(\omega)$ and whose imaginary part is the attenuation $\alpha_p(\omega)$.

Wave number is converted to velocity via $v_p(\omega) = \omega/k_p(\omega)$, and then we can plot the magnitude of A_{xp} , or A_{zp} , versus frequency and velocity as shown in Fig. 2(a). This type of plot makes it easy to obtain the dispersion curves of velocity versus ω for the various modes that make up the signal. The complex amplitudes are used to determine the strength of different wave components, but they can also be used to obtain the parameters for polarization ellipses as discussed next.

B. Polarization Ellipses

The complex amplitudes obtained from two channels can be used to draw polarization ellipses.¹² If the x and z components of the vector complex amplitude are the two complex amplitudes A_x and A_z , then

$$\bar{\mathbf{A}} = A_x \hat{x} + A_z \hat{z} = A_x (\hat{x} + \alpha e^{j\varphi} \hat{z}).$$

Various parameters of the polarization ellipse can be obtained directly from α and φ . The tilt angle ψ (with respect to the x axis) is given by

$$\tan(2\psi) = \frac{2\alpha}{1 - \alpha^2} \cos(\varphi). \quad (10)$$

If major and minor axes are denoted by a and b , then the axial ratio is

$$\frac{b}{a} = \tan(\beta), \quad (11)$$

where β is

$$\sin(2\beta) = \frac{2\alpha}{1 + \alpha^2} \sin(\varphi). \quad (12)$$

The major and minor axes of the ellipse are given by

$$a = \frac{1}{\sqrt{\alpha^2 \cos^2(\psi) + \sin^2(\psi) - 2\alpha \sin(\psi) \cos(\psi) \cos(\varphi)}}, \quad (13)$$

$$b = \frac{1}{\sqrt{\alpha^2 \sin^2(\psi) + \cos^2(\psi) + 2\alpha \sin(\psi) \cos(\psi) \cos(\varphi)}}. \quad (14)$$

In other words, the complex amplitudes derived via vector IQML modeling allow us to quickly calculate polarization parameters for the propagating waves.

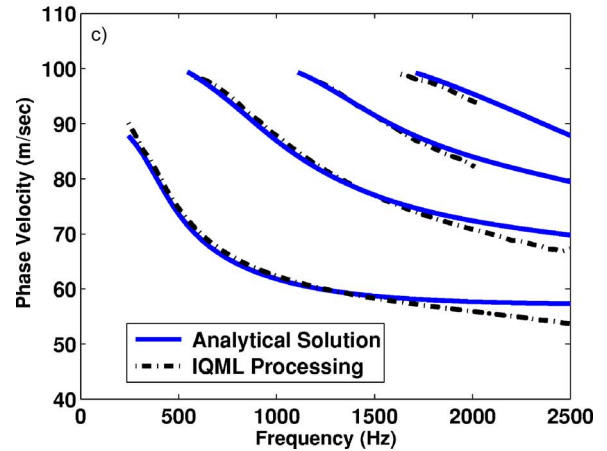
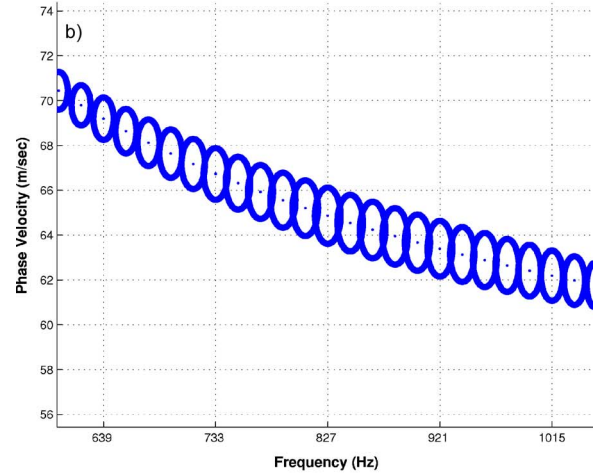
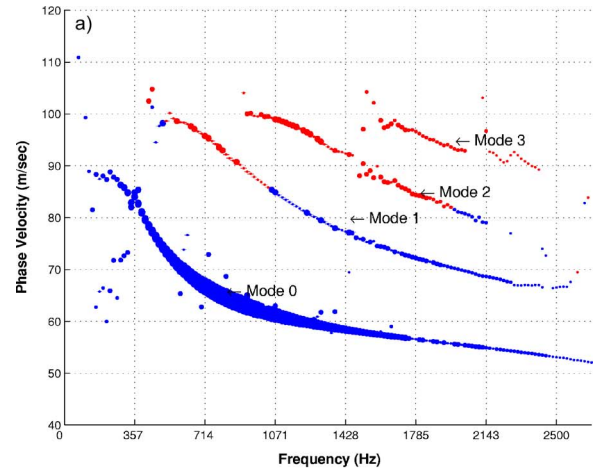


FIG. 2. (Color online) (a) Multimodal dispersion curves extracted with IQML using a fourth-order model ($P=4$), (b) expanded view of mode-0 showing polarization ellipses, and (c) comparison of analytical solution and IQML Processing.

III. PROCESSING DATA FROM LINEAR ARRAYS

A. Synthetic data

Numerical data generated from a 3D finite-difference time-domain (FDTD) model can accurately model elastic wave propagation in a stratified medium.⁵ The seismic source in the model is a vertically oriented point source excited with a differentiated Gaussian pulse with a center frequency of 450 Hz. The profile used for this numerical data is

TABLE I. Soil profile used in numerical data.

Thickness (cm)	V_s (m/s)	V_p (m/s)	Density (kg/m^3)
2.5	60	250	1400
2.5	70	250	1400
2.5	80	250	1400
2.5	90	250	1400
∞	100	250	1400

given in Table I. The data simulate what the sensors would have measured on the surface given the known stratified medium of the model. Examples of synthetic data for the horizontal and vertical channels are shown in Fig. 1, where the

horizontal axis is time and the vertical axis is sensor position. The first sensor lies 110 cm from the source and the intra-sensor distance is 0.5 cm. The total number of sensors is 60, covering an aperture of 30 cm.

Processing of this data set with the vector IQML algorithm and a model order of $P=4$ yields the dispersion curves shown in Fig. 2(a). These multimodal dispersion curves are typical for a layered media.^{1,5} Four different modes can be identified at the higher frequencies, with the strongest one being the Rayleigh wave (mode-0). Traditional two-station methods would only be able to detect the dominant Rayleigh mode. Modes 1 through 3 are guided waves that are due to the layered soil properties used in the model and are essen-

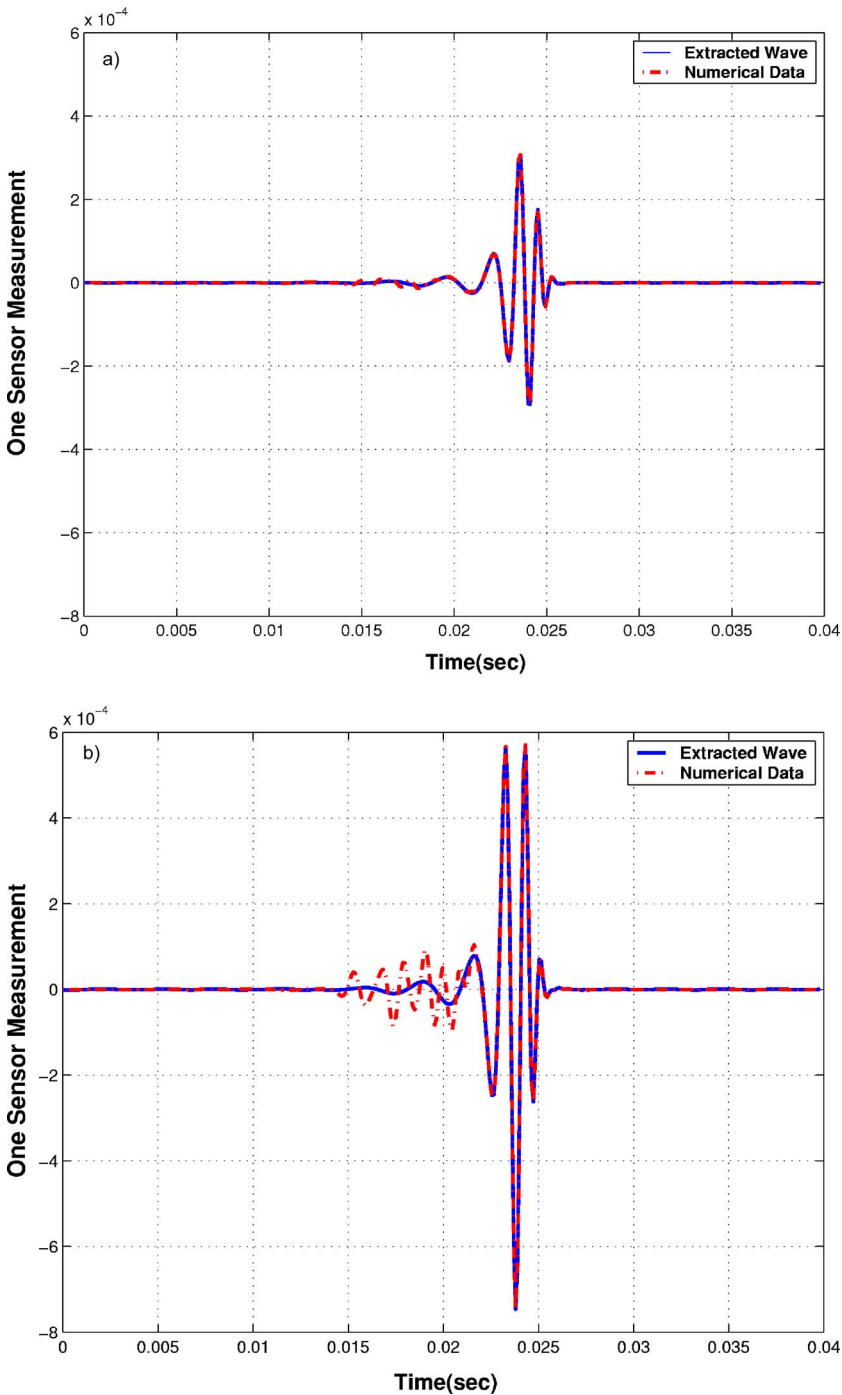


FIG. 3. (Color online) Extraction of mode-0 versus the original: (a) horizontal channel and (b) vertical channel.

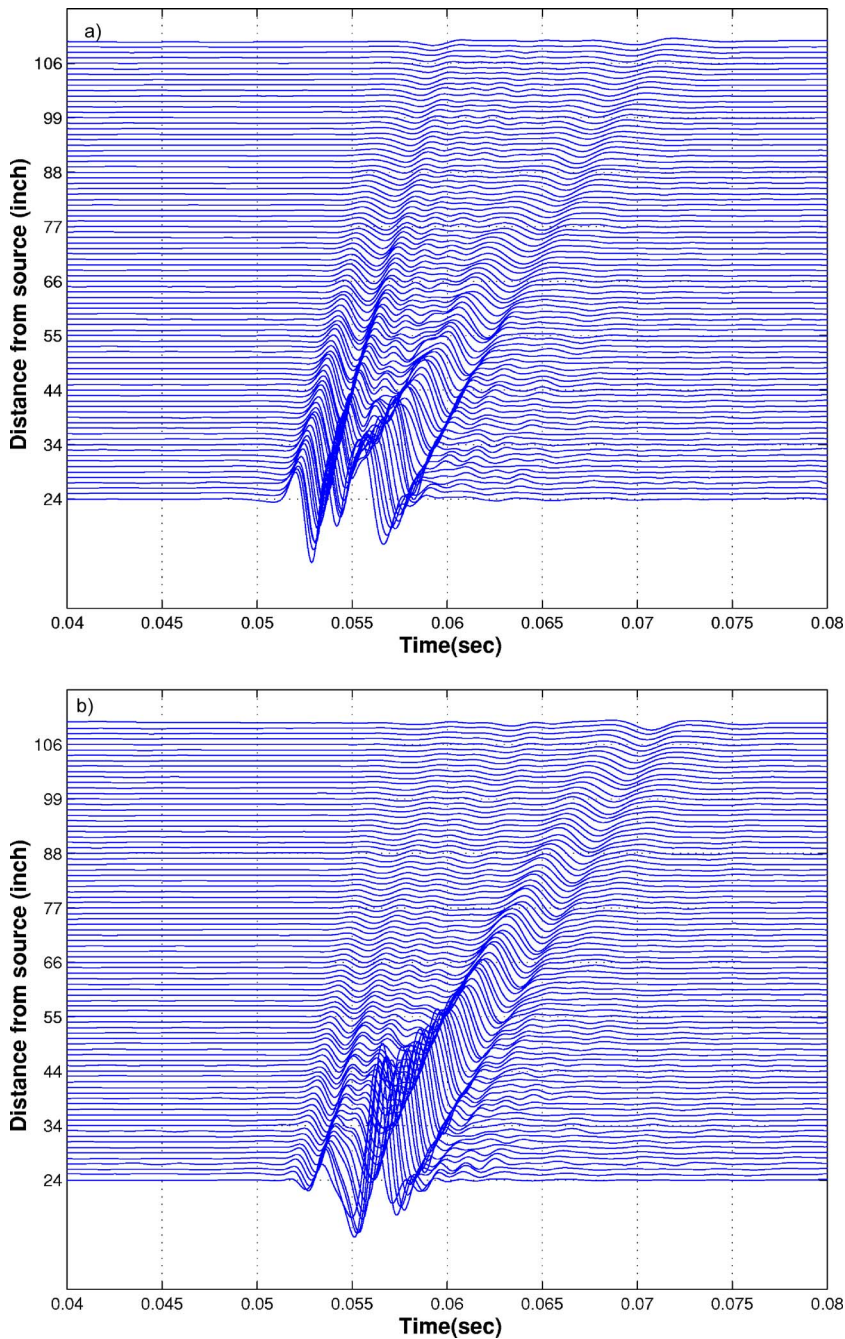


FIG. 4. (Color online) Field data (space time) for (a) horizontal channel $s_x(x, t)$ and (b) vertical channel $s_z(x, t)$.

tially linearly polarized in the vertical direction.^{1,5} The results obtained by IQML processing can be verified by comparing to the analytical solution, which produces the modes from the subsurface profile.¹³ The same profile is used to produce the space-time data. This comparison is shown in Fig. 2(c), in which the modes obtained from IQML processing follows the analytical solution very closely.

The predominant Rayleigh wave exhibits an elliptical polarization which has been calculated from the complex amplitude estimates and plotted in Fig. 2(b). At each frequency an ellipse is plotted at the corresponding phase velocity. The parameters for the ellipse, tilt angle (10), axial ratio (11), major axis (13), and minor axis (14), are obtained by using the complex amplitude estimates for the horizontal and vertical particle motions. The size of each ellipse is proportional to the complex amplitude values in the two chan-

nels. The size is also encoded in the thickness of the line displayed when plotting the ellipse, with the thickness being proportional to $\sqrt{|A_x|^2 + |A_z|^2}$. The vertical channel displacements are larger so the major axis of the ellipse is tilted toward the vertical direction for the Rayleigh wave in Fig. 2(b). The sign of the axial ratio is used to indicate the direction of particle motion, either retrograde or prograde. This polarization direction can be encoded with (color) shading: a dark (blue) shade indicating retrograde motion (as in the Rayleigh wave), and a light (red) shade for prograde.

In Fig. 2(a) it is obvious that individual modes of $s(x, t)$ can be identified and sorted according to velocity $v_p(\omega)$ and frequency. Less obvious is the fact that polarization can also be used for sorting, but the expanded view of Fig. 2(b) shows that polarization is consistent versus frequency for an indi-

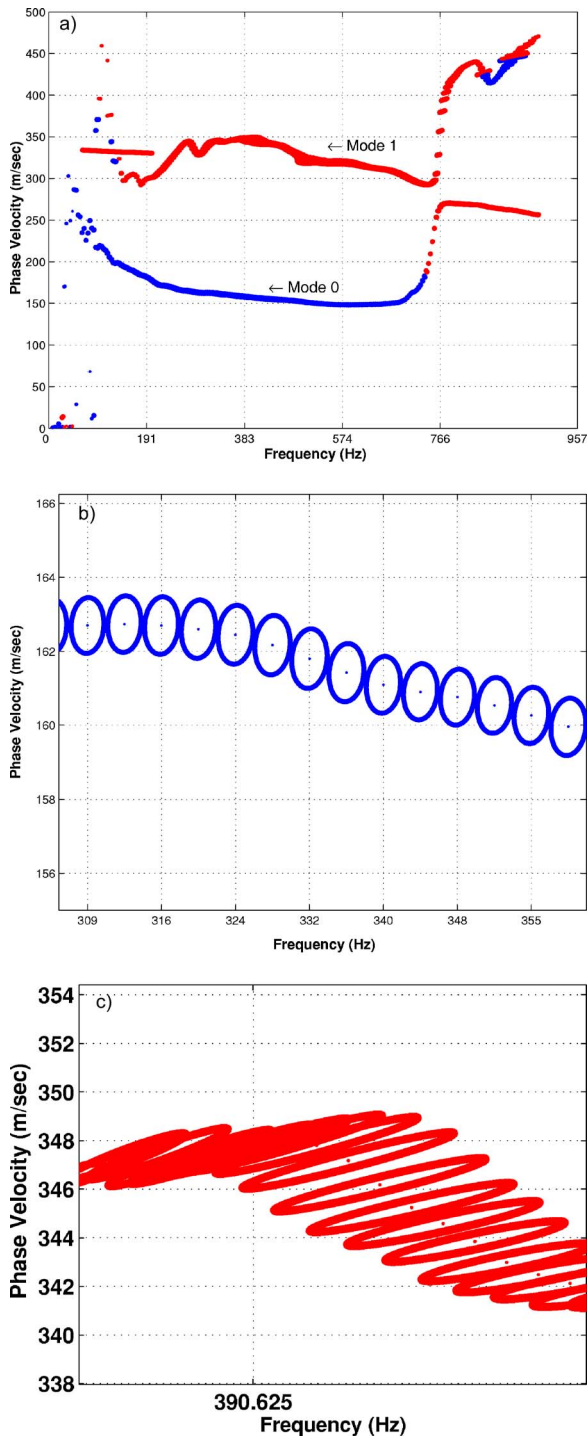


FIG. 5. (Color online) (a) Multimodal dispersion curves. Polarization ellipses for (b) the Rayleigh wave (mode-0), and (c) the Leaky surface wave (mode-1).

vidual mode; in that case, the Rayleigh wave. Once we have sorted out a single mode in the velocity-frequency domain, the wave form for that mode can be reconstructed in the space-time domain by using the model

$$s_z(x, t) = \sum_i A_z(\omega_i) e^{i\{\alpha(\omega_i)x + \int \omega_i dt + k(\omega_i)x\}}, \quad (15)$$

for the z channel; likewise, for the x channel.

Hence, by extracting individual modes from these dispersion curves, along with their complex amplitudes, we can

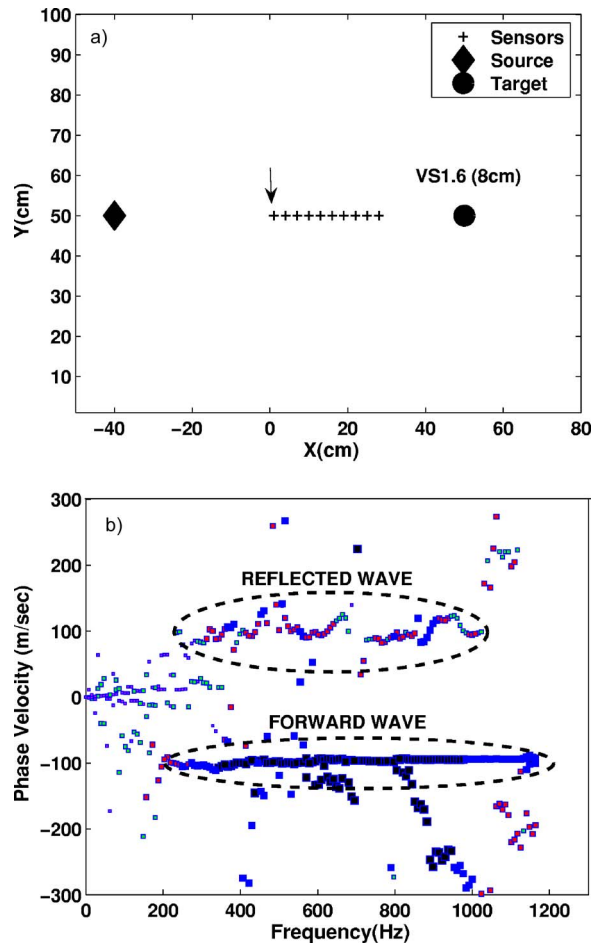


FIG. 6. (a) Array setup with linear array of sensors between the source and target. The first sensor position is indicated by the arrow. (b) (Color online) IQML spectrum analysis: reflected waves have positive velocity; forward waves, negative velocity.

reconstruct signals in the time domain for each mode using Eq. (15). This time-domain resynthesis for the fundamental mode is shown in Figs. 3(a) and 3(b) for the horizontal and vertical channels, respectively. The original numerical data is also shown for comparison. The reconstructed time-domain plot is in close agreement with the original data especially near the main pulse. The small differences between the original and the reconstructed pulses at the leading edge for the vertical channel are due to the higher-order modes which are predominantly vertically polarized. When we reconstructed the pulses from modes 1, 2, and 3 we verified this assertion. Thus we are able to separate the modes, and this separation will be a key step in the imaging methods based on the reflected Rayleigh waves in Sec. IV.

B. Processing of field data

The system used for data collection in the field is described in Refs. 5–7. In these experiments, the seismic waves are created using a surface-contacting electrodynamic transducer which has a rectangular foot. The transducer is excited with a chirp that sweeps from 30 Hz to 1 kHz with a 4 s duration. The response from this chirp is measured and used to synthesize the response to a differentiated Gaussian pulse with a center frequency of 450 Hz. The response to the

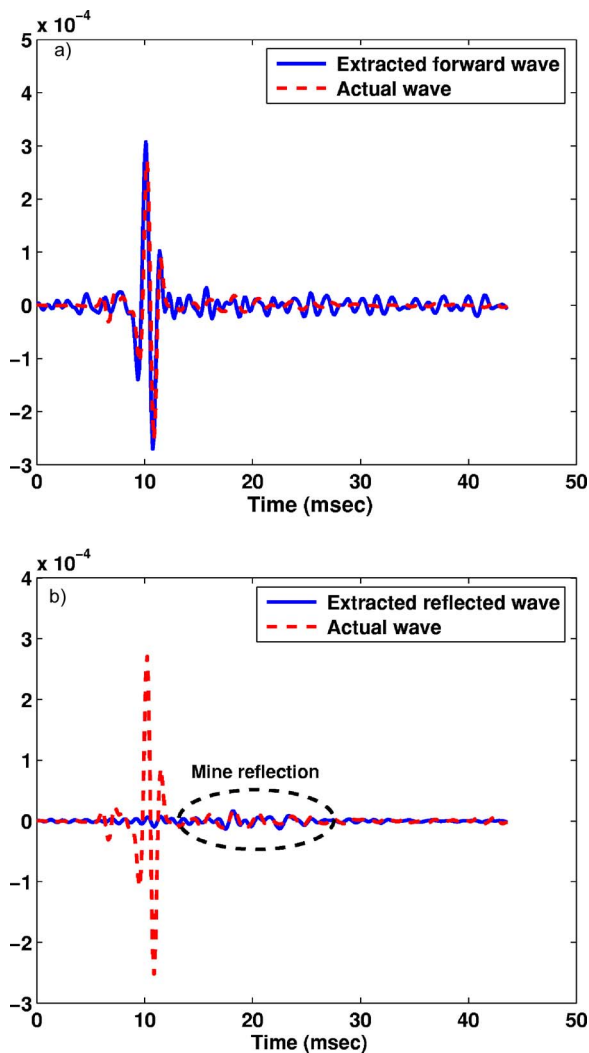


FIG. 7. (Color online) Extracted waves at one sensor: (a) forward wave and original and (b) reflected wave and original.

Gaussian pulse is graphed in the following figures. A set of field data collected in a dirt roadbed at a temperate government field site is shown in Figs. 4(a) and 4(b), for the horizontal and vertical channels, respectively. The first sensor is at a distance of 61 cm from the source with succeeding sensors 2.54 cm apart. Each sensor is a triaxial accelerometer, but only the vertical and horizontal measurements were used. In the IQML processing, the total number of sensors was 85, and the model order was $P=3$. In Fig. 5(a), there are two dispersion curves visible with mode-0 related to the Rayleigh wave. The portion of the spectrum in the frequency range greater than 766 Hz and with velocities between 400 and 450 m/s seems to be related to the pressure wave, which is the fastest body wave, and should appear at higher frequencies. In Fig. 5(b), the polarization ellipses of mode-0 are shown, and we see retrograde elliptical polarization as expected for the Rayleigh wave. Mode-1 is a leaky surface wave^{5,14} with prograde elliptical polarization, and this is confirmed in Fig. 5(c).

IV. PROCESSING IN THE PRESENCE OF BURIED LANDMINES

The IQML processing has been applied to data collected in the presence of buried targets using a single-axis sensor

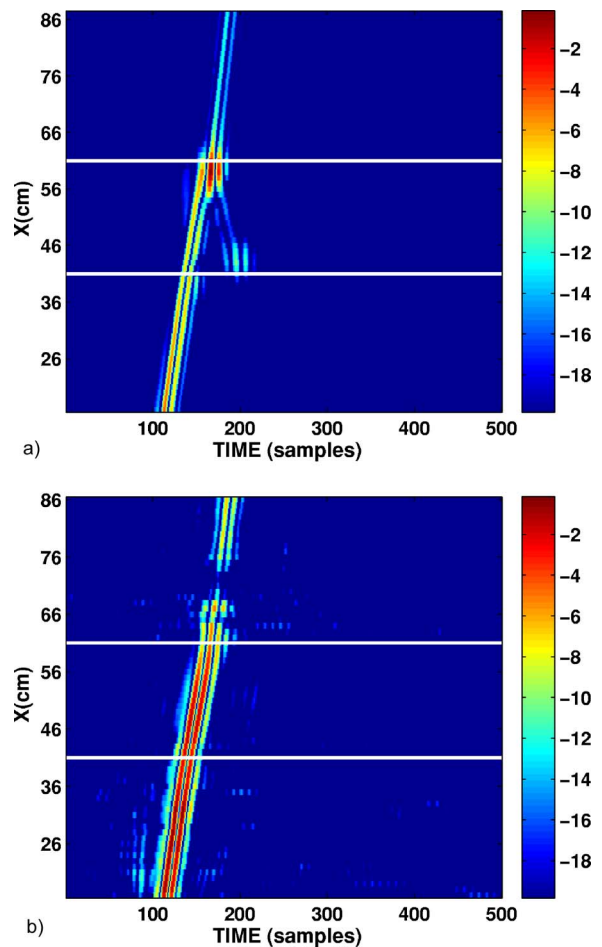


FIG. 8. (Color online) Space-time plots of (a) raw data collected at the center sensor, and (b) extracted forward wave at the center sensor. Front and back edges of the VS-1.6 mine are denoted by horizontal lines. 20-dB scale normalized to the maximum.

that records only the vertical channel.^{6,7,15,16} These experiments were performed in a laboratory facility which consists of a tank that is approximately 6-m wide \times 6-m long \times 1.5-m deep and is filled with damp compacted sand. Damp compacted sand is used as a soil surrogate in the experimental model as its properties closely resemble those of typical soils, and it can easily be reconditioned for repeatable measurements. The Rayleigh wave velocity is approximately 100 m/s in the damp compacted sand resulting in a wavelength of approximately 22 cm at the 450-Hz center frequency of the differentiated Gaussian pulse.

In the setup shown in Fig. 6(a), the sensor array lies between the source and the target, which is a VS-1.6 antitank (AT) landmine buried at a depth of 5 cm. The VS-1.6 is a plastic landmine, 23.3 cm in diameter and 9.2-cm tall. The array consists of ten sensors (ground contacting accelerometers) with an intersensor spacing of 3 cm. The IQML-derived spectrum analysis plot of this collected data is shown in Fig. 6(b). The IQML analysis easily separates the forward and reflected waves on the basis of positive and negative velocities. These waves are identified in velocity-frequency space, and then two groups of frequency-domain components are extracted, followed by reconstruction in the space-time domain using Eq. (15). The extracted forward and re-

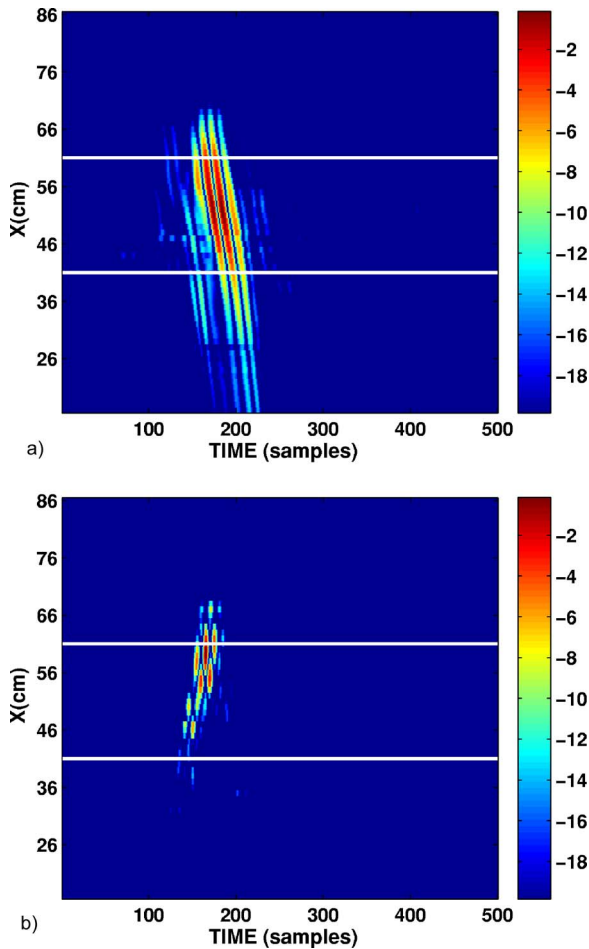


FIG. 9. (Color online) Space-time plots of (a) extracted reflected wave and (b) product of extracted reflected wave and extracted forward wave. Front and back edges of the VS-1.6 mine are denoted by horizontal lines. 20-dB scale normalized to the maximum.

verse waves at the first sensor are shown in Fig. 7. These signal reconstructions in the time domain show how well the IQML method separates these waves.

Imaging algorithm for localizing landmines

A new localization and detection algorithm can be based on the forward and reflected wave separation performed by

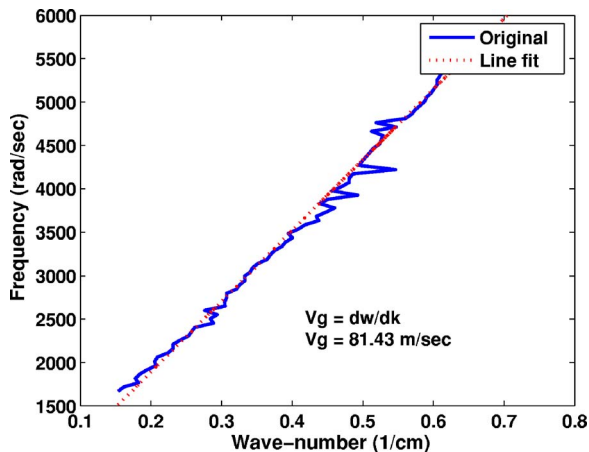


FIG. 10. (Color online) Group velocity estimate obtained from the IQML spectrum analysis.

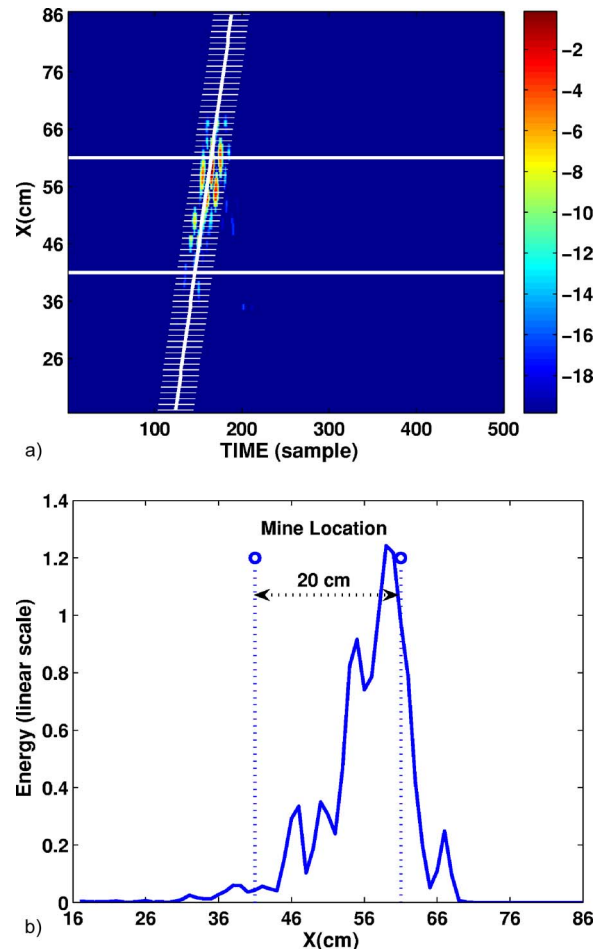


FIG. 11. (Color online) Energy based algorithm: (a) Positions of length-40 time window overlaid on the product of the forward and reflected waves [Fig. 9(b)]. (b) Energy plot, strongest where the mine is located.

IQML. It is also based on the observation that a reflection point exists when the forward and reflected waves are coincident in time and space. Thus we form a space-time image from the product of the forward and reflected waves to enhance such reflection points prior to detection.

In order to reconstruct the reflected wave at many spatial positions, we use a linear array of M sensors to collect the data. The linear array is moved away from the source, in a direction of potential targets, by the intersensor spacing Δx each time. A total of 70 array positions are used. At each array position, IQML separates and reconstructs the forward and reflected Rayleigh waves at the middle sensor of the array. If there is a target near the array then we will see a strong reflected wave and possibly a resonance in the extracted data. This reflected wave can be plotted by itself, or used to form a product image of the extracted forward and reflected waves.

The effectiveness of this product image for locating a target can be demonstrated with the previous example of an antitank mine VS-1.6 buried at a depth of 5 cm. A linear array consisting of $M=31$ sensors is moved 1 cm at a time across the target. At each array position the extracted waves are reconstructed at the middle sensor (the 16th) and are saved for use in the final plot. The raw data collected at the center sensor positions is shown in Fig. 8(a), where the front

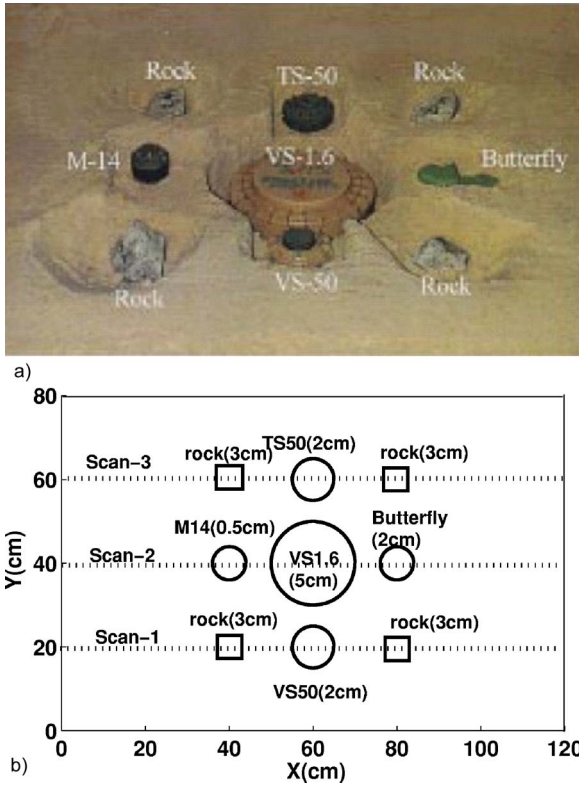


FIG. 12. Experimental setup: (a) (Color online) Photograph of the mines and rocks prior to burial. (b) Map of the relative positions and depths of the mines and rocks.

and back edges of the mine are denoted by two horizontal lines. The extracted forward wave is shown in Fig. 8(b), and the extracted reflected wave in Fig. 9(a). In this example, the extracted reflected wave shows a very strong reflection where the mine is located. The product of the reflected and forward waves is shown in Fig. 9(b), with the back edge giving the strongest response.

The images in Fig. 9 can be used to develop a detection algorithm similar to the energy-based method originally described in Ref. 17. The signal formed from the product of the forward and reflected waves will be used for the energy calculation, but it is important to make the energy calculation in a small region of the space-time plane. We use a time window of length N to extract a small part of the time signal for each spatial position x . The center of the time window is placed at the time corresponding to the travel time of the Rayleigh wave from source to the sensor position x . This travel-time calculation requires a group velocity estimate, which can be calculated directly from the IQML analysis. The group velocity is given by^{1,2}

$$V_g = \left(\frac{d}{d\omega} k(\omega) \right)^{-1}, \quad (16)$$

where $k(\omega)$ is the wave-number estimate at frequency ω . The values of $k(\omega)$ can be extracted directly from the IQML analysis. A line is fit to the mode-shape plot, and the slope of the line is the group velocity estimate. For the VS-1.6 example above, the (ω, k) plot is shown in Fig. 10. The slope of the line fit in Fig. 10 gives a value of 81.43 m/s for the group velocity estimate.

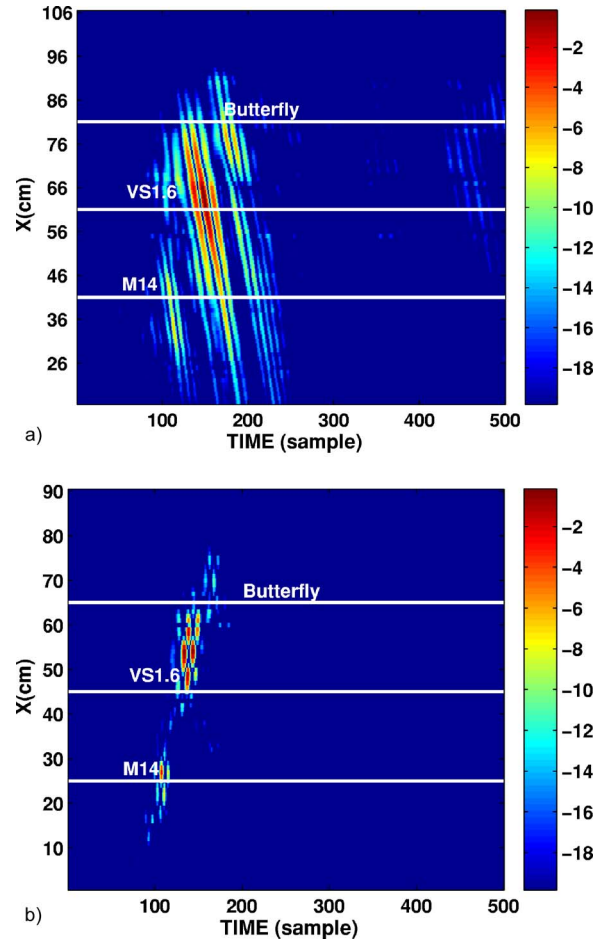


FIG. 13. (Color online) Space-time plots (20-dB scale) of the waves along the scan-2 position: (a) Extracted reflected wave and (b) product of the extracted reflected wave and the extracted forward wave. Positions of the mine centers are denoted by the horizontal lines.

Once we have the group velocity, we form a time window of length N at each spatial position, with the center of the window t_c being calculated via

$$t_c = \frac{D_x}{V_g}, \quad (17)$$

where V_g the group velocity, and D_x the distance between the source and the sensor position at x . At each spatial position the local energy is calculated via

$$E_x = \sum_{t=t_c-N/2}^{t_c+N/2} |p_r(x,t)|^2, \quad (18)$$

where $p_r(x,t)$ is the signal formed from the product of the extracted forward and reflected waves at each spatial position x . A window of length $N=40$ is used which corresponds to 5 ms in time. The window at each spatial position along with the center position line is shown in Fig. 11(a) for all x positions, and the energy calculated is shown in Fig. 11(b). The energy is strongest where the mine is located, thus indicating the spatial position of the mine. Once again, the back edge of the mine gives a stronger response.

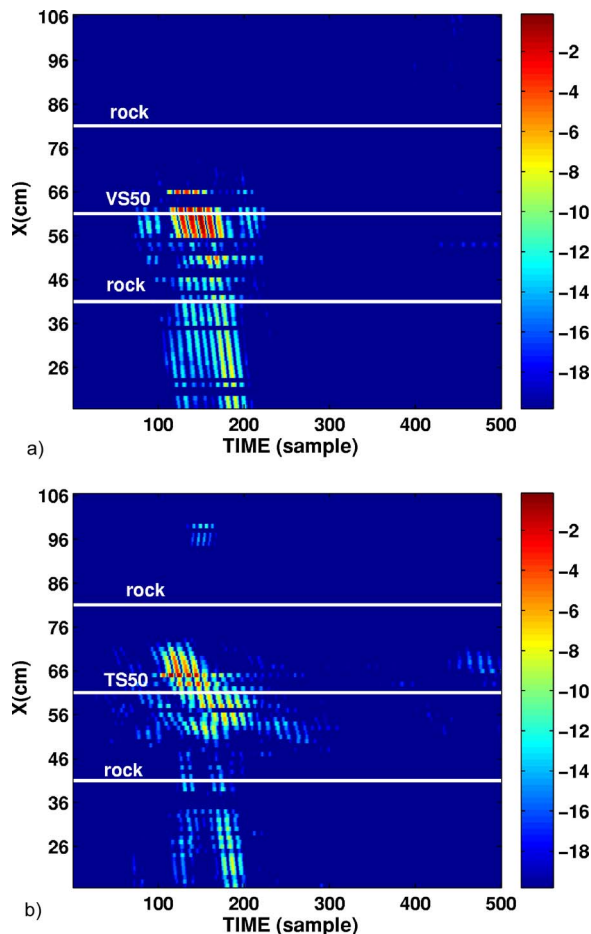


FIG. 14. (Color online) Space-time plots (20-dB scale) of the reflected wave extracted via IQML: (a) across scan-1 and (b) across scan-3. Positions of the mine centers are denoted by horizontal lines.

In the second experiment different kinds of mines are buried along with clutter (rocks). The mines are a M-14 buried at 0.5 cm, a VS-1.6 at 5 cm, a Butterfly (PFM-1) at 2 cm, a VS-50 at 1.5 cm, and a TS-50 at 2 cm. The TS-50 and VS-50 are plastic antipersonnel (AP) landmines, 9 cm in diameter and 4.6 cm in height. The M-14 is a plastic AP landmine, 5.8 cm in diameter and 3.8 cm in height. The PFM-1 is a plastic AP landmine, 12-cm long, 2-cm wide, and 6.1-cm high. The rocks are comparable in size to the TS-50.

The relative sizes, positions, and depths of the mines and rocks are shown in Fig. 12. The reflected Rayleigh waves are extracted for three different linear scans across all x positions: scan-1 at $y=20$ cm, scan-2 at $y=40$ cm, and scan-3 at $y=60$ cm. The space-time reflected wave across scan-2 is shown in Fig. 13(a), and we observe three x positions where reflections are strong. In Fig. 13(b) the product image of the extracted forward and reflected waves is shown, and we can see three positions where there are indications of the presence of targets. Similarly, waves extracted across scan-1 and scan-3 positions are shown in Fig. 14. The rocks give relatively little response, because their reflected waves do not exhibit resonance; the mines, on the other hand, form a resonator with a flexible case of the mine and the soil layer above the mine. The Rayleigh waves couple very effectively into the resonator resulting in enhanced motion above the mine

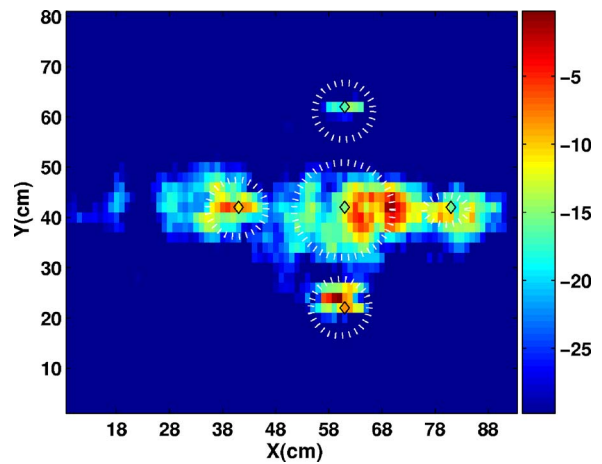


FIG. 15. (Color online) Final energy image obtained after using the entire 2D x - y scan. Positions of the five mines are indicated by circles (30-dB scale).

and an enhanced reflected wave.^{7,17} The final energy image for the whole 2D scan is shown in Fig. 15, where each row consists of an energy calculation as in Fig. 11(b). In Fig. 15 it is easy to pick out the positions of landmines, and there is hardly any evidence of the presence of rocks, which verifies the effectiveness of this technique in discriminating mines from clutter.

V. CONCLUSION

In this paper, a new method is proposed for multi-channel spectrum analysis of surface waves using a vector form of the IQML algorithm. Using this method we are able to separate not only the different modes and their polarization behavior, but also we can reconstruct these modes in the space-time domain. From collected field data we have succeeded in identifying and reconstructing the mode that is the Rayleigh wave. For landmine detection, the same algorithm for single channel data has been applied successfully in recovering the reflected Rayleigh wave from which a new imaging algorithm has been developed for detecting the signature of landmines with and without clutter.

ACKNOWLEDGMENTS

This work is supported by the U.S. Army Research Office under Contract No. DAAD19-02-1-0252. The authors would like to thank Pelham Norville for generating the numerical data, and Dr. G. D. Larson and Dr. J. S. Martin, for the field data collection. Their suggestions also helped to improve the robustness of the algorithm.

¹S. Foti, "Multistation methods for geotechnical characterization using surface waves," Ph.D. thesis, Politecnico di Torino, Italy, 2000.

²D. Zywicki, "Advanced signal processing methods applied to engineering analysis of seismic surface waves," Ph.D. thesis, Georgia Tech, Atlanta, GA, 1999.

³T. Forbriger, "Inversion of shallow-seismic wavefields. Part I: wavefield transformation," *Geophys. J. Int.* **153**, 719–734 (2003).

⁴S. Stotts, R. A. Gramann, and M. S. Bennett, "Source bearing determination from a triaxial seismometer using Rayleigh wave propagation," *J. Acoust. Soc. Am.* **115**, 2003–2012 (2004).

⁵C. T. Schröder, "On the interaction of elastic waves with buried landmines: An investigation using the finite-difference time-domain method,"

Ph.D. thesis, Georgia Tech, Atlanta, GA, 2001.

- ⁶W. R. Scott, Jr., J. S. Martin, and G. D. Larson, "Experimental model for a seismic landmine detection system," *IEEE Trans. Geosci. Remote Sens.* **39**, 1155–1164 (2001).
- ⁷W. R. Scott, Jr., G. D. Larson, J. S. Martin, and P. H. Rogers, "Seismic/electromagnetic system for landmine detection," *J. Acoust. Soc. Am.* **107**, 2897 (2000).
- ⁸E. Smith, P. S. Wilson, F. W. Bacon, J. F. Manning, J. A. Behrens, and T. G. Muir, "Measurement and localization of interface wave reflections from a buried target," *J. Acoust. Soc. Am.* **103**, 2333–2343 (1998).
- ⁹S. W. Lang, A. L. Kurkjian, J. H. McClellan, C. F. Morris, and T. W. Parks, "Estimating slowness dispersion from arrays of sonic logging waveforms," *Geophys. J.* **52**, 530–544 (1987).
- ¹⁰M. Alam, J. H. McClellan, and W. R. Scott, Jr., "Multi-channel spectrum analysis of surface waves," in *Proceedings 37th Asilomar Conference Signals, Systems, and Computers*, 771–775 (IEEE, New York, 2003).
- ¹¹J. H. McClellan and D.-W. Lee, "Exact equivalence of the Steiglitz-McBride iteration and IQML," *IEEE Trans. Signal Process.* **39**, 509–512 (1991).
- ¹²D. Staelin, A. Morgenthaler, and J. Kong, *Electromagnetic Waves* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
- ¹³C. G. Lai and G. J. Rix, "Solution of the Rayleigh eigenproblem in viscoelastic media," *Bull. Seismol. Soc. Am.* **92**, 2297–2309 (2002).
- ¹⁴C. Schröder and W. R. Scott, Jr., "On the complex conjugate roots of the Rayleigh equation: The leaky surface wave," *J. Acoust. Soc. Am.* **110**, 2867–2877 (2001).
- ¹⁵G. D. Larson, J. S. Martin, and W. R. Scott, Jr., "Seismic landmine detection using microphones as near-ground sensors," *J. Acoust. Soc. Am.* **117**, 2385 (2005).
- ¹⁶P. Norville and W. R. Scott, Jr., "Time-reversal focusing of elastic surface waves," *J. Acoust. Soc. Am.* **118**, 735–744 (2005).
- ¹⁷A. Behoodian, J. H. McClellan, and W. R. Scott, Jr., "Signal processing of elastic surface waves for localizing landmines," in *Proceedings 33rd Asilomar Conference Signals, Systems, and Computers*, 827–830 (IEEE, New York, 1999).

Separation of components from impulses in reassigned spectrograms

Sean A. Fulop^{a)}

Department of Linguistics, California State University, Fresno, California 93740-8001

Kelly Fitz^{b)}

School of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington 99164-2752

(Received 13 September 2006; revised 1 December 2006; accepted 6 December 2006)

Two computational methods for pruning a reassigned spectrogram to show only quasisinusoidal components, or only impulses, or both, are presented mathematically and provided with step-by-step algorithms. Both methods compute the second-order mixed partial derivative of the short-time Fourier transform phase, and rely on the conditions that components and impulses are each well-represented by reassigned spectrographic points possessing particular values of this derivative. This use of the mixed second-order derivative was introduced by Nelson [J. Acoust. Soc. Am. **110**, 2575–2592 (2001)] but here our goals are to completely describe the computation of this derivative in a way that highlights the relations to the two most influential methods of computing a reassigned spectrogram, and also to demonstrate the utility of this technique for plotting spectrograms showing line components or impulses while excluding most other points. When applied to speech signals, vocal tract resonances (formants) or glottal pulsations can be effectively isolated in expanded views of the phonation process. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431329]

PACS number(s): 43.60.Hj, 43.72.Ar [EJS]

Pages: 1510–1518

I. INTRODUCTION

A. Overview

This paper presents signal processing techniques which extend previous work² on the reassigned (or *time-corrected instantaneous frequency*) spectrogram. The procedures discussed herein allow a kind of spectrogram (i.e., time-frequency analysis) to be computed which isolates either the line components or the impulses comprising a signal (or both together). Implementable step-by-step algorithms for applying these methods are presented here for the first time, together with examples of fruitful application to speech signals.

Immediately in the sequel we first recap previous descriptions of the reassigned spectrogram with examples. We then in Sec. II describe the theory behind using higher-order derivatives of the short-time Fourier transform phase (complex argument) to “prune” reassigned spectrograms and thereby show only those points meeting certain physical conditions. Numerous example images complement the presentation of these ideas, which were first set down by Nelson.³ In Sec. III, two possible methods for computing the higher-order short-time Fourier transform (STFT) phase derivatives are presented with algorithms that invite the reader’s implementation. The mathematical derivation of an exact transform method for computing the second-order mixed partial

derivative of the STFT phase is included in this section. The paper concludes with Sec. IV which discusses the physical interpretation of this second-order derivative, and relates it to a previously published notion⁴ of *consensus* among reassigned instantaneous frequencies.

B. Digression on the reassigned spectrogram

We here briefly review the reassigned spectrogram, which is discussed more fully in previous work.² We assume our signal $f(t)$ can be modeled as the sum of general AM/FM components:

$$f(t) = \sum_n A_n(t) e^{i[\Omega_n(t) + \phi_n]} \quad (1)$$

and that the STFT is defined in the following way:

$$\text{STFT}_h(\omega, T) = \int_{-\infty}^{\infty} f(t+T)h(-t)e^{-i\omega t} dt. \quad (2)$$

This form of the transform is equivalent modulo a phase factor $e^{i\omega t}$ to the more prevalent form in which the window is time-translated with the signal held to a fixed time [see Eq. (9)].

The *channelized instantaneous frequency* (CIF) of a signal as a function of time and frequency is

$$\text{CIF}(\omega, T) = \frac{\partial}{\partial T} \arg(\text{STFT}_h(\omega, T)), \quad (3)$$

where STFT_h is the short-time Fourier transform using window function h . If there is just one AM/FM component dominant in the neighborhood of a frequency bin, then the CIF spectrum will show the instantaneous frequency of

^{a)} Author to whom correspondence should be addressed. Department of Linguistics PB92, California State University, 5245 N. Backer Ave., Fresno, CA 93740-8001. Phone: (559) 278-4896. Electronic mail: sfulop@csufresno.edu

^{b)} Present address: Starkey Hearing Research Center, 2150 Shattuck Ave., Suite 408, Berkeley, CA 94704.

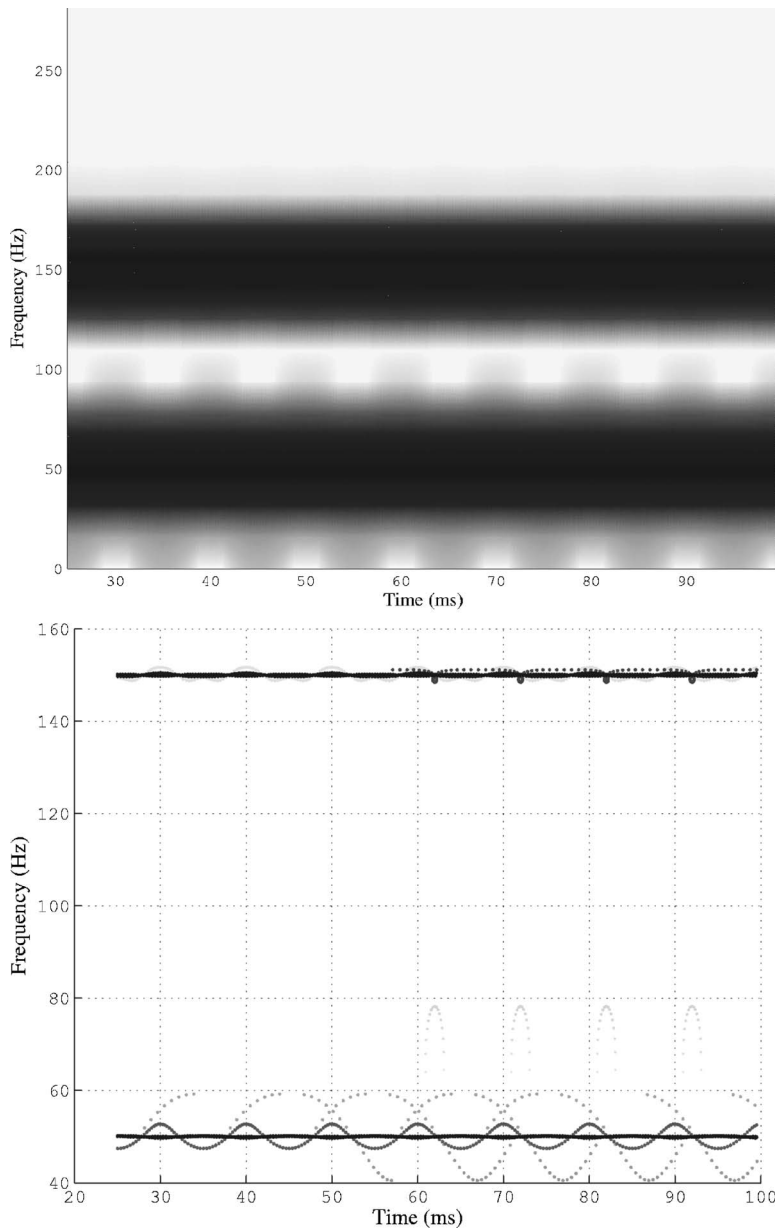


FIG. 1. The figure contrasts the conventional spectrogram (upper panel) with the reassigned spectrogram of a signal composed from a 50 Hz and 150 Hz sine wave for 1/8 second, sampled at 32 kHz. Some interference artifacts can be noted. Both images computed using 1600 point frames and 10 point frame advance.

that component with arbitrary precision (i.e., not quantized by the discrete time-frequency grid).

The *local group delay* (LGD) of a signal is given by

$$\text{LGD}(\omega, T) = -\frac{\partial}{\partial \omega} \arg(\text{STFT}_h(\omega, T)). \quad (4)$$

The reassigned spectrogram plots each STFT magnitude at the new location of its computed CIF, and at a time point corrected by its LGD. The LGD at a time-frequency point (ω_0, T_0) on the STFT matrix may be interpreted as the average true time of $\text{STFT}_h(\omega_0, T_0)$; this is an estimate of the “time correction” to the maximum energy point of the dominant AM/FM signal component observed at (ω_0, T_0) .

The new time-frequency coordinate (computed using the CIF and LGD) is a more meaningful location for the STFT magnitude, and corresponds to the local mean of the complex time-frequency energy distribution of the signal.⁵ The resulting increase in spectrographic imaging precision is illustrated in Figs. 1–3.

II. REASSIGNED SPECTROGRAMS WITH PARTIAL DERIVATIVE THRESHOLDING

Despite the obvious gains in clarity with which the location and frequency modulation of line components can be shown in these reassigned spectrograms, as well as the improved time localization of impulsive events, the images can be disappointingly noisy. In the figures thus far, one can note random speckle and interference not clearly associated with either a line component or an impulse.

This is owing to numerous factors, but in rough sketch, the algorithm employed to locate the AM/FM components in the signal has a meaningful output only in the neighborhood of a component. Where there is no component of significant amplitude, the time-frequency locations of the points to be plotted can become random. There can also be a variety of other computation artifacts resulting from failure of the separability condition, among other circumstances. We next detail a method which has the potential to “denoise” our spectro-

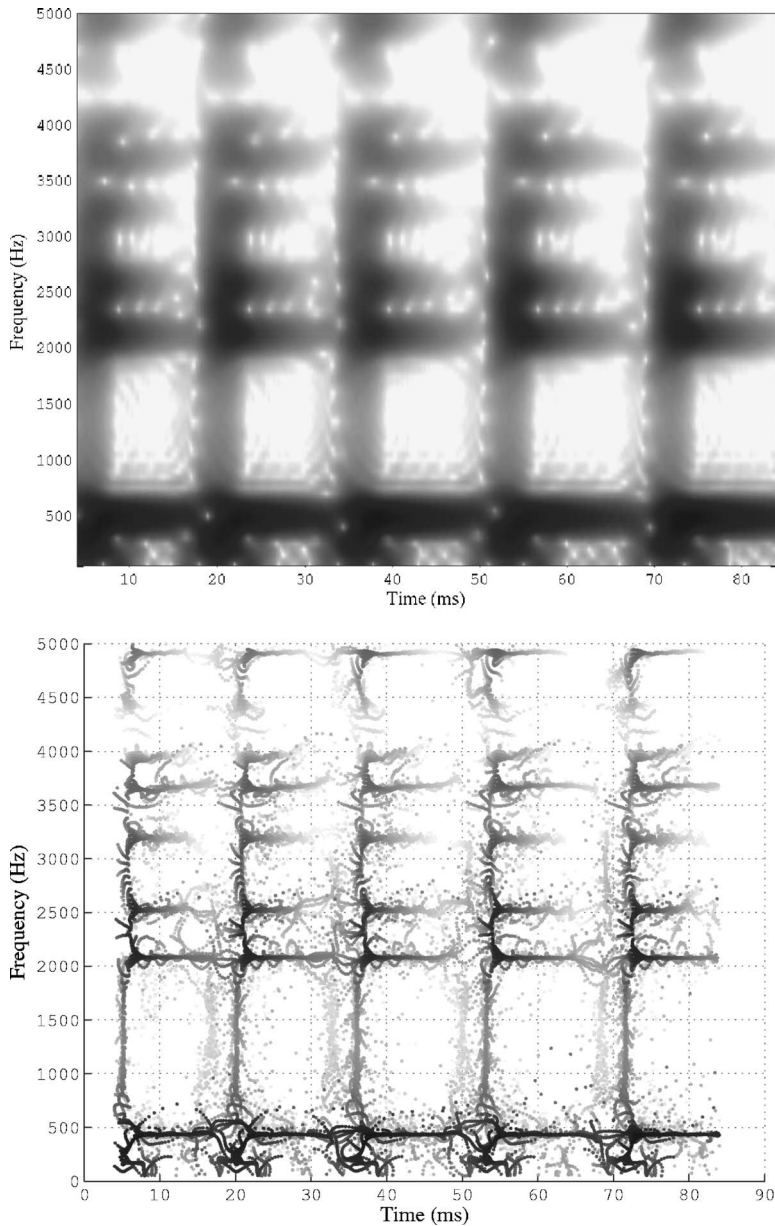


FIG. 2. The conventional spectrogram (upper panel) shows a portion of the vowel [e] (day) pronounced with “creaky voice” for low airflow; the reassigned spectrogram of the lower panel shows the same signal, in which the individual pulsations of the vocal cords are rendered clearly visible along with the mouth resonances (formants) which they excite at each impulse. All speech examples in this paper are sampled at 51.2 kHz. Both images computed using 400 point frames and 4 point frame advance.

grams by isolating quasistationary (low chirp rate) components in a display, or alternatively to permit highly time-localized events (impulses) to be isolated.

A. Interpretation of the mixed partial derivative

The technique to be expounded was first theoretically outlined by Nelson,³ and involves the computation of the second-order mixed partial derivative of the STFT phase. This is equivalent to either the frequency derivative of the CIF, or to the time derivative of the LGD, since it is a fundamental theorem of calculus that the mixed partial derivative can be taken in either order.

Nelson (op. cit.) argued that the nearly stationary AM/FM components of a signal $x(T)$ satisfy

$$\frac{\partial^2}{\partial \omega \partial T} \arg(\text{STFT}_h(\omega, T)) = \frac{\partial}{\partial \omega} \text{CIF}_x(\omega, T) \approx 0. \quad (5)$$

Further explication of this fact is deferred to a later discussion. By plotting just those points in a reassigned spectro-

gram meeting this condition on the second-order mixed partial phase derivative to within a threshold, a spectrogram showing just the line components can be drawn.

Nelson (op. cit.) further asserted that the impulses in a signal $x(T)$ satisfy

$$\frac{\partial^2}{\partial T \partial \omega} \arg(\text{STFT}_h(\omega, T)) = \frac{\partial}{\partial T} \text{LGD}_x(\omega, T) \approx 1. \quad (6)$$

By plotting just those points meeting this condition to within a threshold, a spectrogram showing just the impulsive events in a signal can alternatively be drawn. Plotting all points meeting the disjunction of the above conditions results in a “de-noised” spectrogram showing quasisinusoidal components and impulses together, to the exclusion of most everything else.

B. Application to the reassigned spectrogram

To selectively plot components meeting the condition of a quasisinusoid, one keeps only those points having a value

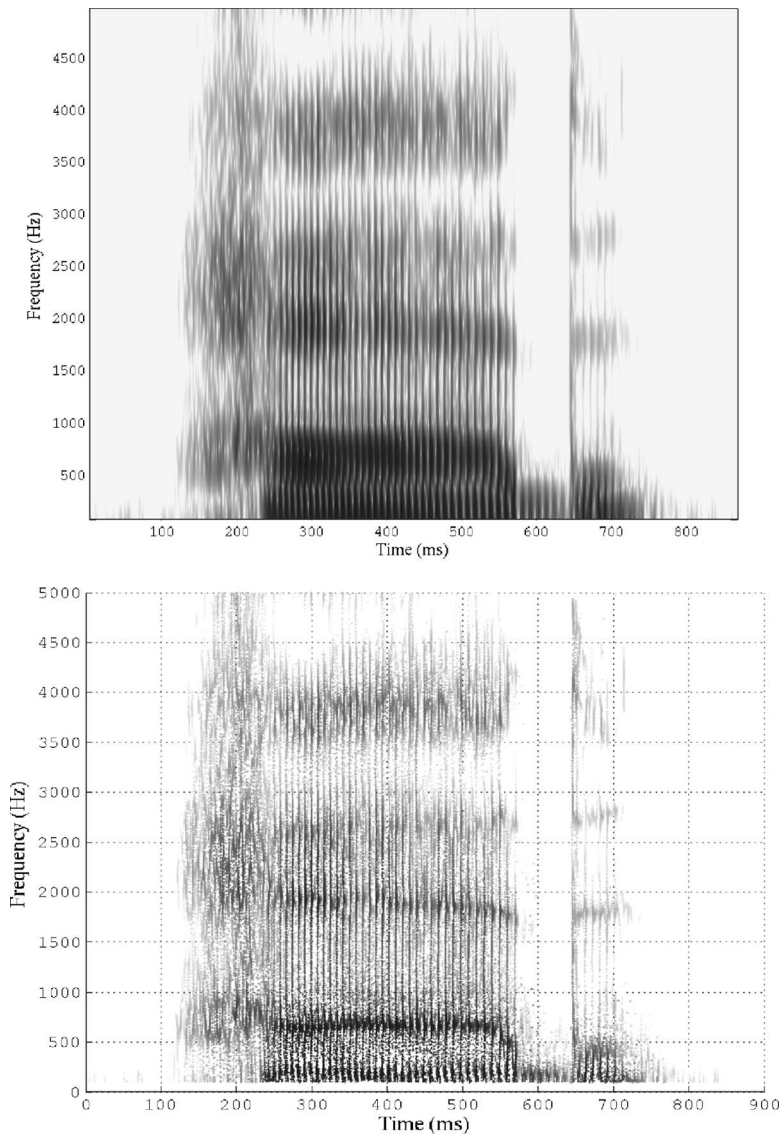


FIG. 3. This figure shows a conventional spectrogram of the English word *had* (upper panel, computed using 300 point frames and 20 point frame advance), along with a reassigned spectrogram below (computed with 40 point frame advance).

of $\frac{\partial^2}{\partial T \partial \omega} \arg(\text{STFT}_h(\omega, T))$ sufficiently near 0, when one computes the derivative using Nelson's definition. The precise threshold can be empirically determined, and will in practice depend on the degree of deviation from a pure sinusoid that is tolerable in the application at hand. This means that greater tolerance in this threshold will be required where line components having high chirp rates are expected—for speech signals an absolute value of the derivative on the order of 0.2 is often a reasonable threshold (see Figs. 4 and 5). On the other hand, a numerical derivative threshold value which is several orders of magnitude smaller can be used to eliminate nearly every point that does not represent a pure sinusoid with no frequency modulation, as is illustrated in Fig. 6.

By Nelson's definition of the short-time Fourier transform, points whose second-order mixed partial phase derivative is near 1 are likely to be associated with impulsive events. For display purposes it is appropriate to be quite tolerant in this threshold, depending on what sort of signal content we desire to regard as "impulsive." A derivative value between 0.75 and 1.25 usually yields good results for

speech signals, without straying too far from identifiably impulse-like events (see Figs. 7 and 8).

III. COMPUTING THE HIGHER-ORDER MIXED PARTIAL DERIVATIVE

A. Cross-spectral method

The mathematical theory behind cross-spectral expressions for all higher-order partial derivatives of the STFT phase is completely presented in prior literature,³ and we have relied on this in developing the particular algorithm for the second-order mixed partial derivative that is presented below. The steps in the computational method will be based on the "Nelson method" algorithm for the reassigned spectrogram published by Fulop and Fitz.² Readers are invited to refer to that algorithm to complement that presented below. It is important to note that, just as with the cross-spectral method for computing the first-order STFT phase derivatives (and thereby the reassigned spectrogram) discussed in previous work,^{1,2} the method presented here will compute an *approximation* of the second-order mixed partial STFT phase

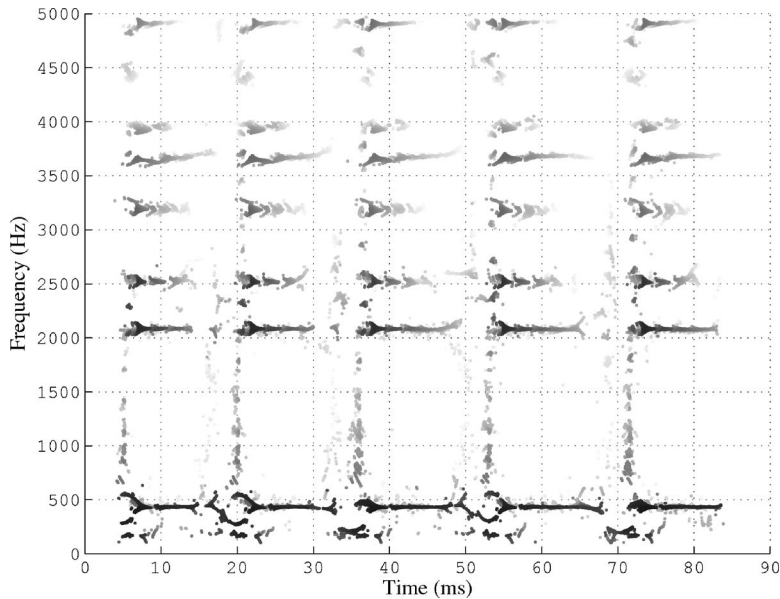


FIG. 4. Another view of the creaky voiced vowel shown in Fig. 2, showing only those points whose absolute value mixed partial derivative of the STFT phase is less than 0.1.

derivative. This approximation is generally so close that it might not matter for practical purposes.

(1) First, one builds two matrices S and S_{del} of windowed signal frames of length `win_size` (user-supplied to the procedure) time samples, with S_{del} having frames that are delayed by one sample with respect to S . For the present purposes a standard Hann window function will suffice, but other windows may be more appropriate for other applications. The windowed signal frames overlap by the same user-input number of points in each of the matrices.

(2) One next computes four short-time Fourier transform matrices; each column is an `fftn`-length-Fourier transform of a signal frame, computed with a fast Fourier transform function called `fft`. The length value `fftn` is supplied by the user. The difference between `fftn` and `win_size` is zero-padded up to `fftn` for the computation.

$STFT_{\text{del}} = \text{fft}(S_{\text{del}})$

$STFT = \text{fft}(S)$

$STFT_{\text{freqdel}}$ is just $STFT$ rotated by one frequency

bin—this can be accomplished by shifting the rows in $STFT$ up by one step and moving the former last row to the new first row.

$STFT_{\text{ftrtimedel}}$ is $STFT_{\text{del}}$ similarly rotated by one frequency bin.

(3) Next compute a cross-spectral surface by applying Nelson's³ theory:

$$\text{MixCIF} = STFT \times STFT_{\text{del}}^* \times (STFT_{\text{freqdel}} \times STFT_{\text{ftrtimedel}}^*)^*, \quad (7)$$

where the notation X^* for complex X indicates the complex conjugate (pointwise if X is a matrix of complex numbers). The notation $A \times B$ for matrices A, B denotes a point-by-point product, not a matrix multiplication.

(4) Now the partial frequency derivative of the channelized instantaneous frequency can be computed:

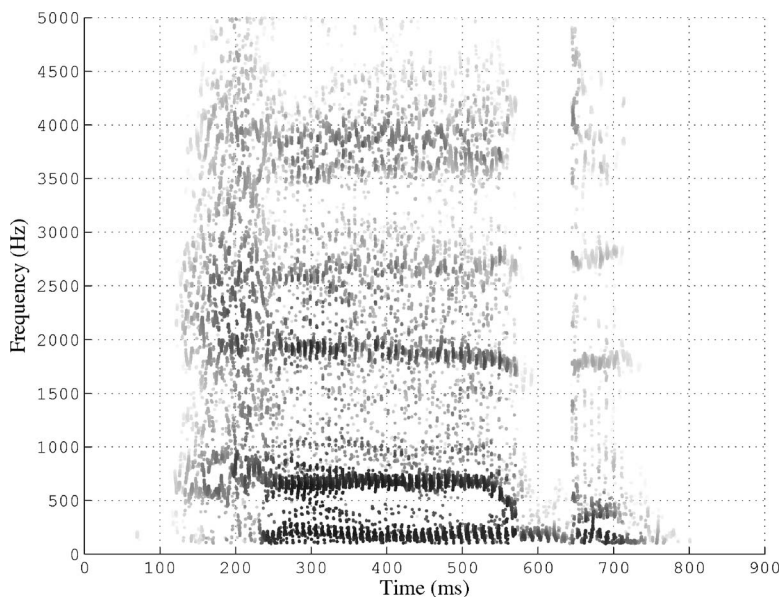


FIG. 5. Another view of the word *had* from Fig. 3, this time showing only points having $|\frac{\partial^2}{\partial \omega \partial T} \arg(STFT_h(\omega, T))| < 0.1$.

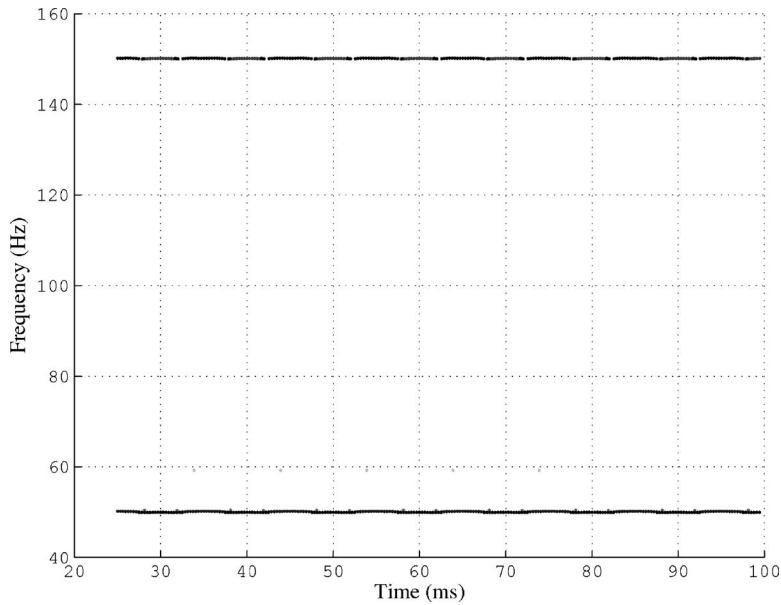


FIG. 6. Another view of the double sine wave shown in Fig. 1, this time showing only those points whose absolute value second-order mixed partial derivative of the STFT phase is less than 10^{-4} .

$$\text{CIFderiv} = \frac{\text{fftn} \cdot \text{Fs}}{2\pi \cdot \text{win_size}} \times \arg(\text{MixCIF}) \times \arg(\text{MixCIF}), \quad (8)$$

where the $\arg(\cdot)$ function is valued on the range $(0, 2\pi)$, and Fs is the sampling rate (in Hz) of the signal.

The final quantity computed by the above algorithm is equivalent to the partial time derivative of the local group delay, and either of these represents the (unique) second-order mixed partial derivative of the STFT phase.

B. Exact transform method

This subsection presents the mathematical theory behind an alternative approach to Nelson's cross-spectral theory,

$$X(t, \omega) = \int x(\tau) h^*(t - \tau) e^{-j\omega\tau} d\tau \quad (9)$$

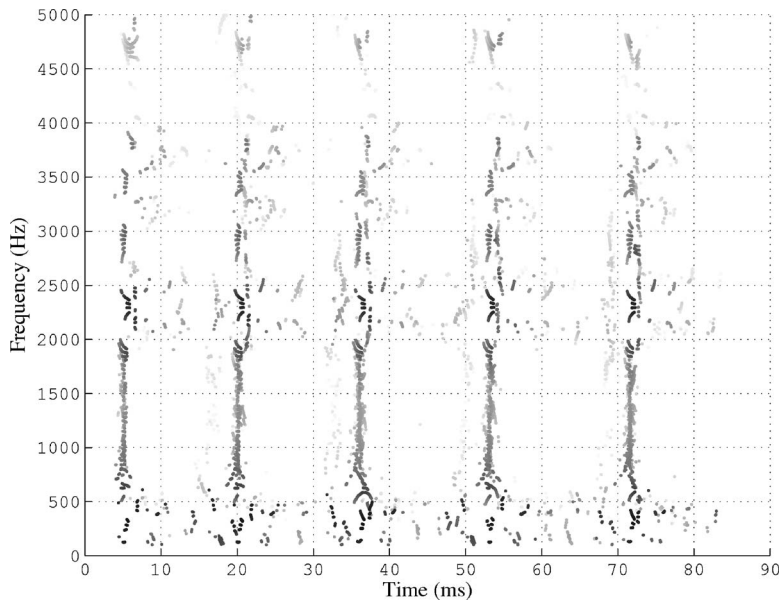


FIG. 7. This reassigned spectrogram shows the same vowel as in Figs. 2 and 5, but now a partial derivative threshold of time-correction characteristic of impulses has been applied, so only those points having $|\frac{\partial^2}{\partial\omega\partial T} \arg(\text{STFT}_h(\omega, T)) - 1| < 0.25$ are plotted.

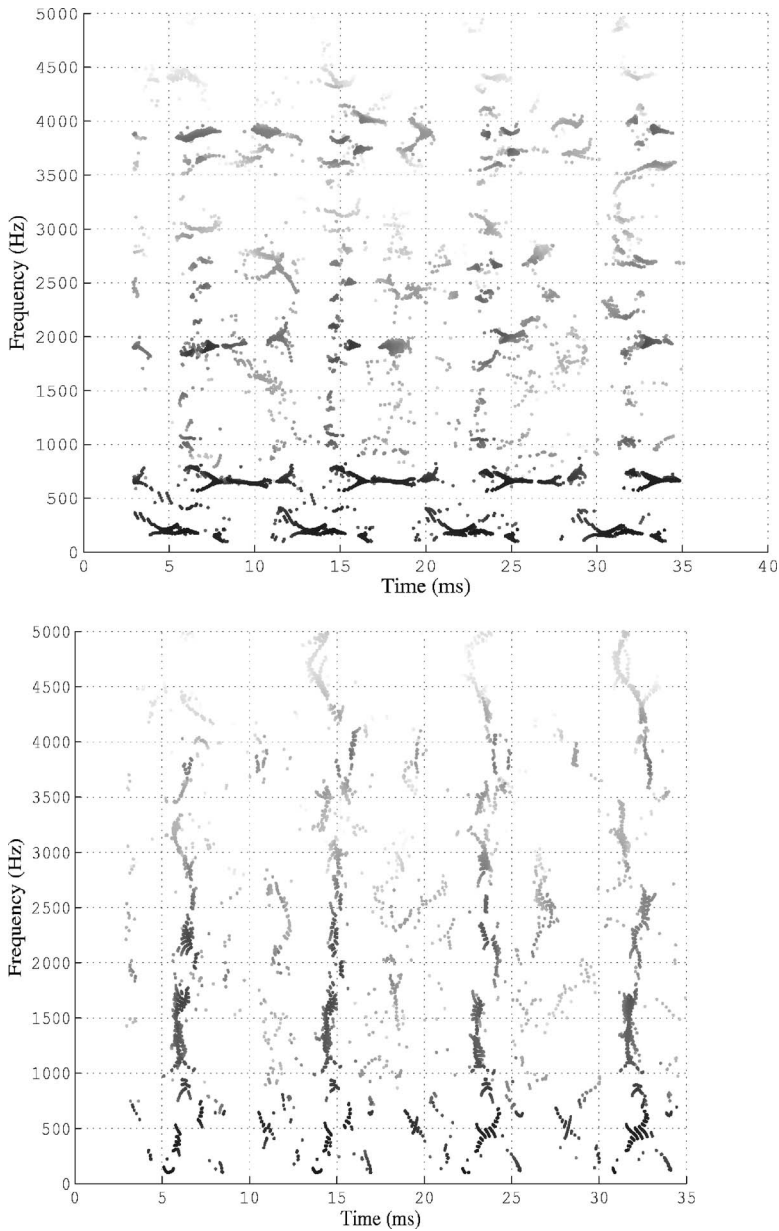


FIG. 8. These spectrograms zoom in on a few vocal cord pulsations from the vowel [æ] in the utterance of *had* shown in Figs. 3 and 6. The upper panel shows just the line components, while the lower panel focuses instead on the impulsive events. Both images computed with 300 point frames and 4 point frame advance.

$$=M(t,\omega)e^{j\phi(t,\omega)} \quad (10)$$

where $h(t)$ is a finite-length, real-valued window function, $M(t,\omega)$ is the magnitude of the short-time Fourier transform, and $\phi(t,\omega)$ is its phase.

Since

$$\frac{\partial X(t,\omega)}{\partial t} = \frac{\partial M(t,\omega)}{\partial t} e^{j\phi(t,\omega)} + j \frac{\partial \phi(t,\omega)}{\partial t} X(t,\omega), \quad (11)$$

the mixed partial derivative of the short-time Fourier transform with respect to time t and frequency ω can be expressed as a sum of two terms:

$$\begin{aligned} \frac{\partial^2 X(t,\omega)}{\partial t \partial \omega} &= \frac{\partial}{\partial \omega} \left[\frac{\partial M(t,\omega)}{\partial t} e^{j\phi(t,\omega)} \right] \\ &+ j \frac{\partial}{\partial \omega} \left[\frac{\partial \phi(t,\omega)}{\partial t} X(t,\omega) \right]. \end{aligned} \quad (12)$$

The first term on the right-hand side of Eq. (12) can be written

$$\begin{aligned} \frac{\partial}{\partial \omega} \left[\frac{\partial M(t,\omega)}{\partial t} e^{j\phi(t,\omega)} \right] &= \frac{\partial^2 M(t,\omega)}{\partial t \partial \omega} e^{j\phi(t,\omega)} \\ &+ jX(t,\omega) \Re \left\{ \frac{X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)} \right\} \Re \left\{ \frac{X_{\mathcal{T}h}(t,\omega)}{X(t,\omega)} \right\} \\ &- jX(t,\omega) \Re \left\{ \frac{X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)} \right\} \end{aligned} \quad (13)$$

and the second term on the right-hand side of Eq. (12) can be written

$$\begin{aligned}
j \frac{\partial}{\partial \omega} \left[\frac{\partial \phi(t, \omega)}{\partial t} X(t, \omega) \right] &= tX(t, \omega) \mathfrak{I} \left\{ \frac{X_{\mathcal{D}h}(t, \omega)}{X(t, \omega)} \right\} \\
&\quad - X_{\mathcal{T}h}(t, \omega) \mathfrak{I} \left\{ \frac{X_{\mathcal{D}h}(t, \omega)}{X(t, \omega)} \right\} \\
&\quad + jX(t, \omega) \frac{\partial^2 \phi(t, \omega)}{\partial t \partial \omega}. \quad (14)
\end{aligned}$$

Here, $X_{\mathcal{T}h}(t, \omega)$ is the short-time Fourier transform computed using a time-weighted analysis window $h_{\mathcal{T}}(t) = t \cdot h(t)$ and $X_{\mathcal{D}h}(t, \omega)$ is the short-time Fourier transform computed using a time-derivative analysis window $h_{\mathcal{D}}(t) = \frac{d}{dt}h(t)$.

The complicated expression that we obtain by substituting Eqs. (13) and (14) into Eq. (12) is useful because one of its terms contains the mixed partial derivative of the short-time Fourier transform phase, $\frac{\partial^2 \phi(t, \omega)}{\partial t \partial \omega}$, which is the quantity of interest. A much simpler expression for the mixed partial derivative of the short-time Fourier transform with respect to time t and frequency ω is

$$\frac{\partial^2 X(t, \omega)}{\partial t \partial \omega} = -j t X_{\mathcal{D}h}(t, \omega) + j X_{\mathcal{T}Dh}(t, \omega), \quad (15)$$

where $X_{\mathcal{T}Dh}(t, \omega)$ is the short-time Fourier transform computed using the analysis window

$$h_{\mathcal{T}D}(t) = t \frac{d}{dt}h(t) = t h_{\mathcal{D}}(t). \quad (16)$$

Equating the right-hand sides of Eqs. (12) and (15), using the substitutions given by Eqs. (13) and (14), and taking advantage of the fact that the short-time Fourier transform phase is a real quantity, so its mixed partial derivative must also be real, we can obtain the simplified expression for the mixed partial derivative of short-time Fourier transform phase:

$$\frac{\partial^2 \phi(t, \omega)}{\partial t \partial \omega} = \Re \left\{ \frac{X_{\mathcal{T}Dh}(t, \omega)}{X(t, \omega)} \right\} - \Re \left\{ \frac{X_{\mathcal{T}h}(t, \omega) X_{\mathcal{D}h}(t, \omega)}{X(t, \omega)^2} \right\}. \quad (17)$$

Equation (17) gives a method of computing the mixed partial derivative of short-time Fourier transform phase at discrete time-frequency coordinates using only discrete short-time Fourier transform data, with no approximations of partial derivatives. The different initial definition of the short-time Fourier transform in Eq. (9) has the effect of shifting the relevant values of the mixed partial phase derivatives, so that line components and impulses are now characterized by derivatives near -1 and 0 , respectively.

The algorithm invoking the above considerations is as follows:

(1) A time ramp and frequency ramp are constructed for the modified window functions, and these depend in detail on whether there is an odd or even number of data points in each frame. Accordingly, the following algorithm should be used to obtain the ramps and the special windows:

1: **if** mod(win_size,2) **then**
2: $Mw = (\text{win_size} - 1)/2$
3: $\text{framp} = [(0:Mw), (-Mw:-1)]$ (using Matlab colon no-

tation for a sequence of numbers stepping by one over the specified range)

4: $\text{tramp} = (-Mw:Mw)$

5: **else**

6: $Mw = \text{win_size}/2$

7: $\text{framp} = [(0.5:Mw-0.5), (-Mw+0.5:-0.5)]$

8: $\text{tramp} = (-Mw+0.5:Mw-0.5)$

9: **end if**

10: $\text{tramp} = \text{tramp}/Fs$, where Fs is the sampling rate (in Hz) of the signal

11: $\text{framp} = (Fs/\text{win_size}) \times \text{framp}$

12: $Wt = \text{tramp} \times \text{window}$

13: $Wdt = -\text{imag}(\text{ifft}(\text{framp} \times \text{fft}(\text{window})));$ (ifft is the inverse transform function to fft)

14: $Wtdt = \text{tramp} \times Wdt$

2. One next builds four matrices of windowed signal frames of length win_size time samples. The matrix S has its frames windowed by the nominal function window. The matrices S_{time} , S_{deriv} , and S_{td} have their frames windowed, respectively, by Wt , Wdt , and $Wtdt$. The windowed signal frames overlap by a user-supplied number of points.

3. One next computes four corresponding short-time Fourier transform matrices in the customary manner described in the previous algorithm:

$\text{STFT} = \text{fft}(S)$

$\text{STFT}_{\text{time}} = \text{fft}(S_{\text{time}})$

$\text{STFT}_{\text{deriv}} = \text{fft}(S_{\text{deriv}})$

$\text{STFT}_{\text{td}} = \text{fft}(S_{\text{td}})$

4. $\text{STFTeps} = \max(\text{STFT}, \text{eps})$; this adds the minimum floating point precision value defined by Matlab to any zero values.

5.

$$\begin{aligned}
\text{MixPD} &= 2\pi \left(\Re \left(\frac{\text{STFT}_{\text{td}}}{\text{STFTeps}} \right) - \Re \left(\frac{\text{STFT}_{\text{deriv}}}{\text{STFTeps}} \right) \right) \\
&\quad \times \Re \left(\frac{\text{STFT}_{\text{time}}}{\text{STFTeps}} \right) \\
&\quad + \Im \left(\frac{\text{STFT}_{\text{deriv}}}{\text{STFTeps}} \right) \times \Im \left(\frac{\text{STFT}_{\text{time}}}{\text{STFTeps}} \right). \quad (18)
\end{aligned}$$

Once again the notation $A \times B$ for matrices A, B denotes a point-by-point product.

Comparing the computational cost of the two algorithms, the chief difference comes from the exact method's requiring four fast Fourier transforms of various windowings of the entire signal matrix, while Nelson's method requires only two. These computations can take non-negligible amounts of time for a large number of signal frames, while on the other hand the remaining calculations to be performed in either method are simple pointwise multiplications and additions that will not differentiate the methods noticeably. In practice, it has been found that by far the most time-consuming aspect of obtaining a reassigned spectrogram by either method is the process of displaying all the points, a

process that depends entirely on the implementational details of the visualization component of one's software.

The images in this paper have been produced using Matlab software, and the form of the algorithms somewhat reflects the architecture of this language. For example, it is probably unwise in any computer language to literally construct the large matrices in the algorithms to perform the pointwise multiplications and other operations, and Matlab indeed handles matrices more efficiently than is implied by the simple representation. Readers may wish to refer to the Matlab implementations of these routines which are linked on the first author's web page.

IV. DISCUSSION

A. Justifying the interpretation of the phase derivative

In regions where the CIF is not changing with frequency, the spectrum is dominated by a single component that is highly concentrated in frequency (i.e., a sinusoid). In these regions, all nearby spectral data are mapped to the frequency of the dominant sinusoid, so that the variation (partial derivative) with respect to frequency is near zero. Similarly, in regions in which all spectral data are mapped to the time of a dominant component that is highly concentrated in time (i.e., an impulse), the variation (partial derivative) of the reassigned time with respect to time is near zero. Since the reassigned time is computed by adding the LGD to the nominal time t ,

$$0 \approx \frac{\partial}{\partial t}[t + \text{LGD}(t, \omega)] = 1 + \frac{\partial}{\partial t}\text{LGD}(t, \omega) \quad (19)$$

so

$$1 \approx -\frac{\partial}{\partial t}\text{LGD}(t, \omega). \quad (20)$$

That is, as the nominal time t increases, the time correction (LGD) for data in the neighborhood of a dominant impulse decreases proportionally.

B. The related notion of *consensus*

Figure 9 demonstrates the effect of frequency reassignment for a fragment of voiced speech. The upper plot shows the conventional (dashed lines) and reassigned (crosses) magnitude spectra. The lower plot shows the mapping of nominal (Fourier transform bin) frequency to reassigned frequency for the same fragment of speech. Near the frequencies of strong harmonics, the mapping is flat, as all nearby transform data are reassigned to the frequency of the dominant harmonic component. This *consensus*, or clustering among reassigned frequency estimates in the vicinity of spectral peaks can be used as an indicator of the reliability of the time-frequency data.⁴ If the reassigned frequencies for neighboring short-time Fourier transform channels are all very similar, then there is said to be a high degree of con-

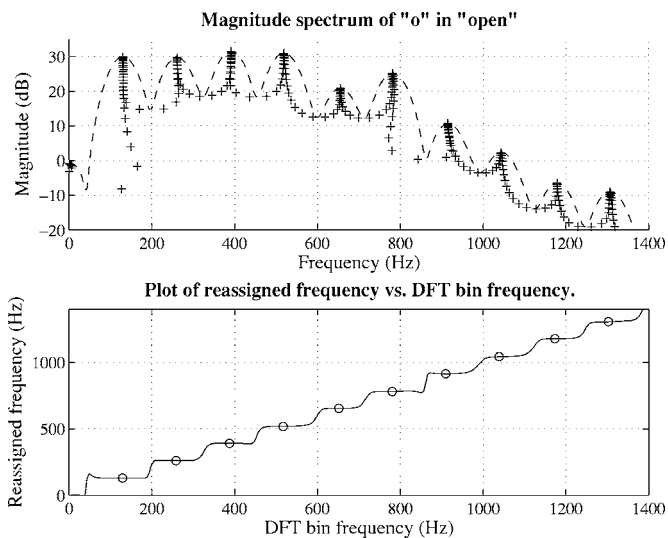


FIG. 9. Demonstration of frequency reassignment in a single spectrum for a fragment of speech (the "o" in "open"), computed using a 33.6 ms Kaiser analysis window with a shaping parameter of 12. The upper plot shows the conventional (dashed lines) and reassigned (crosses) magnitude spectra. The lower plot shows the mapping of nominal (Fourier transform bin) frequency to reassigned frequency for the same fragment of speech. The flat portions of the lower curve represent regions in which the energy in many transform bins is reassigned to the same frequency. The circled points show the samples having locally minimal frequency reassignments.

sensus and the quality of the frequency estimates is assumed to be good. In fact, consensus is exactly what is measured, in a very local way, by the mixed partial phase derivative. Moreover, an analogous argument can be made in the time dimension, wherein consensus would be a measure of impulsiveness, or more specifically a measure of the appropriateness of the analysis window when the signal is sparse in time.

ACKNOWLEDGMENTS

Thanks to Doug Nelson for his encouragement, and for work worth following.

- ¹D. J. Nelson, "Cross-spectral methods for processing speech," J. Acoust. Soc. Am. **110**, 2575–2592 (2001).
- ²S. A. Fulop and K. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," J. Acoust. Soc. Am. **119**, 360–371 (2006).
- ³D. J. Nelson, "Instantaneous higher order phase derivatives," Digit. Signal Process. **12**, 416–428 (2002).
- ⁴T. J. Gardner and M. O. Magnasco, "Instantaneous frequency decomposition: An application to spectrally sparse sounds with fast frequency modulations," J. Acoust. Soc. Am. **117**, 2896–2903 (2005).
- ⁵K. Kodera, C. de Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," Phys. Earth Planet. Inter. **12**, 142–150 (1976).
- ⁶F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," IEEE Trans. Signal Process. **43**, 1068–1089 (1995).
- ⁷H. A. Murthy, K. V. Madhu Murthy, and B. Yegnanarayana, "Formant extraction from phase using weighted group delay function," Electron. Lett. **25**, 1609–1611 (1989).
- ⁸A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

A Bayesian method to estimate the depth and the range of phonating sperm whales using a single hydrophone

Christophe Laplanche^{a)}

Laboratoire Images, Signaux et Systèmes Intelligents, Groupe Ingénierie des Signaux Neuro-Sensoriels,
Université Paris 12, Créteil, France

(Received 28 September 2006; revised 30 December 2006; accepted 2 January 2007)

Some bioacousticians have used a single hydrophone to calculate the depth/range of phonating diving animals. The standard one-hydrophone localization method uses multipath transmissions (direct path, sea surface, and seafloor reflections) of the animal phonations as a substitute for a vertical hydrophone array. The standard method requires three multipath transmissions per phonation. Bioacousticians who study foraging sperm whales usually do not have the required amount of multipath transmissions. However, they usually detect accurately (using shallow hydrophones towed by research vessels) direct path transmissions and sea surface reflections of sperm whale phonations (*clicks*). Sperm whales emit a few thousand clicks per foraging dive, therefore researchers have this number of direct path transmissions and this number of sea surface reflections per dive. The author describes a Bayesian method to combine the information contained in those acoustic data plus visual observations. The author's tests using synthetic data show that the accurate estimation of the depth/range of sperm whales is possible using a single hydrophone and without using any seafloor reflections. This method could be used to study the behavior of sperm whales using a single hydrophone in any location no matter what the depth, the relief, or the constitution of the seafloor might be. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2436644]

PACS number(s): 43.60.Jn [WWA]

Pages: 1519–1528

I. INTRODUCTION

Sperm whales undertake long foraging dives to catch their prey. They breathe at the sea surface, fluke-up and swim downwards to reach their prey, hunt at depth, and reascend back to the sea surface (Miller *et al.*, 2004a). During foraging dives, sperm whales emit echolocation clicks (Backus and Schevill, 1966). They emit echolocation clicks at a tremendous source level (Møhl *et al.*, 2003, 2000) and in series (Whitehead and Weilgart, 1990).

Since sperm whales emit long series of clicks of high source level, passive acoustics is an effective tool to study the foraging behavior of these animals. Researchers have developed and used different passive acoustic localization techniques. These techniques require synchronous recordings made on tridimensional (Watkins and Schevill, 1972), bidimensional (Thode, 2004), or unidimensional (Villadsgaard *et al.*, 2007) arrays of hydrophones.

To locate the sound source using an array of n hydrophones, one would need to isolate one signal emitted by the source and to measure the n times of arrival (TOA) of this signal on the n hydrophones of the array. The differences between TOAs (TOAD) are calculated. A TOAD provides information on the location of the source: The source is on a sheet of a two-sheeted hyperboloid of geometry given by the TOAD itself and by the location of the hydrophones used to calculate the TOAD. By repeating this localization process using different TOADs, one can, using the required amount of hydrophones (Spiesberger, 2001), geometrically or ana-

lytically compute the intersection of the hyperboloid sheets and therefore be able to more accurately identify the location of the source. A unidimensional array requires at least three hydrophones (Villadsgaard *et al.*, 2007). In this case the hyperboloid sheets defined by the TOADs intersect into a circle. The plane containing this circle is perpendicular to the line of the array and the center of the circle sits on this line. Therefore, if the unidimensional array is vertical, then the circle is horizontal, and its depth and radius give the depth and the horizontal range of the sound source.

One noteworthy passive acoustic localization technique requires a single hydrophone [see for instance Thode *et al.* (2002) or Laplanche *et al.* (2005)]. Signals emitted by sound sources may reflect on the sea surface and the seafloor while propagating to the hydrophones. The detection on a hydrophone of the echoes from the surface/seafloor serves as a substitute of an unidimensional vertical hydrophone array (Urlick, 1983). By measuring the TOADs of the echoes (later referred to as echo delays) relative to the nonreflected transmitted signal, one can calculate the depth and the range of the sound source. If the source is a phonating sperm whale, one can theoretically, by repeating the localization process on every click emitted by the whale during a whole dive, plot the values of the depth and the range of the whale during this dive.

Unfortunately, sperm whales usually forage above continental slopes or abyssal plains, i.e., areas of either deep or high relief seafloor. One can seldom clearly detect the seafloor echoes of clicks emitted by sperm whale using a hydrophone (close to the sea surface) towed by a research vessel. Usually one can only detect seafloor echoes at the beginning

^{a)}Electronic mail: laplanche@gmail.com

of the whale's dive, while the whale both swims and clicks downwards (Thode *et al.*, 2002). In low sea state conditions, using a hydrophone close to the sea surface, one can however clearly detect surface echoes during the whole dive [see for instance Thode (2004) or Laplanche *et al.* (2005)].

Nevertheless, the measurement of a single echo delay (e.g., a surface echo delay) is not enough to calculate the depth and the range of a phonating sperm whale, since, as aforementioned, by using a single TOAD one can only know that the whale is on a hyperboloid sheet. Is it not possible to more accurately identify the location of the whale using a single hydrophone but not using seafloor echoes?

Actually one can still estimate the depth and the range of the whale under such constraints, and the aim of this work is to demonstrate the feasibility of this process. Every single surface echo delay contains information regarding the location of the whale. By combining the pieces of information contained in the set of the surface echo delays of the clicks emitted by the whale during a dive, one should be able to give a more accurate description of the location of the whale during this dive. The efficient combining of these pieces of information can be achieved in a Bayesian frame. The Bayesian approach has already proven to be efficient to locate sound sources using TOADs (Spiesberger, 2005).

The author proposes a Bayesian passive acoustic technique to estimate the depth and the range of foraging sperm whales using a single hydrophone and without using any seafloor echoes. To be applied, this technique requires a set of values of the sea surface echo delay of clicks emitted by the whale during the whole dive. It also requires the measurement, by a visual observer, of the approximate location (i.e., range and azimuth relative to the research vessel) of the whale when beginning and ending the foraging dive.

II. MATERIAL AND METHODS

A. Trajectory model

First, one would need a representation of the underwater trajectory of the whale using a mathematical model. Let $t^{(A)}$ be the time when the sperm whale flukes-up and starts diving, $t^{(B)}$ the time when the whale starts clicking, $t^{(C)}$ the time when the whale stops clicking, and $t^{(D)}$ the time when the whale resurfaces. The author will decompose the trajectory of the whale for $t \in [t^{(B)}, t^{(C)}]$ into $n \in \mathbb{N}^*$ pieces of equal duration $\tau_0 = (t^{(C)} - t^{(B)})/n$ (Fig. 1). Let $t^{(s)} = t^{(B)} + (s-1)\tau_0$ be the time when the whale is at the beginning ($s \in \{1, \dots, n\}$) or at the end ($s \in \{2, \dots, n+1\}$) of such trajectory pieces.

Let $E^{(s)}$ be the location in the terrestrial reference frame, let $z^{(s)}$ be the depth, let $r^{(s)}$ be the horizontal range, and let $\theta^{(s)}$ be the azimuth of the whale at time $t^{(s)}$ (Fig. 2). Let $E_p^{(s)}$ and H_p be the projections of $E^{(s)}$ and H (H is the location of the hydrophone) on a horizontal plane. Let $\gamma_b^{(s)}$ and $\gamma_e^{(s)}$ be the angles $(E_p^{(s)} \vec{E}_p^{(s+1)}, H_p \vec{E}_p^{(s)})$ and $(E_p^{(s)} \vec{E}_p^{(s+1)}, H_p \vec{E}_p^{(s+1)})$.

The value of n is chosen high enough to be able to assume that the whale moves at constant speed and constant heading for $t \in [t^{(s)}, t^{(s+1)}]$ ($s \in \{1, \dots, n\}$). Let $\mathcal{S}^{(s)} = [E^{(s)} E^{(s+1)}]$ be the segment defining the location of the whale for $t \in [t^{(s)}, t^{(s+1)}]$. Let $v_z^{(s)} \in \mathbb{R}$ and $v_r^{(s)} \in \mathbb{R}^+$ be the vertical and horizontal speeds of the whale along the seg-

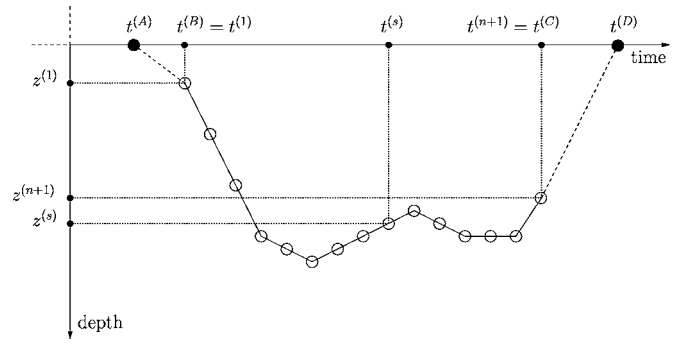


FIG. 1. The whale dives at $t=t^{(A)}$, starts clicking at $t=t^{(B)}$, stops clicking at $t=t^{(C)}$, and resurfaces at $t=t^{(D)}$. The whale is at the depth $z=z^{(s)}$ at $t=t^{(s)} = t^{(B)} + (s-1)\tau_0$ ($s \in \{1, \dots, n+1\}$, in this example $n=14$). The vertical speed of the whale is constant and equal to $v_z^{(s)}$ for $t \in [t^{(s)}, t^{(s+1)}]$.

ment $\mathcal{S}^{(s)}$. Each segment $\mathcal{S}^{(s)}$ is entirely defined by the coordinates of the points $E^{(s)}$ and $E^{(s+1)}$. The trajectory of the whale for $t \in [t^{(B)}, t^{(C)}]$ is labeled \mathcal{T} . It itself is entirely defined by the location of the summits $E^{(s)}$ ($s \in \{1, \dots, n+1\}$), that is to say $\mathcal{T} \equiv (E^{(1)}, \dots, E^{(n+1)})$. This definition of \mathcal{T} requires, by using the coordinates in the terrestrial reference frame of the $n+1$ summits $E^{(s)}$, a set of $3n+3$ parameters.

One can then define the trajectory \mathcal{T} using the depth, range, and heading of the whale. Each segment $\mathcal{S}^{(s)}$ can be recursively defined by writing

$$\mathcal{S}^{(1)} \equiv (z^{(1)}, r^{(1)}, \theta^{(1)}, z^{(2)}, r^{(2)}, \gamma_b^{(1)}),$$

$$\mathcal{S}^{(s)} \equiv (z^{(s+1)}, r^{(s+1)}, \gamma_b^{(s)} | \mathcal{S}^{(s-1)}) \quad \text{for } s \in \{2, \dots, n\} \quad (1)$$

and the trajectory \mathcal{T} is entirely defined by the set of $3n+3$ parameters

$$\mathcal{T} \equiv (z^{(1)}, r^{(1)}, \theta^{(1)}, z^{(2)}, r^{(2)}, \gamma_b^{(1)}, \dots, z^{(n+1)}, r^{(n+1)}, \gamma_b^{(n)}). \quad (2)$$

This leads to the following definition of \mathcal{T} , which is required by the algorithm described later. Let $\Delta_\gamma^{(s)} = \gamma_e^{(s-1)} - \gamma_b^{(s)}$ be the change of heading of the whale at time $t^{(s)}$. Each segment $\mathcal{S}^{(s)}$ is recursively defined by writing

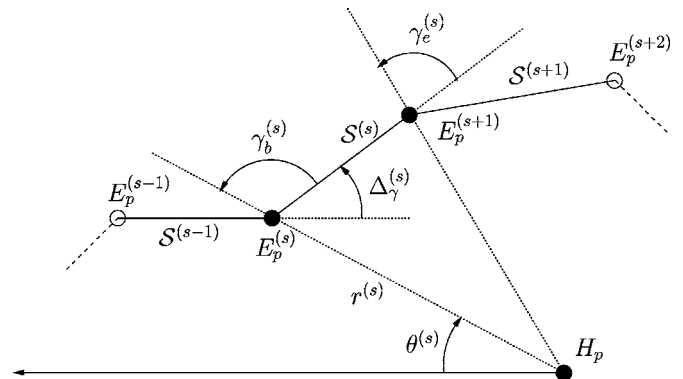


FIG. 2. The whale is at $E^{(s)}$ at $t=t^{(s)}$. The hydrophone is at H . $E_p^{(s)}$ and H_p are the projections of $E^{(s)}$ and H on a horizontal plane. The whale moves in a constant heading and at a constant horizontal speed $v_r^{(s)}$ for $t \in [t^{(s)}, t^{(s+1)}]$ along the segment $\mathcal{S}^{(s)}$ to reach $E^{(s+1)}$ at $t=t^{(s+1)}$. The angles $\gamma_b^{(s)}$ and $\gamma_e^{(s)}$ are defined as the angles between $(E_p^{(s)} E_p^{(s+1)})$ and $(H_p E_p^{(s)})$, and $(E_p^{(s)} E_p^{(s+1)})$ and $(H_p E_p^{(s+1)})$, respectively. The change of heading from $\mathcal{S}^{(s-1)}$ to $\mathcal{S}^{(s)}$ is $\Delta_\gamma^{(s)} = \gamma_e^{(s-1)} - \gamma_b^{(s)}$. The horizontal range and the azimuth of the whale at $t=t^{(s)}$ are $r^{(s)} = H_p E_p^{(s)}$ and $\theta^{(s)}$, respectively.

$$\mathcal{S}^{(1)} \equiv (z^{(1)}, r^{(1)}, \theta^{(1)}, v_z^{(1)}, v_r^{(1)}, \gamma_b^{(1)}),$$

$$\mathcal{S}^{(s)} \equiv (v_z^{(s)}, v_r^{(s)}, \Delta_\gamma^{(s)} | \mathcal{S}^{(s-1)}) \quad \text{for } s \in \{2, \dots, n\} \quad (3)$$

given the coordinates of the first summit [$E^{(1)}$, defined by the triplet $(z^{(1)}, r^{(1)}, \theta^{(1)})$] and the whale vertical speed, horizontal speed, and change of heading in the n segments. Alternatively, given the coordinates of the last summit [$E^{(n+1)}$, defined by the triplet $(z^{(n+1)}, r^{(n+1)}, \theta^{(n+1)})$] and the whale vertical speed, horizontal speed, and change of heading in the n segments, each segment $\mathcal{S}^{(s)}$ is recursively defined by writing

$$\mathcal{S}^{(s)} \equiv (v_z^{(s)}, v_r^{(s)}, \Delta_\gamma^{(s+1)} | \mathcal{S}^{(s+1)}) \quad \text{for } s \in \{1, \dots, n-1\}$$

$$\mathcal{S}^{(n)} \equiv (v_z^{(n)}, v_r^{(n)}, \gamma_e^{(n)}, z^{(n+1)}, r^{(n+1)}, \theta^{(n+1)}). \quad (4)$$

Using Eq. (3), the trajectory \mathcal{T} is also entirely defined by the set of $3n+3$ parameters

$$\begin{aligned} \mathcal{T} \equiv & (z^{(1)}, r^{(1)}, \theta^{(1)}, v_z^{(1)}, v_r^{(1)}, \gamma_b^{(1)}, \\ & v_z^{(2)}, v_r^{(2)}, \Delta_\gamma^{(2)}, \dots, v_z^{(n)}, v_r^{(n)}, \Delta_\gamma^{(n)}), \end{aligned} \quad (5)$$

or alternatively, using Eq. (4), by the set of $3n+3$ parameters

$$\begin{aligned} \mathcal{T} \equiv & (v_z^{(1)}, v_r^{(1)}, \Delta_\gamma^{(2)}, \dots, v_z^{(n-1)}, \\ & v_r^{(n-1)}, \Delta_\gamma^{(n)}, v_z^{(n)}, v_r^{(n)}, \gamma_e^{(n)}, z^{(n+1)}, r^{(n+1)}, \theta^{(n+1)}). \end{aligned} \quad (6)$$

B. Prior information

By choosing $\theta^{(1)}=0$, $\theta^{(n+1)}$ represents the change in azimuth of the whale, relative to the research vessel, between the points $E^{(1)}$ and $E^{(n+1)}$. One can combine the definitions of \mathcal{T} given in Eqs. (5) and (6) by writing

$$\begin{aligned} \mathcal{T} \equiv & (z^{(1)}, r^{(1)}, v_z^{(1)}, v_r^{(1)}, \gamma_b^{(1)}, v_z^{(2)}, v_r^{(2)}, \\ & \Delta_\gamma^{(2)}, \dots, v_z^{(n)}, v_r^{(n)}, \\ & \Delta_\gamma^{(n)}, z^{(n+1)}, r^{(n+1)}, \theta^{(n+1)}). \end{aligned} \quad (7)$$

This latter definition of \mathcal{T} uses $3n+5$ parameters, that is to say redundant information. Indeed, the coordinates of the last summit $E^{(n+1)}$ can be calculated given the coordinates of the first summit $E^{(1)}$ and using the values of speeds and change of heading in the n segments. The consequences of such redundancy during the estimation process will be discussed later. The aim of combining Eqs. (5) and (6) into Eq. (7) is to gather in a single definition of \mathcal{T} information on the fluking and resurfacing points of the whale.

Let $z^{(A)}=0$ m and $r^{(A)}$ be the depth and the horizontal range of the whale at the time $t=t^{(A)}$. Let $v_z^{(A)}$ and $v_r^{(A)}$ be the average vertical and horizontal speeds of the whale for $t \in [t^{(A)}, t^{(B)}]$. The depth and the range of the whale at $t=t^{(1)}$ are then $z^{(1)}=v_z^{(A)}(t^{(B)}-t^{(A)})$ and $r^{(1)}$. Let $z^{(D)}=0$ m and $r^{(D)}$ be the depth and the horizontal range of the whale at the time $t=t^{(D)}$. Let $v_z^{(D)}$ and $v_r^{(D)}$ be the average vertical and horizontal speeds of the whale for $t \in [t^{(C)}, t^{(D)}]$. The depth and the range of the whale at $t=t^{(n+1)}$ are then $z^{(n+1)}=v_z^{(D)}(t^{(C)}-t^{(D)})$ and $r^{(n+1)}$.

Sperm whales initiate their foraging dives by fluking-up and diving vertically (as observed from the sea surface at the

very beginning of the dive). Sperm whales usually keep a constant vertical speed while descending to reach their prey and while ascending to reach the sea surface (Miller *et al.*, 2004a). There are exceptions however: Sperm whales may, for instance, horizontally translate during the ascent likely due to the presence of conspecifics (Miller *et al.*, 2004b). Assuming that the main objective of the whale while descending is to reach bathypelagic prey and that the main objective of the whale while ascending is to reach oxygen at the sea surface, the whale would swim vertically for $t \in [t^{(A)}, t^{(B)}]$ and $t \in [t^{(C)}, t^{(D)}]$. This leads to $v_r^{(A)} \simeq 0$ m s⁻¹, $v_r^{(D)} \simeq 0$ m s⁻¹, $r^{(1)} \simeq r^{(A)}$, and $r^{(n+1)} \simeq r^{(D)}$. In this case $\theta^{(n+1)}$ represents the change in azimuth of the whale, relative to the research vessel, between the fluking and the resurfacing points. A visual observer can measure the parameters $z^{(1)}$, $r^{(1)}$, $z^{(n+1)}$, $r^{(n+1)}$, and $\theta^{(n+1)}$ from the research vessel. The visual measurement process is inaccurate and thus results in uncertainties on $z^{(1)}$, $r^{(1)}$, $z^{(n+1)}$, $r^{(n+1)}$, and $\theta^{(n+1)}$. The assumption of verticalness may also be inaccurate resulting in additional uncertainties on these variables. Such uncertainties are modeled in the following section using random variables.

C. Likelihood

The algorithm which will be described later is used to estimate the depth and the range of the whale during a dive, and requires the set of values of the surface echo delay of the clicks that the whale has emitted during this dive. Let K_e be the number of echolocation clicks that the whale has emitted for $t \in [t^{(B)}, t^{(C)}]$. The author assumes that $K \leq K_e$ clicks are correctly detected (both the direct path and the surface echo). The observer then makes K consistent measurements of the surface echo delay at the time $t \in \{t_1, \dots, t_K\}$ ($t_1=t^{(B)}$ and $t_K=t^{(C)}$), labeled

$$\mathcal{M} = (\tau(t_1), \dots, \tau(t_K)). \quad (8)$$

The trajectory model \mathcal{T} is close to the true trajectory that the whale follows. Let $f(\mathcal{T}, t_k)$ be the value at the time t_k ($k \in \{1, \dots, K\}$) of the surface echo delay if the whale were on the trajectory \mathcal{T} . Such value, however, due to the inaccuracy of the measurement and modeling processes, is not exactly equal to $\tau(t_k)$. The difference $\epsilon(\mathcal{T}, t_k)$ between the data $\tau(t_k)$ and the model $f(\mathcal{T}, t_k)$ is defined as

$$\tau(t_k) = f(\mathcal{T}, t_k) + \epsilon(\mathcal{T}, t_k) \quad \text{for } k \in \{1, \dots, K\}. \quad (9)$$

The author assumes that the errors due to such inaccuracies are centered (the mean of the error is equal to zero), independent [the error made on $\tau(t_i)$ is independent with the error made on $\tau(t_j)$, $i \neq j$], and of equal variance. In that case, one can model the above-described inaccuracies by interpreting them as an additive white Gaussian noise. Let ϵ be a centered, white Gaussian noise of standard deviation σ_τ , $\epsilon \sim \mathcal{N}(0, \sigma_\tau)$.

The noise on the data is modeled using the random variable ϵ . One can model the fluctuations in the values of the parameters previously defined using random variables. Let $\mathbf{E}^{(s)}$, $\mathbf{E}_p^{(s)}$, $\mathbf{z}^{(s)}$, $\mathbf{r}^{(s)}$, and $\boldsymbol{\theta}^{(s)}$ be the random variables corresponding to the parameters $E^{(s)}$, $E_p^{(s)}$, $z^{(s)}$, $r^{(s)}$, and $\theta^{(s)}$ (for

$s \in \{1, \dots, n+1\}$). Let $\mathcal{S}^{(s)}$, $\mathbf{v}_z^{(s)}$, $\mathbf{v}_r^{(s)}$, $\gamma_b^{(s)}$, $\gamma_e^{(s)}$, and $\Delta_\gamma^{(s)}$ be the random variables corresponding to the parameters $\mathcal{S}^{(s)}$, $v_z^{(s)}$, $v_r^{(s)}$, $\gamma_b^{(s)}$, $\gamma_e^{(s)}$ (for $s \in \{1, \dots, n\}$), and $\Delta_\gamma^{(s)}$ (for $s \in \{2, \dots, n\}$). Let \mathcal{T} , \mathcal{M} , and $\tau(t_k)$ be the random variables corresponding to the parameters \mathcal{T} , \mathcal{M} , and $\tau(k)$ (for $k \in \{1, \dots, K\}$). Using such definitions, relationship (9) can be rewritten $\tau(t_k) = f(\mathcal{T}, t_k) + \epsilon$. The likelihood of a single measurement $\tau(t_k)$ is then

$$p(\tau(t_k)|\mathcal{T}) = \frac{1}{\sqrt{2\pi\sigma_\tau^2}} \exp\left[-\frac{[\tau(t_k) - f(\mathcal{T}, t_k)]^2}{2\sigma_\tau^2}\right] \quad (10)$$

by noting $p(x|y) = p(\mathbf{x}=x|\mathbf{y}=y)$, the value at x of the conditional probability density function of the random variable \mathbf{x} given the value y of the random variable \mathbf{y} . The aforementioned assumptions on the noise ϵ lead to the following expression of the likelihood of the measurement set \mathcal{M} :

$$p(\mathcal{M}|\mathcal{T}) = \frac{1}{\sqrt{2\pi\sigma_\tau^2}} \exp\left[-\frac{1}{2\sigma_\tau^2} \sum_{k=1}^K [\tau(t_k) - f(\mathcal{T}, t_k)]^2\right]. \quad (11)$$

D. Independence between the variables

Using the definition of the trajectory given in Eq. (7), the prior probability of a trajectory \mathcal{T} is

$$p(\mathcal{T}) = p(z^{(1)}, r^{(1)}, v_z^{(1)}, v_r^{(1)}, \gamma_b^{(1)}, v_z^{(2)}, v_r^{(2)}, \Delta_\gamma^{(2)}, \dots, v_z^{(n)}, v_r^{(n)}, \Delta_\gamma^{(n)}, z^{(n+1)}, r^{(n+1)}, \theta^{(n+1)}), \quad (12)$$

by noting $p(x) = p(\mathbf{x}=x)$, the value at x of the probability density function of the random variable \mathbf{x} . In reality the $3n+5$ variables are statistically dependent, but one does not know what the dependency is. One needs to make some assumptions and simplifications with the aim of giving an analytical expression of the prior $p(\mathcal{T})$.

As mentioned earlier, the definition given in Eq. (7) is not minimal, and one can theoretically calculate $(\mathbf{z}^{(n+1)}, \mathbf{r}^{(n+1)}, \theta^{(n+1)})$ [related to the location of the resurfacing point] using $(\mathbf{z}^{(1)}, \mathbf{r}^{(1)})$ [related to the location of the fluking point] and $(\mathbf{v}_z^{(1)}, \mathbf{v}_r^{(1)}, \gamma_b^{(1)}, v_z^{(2)}, v_r^{(2)}, \Delta_\gamma^{(2)}, \dots, v_z^{(n)}, v_r^{(n)}, \Delta_\gamma^{(n)})$ [related to the underwater behavior of the whale]. The author, nevertheless, assumes from now on that the relationship among the fluking point, the resurfacing point, and the underwater behavior of the whale is too intricate to be comprehensible to an observer. In this case, there is no apparent prior relationship between $(\mathbf{z}^{(1)}, \mathbf{r}^{(1)})$, $(\mathbf{v}_z^{(1)}, \mathbf{v}_r^{(1)}, \gamma_b^{(1)}, v_z^{(2)}, v_r^{(2)}, \Delta_\gamma^{(2)}, \dots, v_z^{(n)}, v_r^{(n)}, \Delta_\gamma^{(n)})$, and $(\mathbf{z}^{(n+1)}, \mathbf{r}^{(n+1)}, \theta^{(n+1)})$, that is to say that these three sets of variables are assumed to be statistically independent.

One can assume that (1) the range of the whale when fluking up, (2) the vertical speed of the whale when initiating a dive, and (3) the period of time the whale waits before clicking are statistically independent, hence the independence between $\mathbf{r}^{(1)}$ and $\mathbf{z}^{(1)}$. One can also make similar assumptions regarding the final ascent, hence the independence among $\mathbf{r}^{(n+1)}$, $\mathbf{z}^{(n+1)}$, and $\theta^{(n+1)}$.

Finally, there must be a strong dependence between the parameters $\mathbf{v}_z^{(i)}$, $\mathbf{v}_r^{(i)}$ ($i \in \{1, \dots, n\}$), and $(\Delta_\gamma^{(2)}, \dots, \Delta_\gamma^{(n)})$ de-

scribing the underwater behavior of the whale. Such a relationship is however unknown to an observer. One can assume that the changes in vertical speed, horizontal speed, and heading are statistically independent. One can also assume that the speed of the whale in each segment is independent of the speed of the whale in any other segment. In view of the above-presented statements, the prior probability given in Eq. (12) can be rewritten

$$p(\mathcal{T}) = p(z^{(1)})p(r^{(1)})p(\gamma_b^{(1)})p(z^{(n+1)})p(r^{(n+1)})p(\theta^{(n+1)}) \times \left[\prod_{s=1}^n p(v_z^{(s)})p(v_r^{(s)}) \right] p(\Delta_\gamma^{(2)}, \dots, \Delta_\gamma^{(n)}). \quad (13)$$

E. Probability distribution of the prior

The author will give a more accurate description of the prior distributions appearing in Eq. (13). Using the Hammersley-Clifford theorem (Perez, 1998), choosing a joint probability density function ($\alpha > 0$)

$$p(\Delta_\gamma^{(2)}, \dots, \Delta_\gamma^{(n)}) \propto \exp\left[-\alpha \sum_{s=2}^n |\Delta_\gamma^{(s)}|\right] \quad (14)$$

ensures that $(\Delta_\gamma^{(2)}, \dots, \Delta_\gamma^{(n)})$ is a Markov chain. By choosing such a prior, the author indicates that the underwater movement of the whale is not erratic, and that it is likely that the whale tends to swim in a given direction. The strength of this constraint (later referred to as constraint of rectilinearity) increases with α .

One has no prior information on the heading of the whale when it starts to dive, hence $p(\gamma_b^{(1)}) = 1/2\pi$ for $\gamma_b^{(1)} \in [-\pi, \pi]$. The prior information regarding the values of $\mathbf{r}^{(1)}$, $\mathbf{z}^{(1)}$, $|\mathbf{v}_z^{(s)}|$, $\mathbf{v}_r^{(s)}$ ($s \in \{1, \dots, n\}$), $\mathbf{r}^{(n+1)}$, $\mathbf{z}^{(n+1)}$, and $\theta^{(n+1)}$ is introduced by assigning to these random variables bell-shaped probability density functions. The probability density functions have a maximum at the most likely prior value of the parameters, and its width illustrates the confidence we grant to this most likely prior value. One could choose for instance gamma probability distributions. For practical reasons, the author has chosen truncated normal distributions. Such probability distributions are denoted $\mathcal{N}_{[a,b]}(\mu, \sigma)$ in the following $[(a, b) \in \bar{\mathbb{R}}^2, a < b, \mu \in [a, b], \sigma > 0]$. Their probability density functions are equal to, for $x \in [a, b]$,

$$g(x) = f(x) \Big/ \int_a^b f(t) dt$$

with

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \quad (15)$$

The author assumes that $\mathbf{v}_z^{(s)}$ is positive or negative with an equal probability 1/2.

F. Probability distribution of the posterior

Using Bayes' formula (Robert and Casella, 2004), the posterior of the trajectory \mathcal{T} is known modulo a multiplicative constant

$$p(\mathcal{T}|\mathcal{M}) = \frac{p(\mathcal{M}|\mathcal{T})p(\mathcal{T})}{p(\mathcal{M})} \propto p(\mathcal{M}|\mathcal{T})p(\mathcal{T}). \quad (16)$$

Using Eqs. (11) and (13), one can calculate the posterior $p(\mathcal{T}|\mathcal{M})$ of the trajectory \mathcal{T} modulo a multiplicative constant. The mean $\hat{\mathcal{T}}$ of the posterior $p(\mathcal{T}|\mathcal{M})$ is the trajectory which best fits the data \mathcal{M} in the minimal mean square error (MMSE) sense (Robert and Casella, 2004).

G. MCMC algorithm

The analytical expression of the probability $p(\mathcal{T}|\mathcal{M})$ is known modulo a multiplicative constant. One cannot directly draw samples using Eqs. (11), (13), and (16). However, one can use a Markov Chain Monte Carlo (MCMC) algorithm to draw samples of the posterior. Such an algorithm can generate samples $\mathcal{T}[i]$ ($i \in \{I_1, \dots, I_2\}$, $(I_1, I_2) \in \mathbb{N}^2$, $I_2 > I_1$ distributed according to the posterior $p(\mathcal{T}|\mathcal{M})$). The mean of such samples

$$\hat{\mathcal{T}} = \frac{1}{I_2 - I_1 + 1} \sum_{i=I_1}^{I_2} \mathcal{T}[i] \quad (17)$$

is then an estimate of the optimal trajectory $\hat{\mathcal{T}}$ in the MMSE sense. One can use the Metropolis-Hastings algorithm (Robert and Casella, 2004) to draw such samples. Choose an initial value $\mathcal{T}[0]$ of the trajectory, and, at iteration i ($i \in \{1, \dots, I_2\}$), draw a trajectory sample $\mathcal{T}_d[i]$. Calculate the acceptance ratio of this sample to determine if the new sample $\mathcal{T}_d[i]$ is accepted ($\mathcal{T}[i] \equiv \mathcal{T}_d[i]$) or rejected ($\mathcal{T}[i] \equiv \mathcal{T}[i-1]$) [see Robert and Casella (2004) for a description on how to draw samples and how to calculate acceptance ratios]. The algorithm converges after I_1 iterations, and samples drawn for $i \geq I_1$ will be distributed according to $p(\mathcal{T}|\mathcal{M})$.

To accelerate convergence, one can do Gibbs sampling. Instead of drawing a whole new trajectory at each iteration i , draw the summits one by one. For $i \in \{1, \dots, I_2\}$ and for $s \in \{1, \dots, n+1\}$, draw a trajectory sample $\mathcal{T}_d^{(s)}[i]$ of \mathcal{T} . Calculate the acceptance ratio of this sample to determine if it is accepted ($\mathcal{T}^{(s)}[i] \equiv \mathcal{T}_d^{(s)}[i]$) or rejected [see Robert and Casella (2004) for a description of the Metropolis-Hastings algorithm using Gibbs sampling and try the software SBPLASH to see how the algorithm is implemented by the author]. The estimate of the trajectory is

$$\hat{\mathcal{T}} = \frac{1}{(n+1)(I_2 - I_1 + 1)} \sum_{i=I_1}^{I_2} \sum_{s=1}^{n+1} \mathcal{T}^{(s)}[i]. \quad (18)$$

Instead of using samples for $i \in \{I_1, \dots, I_2\}$ of a single large Markov chain to calculate $\hat{\mathcal{T}}$, one can use samples for $i \in \{I_1, \dots, I_2'\}$ of C smaller subchains, with $I_2' = I_1 + \Delta_I'$, $\Delta_I' = \Delta_I / C$, $\Delta_I = I_2 - I_1$. In theory, if convergence is reached (that is to say I_1 is large enough), each method will provide the same result. In practice, the second method requires a lower total amount of samples, as it requires less than Δ_I / C samples per subchain to lead to a correct estimate $\hat{\mathcal{T}}$. The author has used the subchain method. In this case, by label-

ing $\mathcal{T}^{(s)}[i][c]$ the trajectory samples ($c \in \{1, \dots, C\}$, $i \in \{1, \dots, I_2\}$, $s \in \{1, \dots, n+1\}$), the final estimate will be

$$\hat{\mathcal{T}} = \frac{1}{C(n+1)(I_2' - I_1 + 1)} \sum_{c=1}^C \sum_{i=I_1}^{I_2'} \sum_{s=1}^{n+1} \mathcal{T}^{(s)}[i][c]. \quad (19)$$

H. Data set

The author has run simulations using the free software SBPLASH implemented in MATLAB. Source and data files are available at <http://www.ensat.fr/pers/laplanche/sbplash>. The author did not have complete data (i.e., both trajectory and acoustic data) gathered during field experiments to test the algorithm. The efficiency of the algorithm is illustrated using synthetic data. The following trajectory was used in the simulations.

The whale flukes up and vertically dives at $t^{(A)}=0$ min, at a speed $v_z^{(A)}=1$ m s⁻¹, and at a range $r^{(A)}=500$ m west of the hydrophone. The whale starts clicking at $t^{(B)}=0.5$ min. The descent lasts 10 min. At $t=10$ min, the whale starts a 10 min swim, north at $v_r=1$ m s⁻¹ and upwards at $v_z=-0.5$ m s⁻¹. At $t=20$ min, the whale starts a 10 min swim, east at $v_r=1$ m s⁻¹ and downwards at $v_z=0.5$ m s⁻¹. At $t^{(C)}=30$ min, the whale stops clicking and starts a 10 min vertical ascent to the sea surface at $v_z=-1$ m s⁻¹. The whale resurfaces at $t^{(D)}=40$ min. The depth, range, vertical speed, and horizontal speed of the whale when following this trajectory are given in Figs. 3–6, respectively.

The acoustic activity of the whale is modeled by assuming that it emitted $K=1770$ clicks for $t \in [t^{(B)}, t^{(C)}]$ at 1 click/s with no interruption. The hydrophone is at a 70 m depth and the speed of sound is $c=1500$ m s⁻¹. Every click that the whale emitted is assumed to have been detected (both direct path and surface reflection) leading to K measurements of the surface echo delay τ . The measurement process of the delays is not free from errors (Wahlberg *et al.*, 2001). A centered white Gaussian noise of standard deviation $\sigma_\tau=1$ ms has been added to the τ data to simulate such inaccuracy. The noisy values of τ compose the data set denoted earlier as \mathcal{M} . Such values are plotted in Fig. 7.

The above-described trajectory is unknown to the observer. He knows however the data set \mathcal{M} and has some visual information on the fluking-up and resurfacing points. The range of the whale when fluking-up is measured as $r^{(A)}=500 \pm 25$ m [$\mathbf{r}^{(A)} \sim \mathcal{N}_{[0,+\infty]}(500, 25)$]. The vertical speed of the whale for $t \in [t^{(A)}, t^{(B)}]$ (when initiating a dive) is assumed to be equal to $v_z^{(A)}=1 \pm 0.1$ m s⁻¹ [$\mathbf{v}_z^{(A)} \sim \mathcal{N}_{[0,+\infty]}(1, 0.1)$]. The depth of the whale when it starts clicking is then $z^{(B)}=30 \pm 3$ m [$\mathbf{z}^{(B)} \sim \mathcal{N}_{[0,+\infty]}(30, 3)$]. The prior vertical and horizontal speeds of the whale for $t \in [t^{(B)}, t^{(C)}]$ follow the same truncated normal probability distribution $\mathcal{N}_{[0,+\infty]}(0, 0.75)$. The whale resurfaces at a range $r^{(D)}=630 \pm 25$ m [$\mathbf{r}^{(D)} \sim \mathcal{N}_{[0,+\infty]}(630, 25)$]. The vertical speed of the whale for $t \in [t^{(C)}, t^{(D)}]$ (when ending a dive) is assumed to be equal to $v_z^{(D)}=-1 \pm 0.1$ m s⁻¹ [$\mathbf{v}_z^{(D)} \sim \mathcal{N}_{[0,+\infty]}(1, 0.1)$]. The whale stops clicking at a depth $z^{(C)}=600 \pm 60$ m [$\mathbf{z}^{(C)} \sim \mathcal{N}_{[0,+\infty]} \times (600, 60)$]. The change of azimuth of the whale between

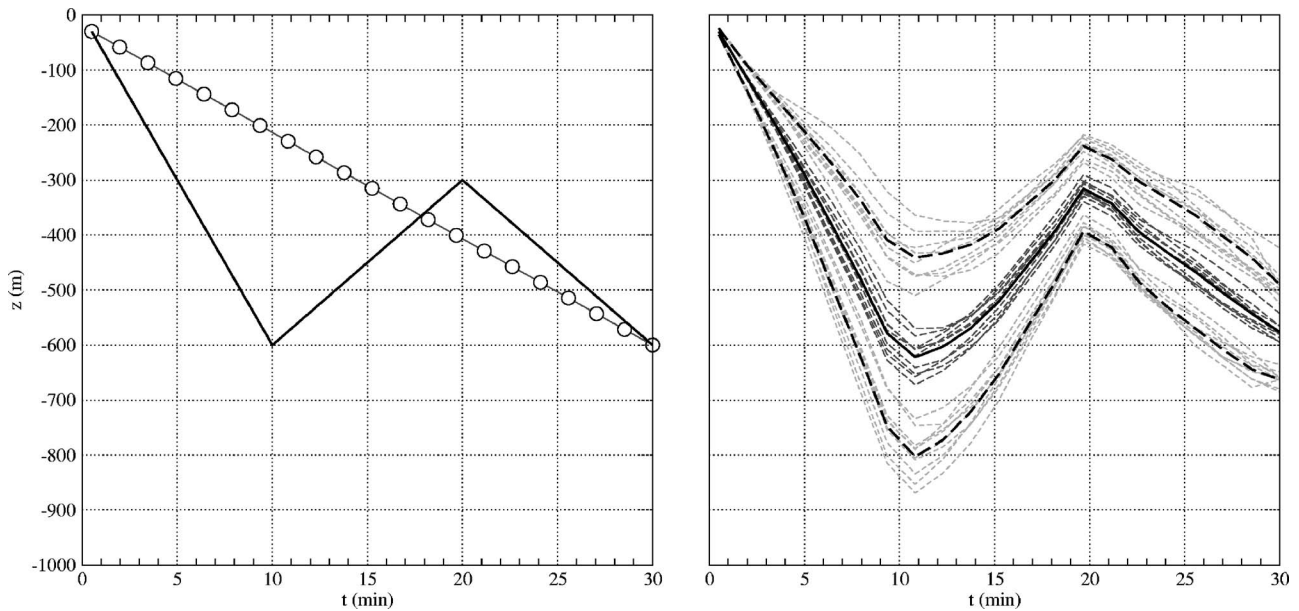


FIG. 3. True value (left, thick solid line) and initial value used in the MCMC algorithm (left, circles) of the depth z of the sperm whale throughout the dive. Estimation of the depth of the sperm whale (right, thick line) plus/minus twice its standard deviation (right, thick dashed line) for $t \in [t^{(B)}, t^{(C)}]$ using the data set presented in Fig. 7 and the ten 21 000-iteration Markov subchains. Partial estimations of the depth of the sperm whale (right, thin dark dashed lines) plus/minus twice their standard deviation (right, thin bright dashed lines) for each 21,000-iteration Markov subchain.

the fluking-up and resurfacing points is $\theta^{(n+1)} = \theta^{(D)} = 71.5 \pm 1^\circ$ [$\theta^{(n+1)} \sim \mathcal{N}_{[0,360]}(71.5, 1)$]. The choice of these priors is discussed later.

A MCMC algorithm theoretically converges no matter what initial value is chosen (Robert and Casella, 2004). The initial value of the trajectory is in the author's simulations the rectilinear, constant speed trajectory linking the prior locations of the points $E^{(1)}$ and $E^{(n+1)}$ (Figs. 3–6). The corresponding values of τ are given in Fig. 7. The trajectory is decomposed into $n=20$ segments. The parameters (α , σ_r and the parameters of the truncated normal distributions) could be estimated while estimating the trajectory parameters

(Robert and Casella, 2004). These parameters here are constants which are empirically chosen. The author has chosen $\alpha=2$ and $\sigma_\tau=10$ ms. Such a choice is discussed later.

III. RESULTS

A. Convergence

In this example, the algorithm generates trajectories with τ values close to the data after $I_0=100$ iterations. Samples are considered after $I_1=1000$ iterations. The following results required the use of $C=10$ Markov subchains, $I_2=21\,000$ iterations per chain ($\Delta_I=200\,000$). The acceptance

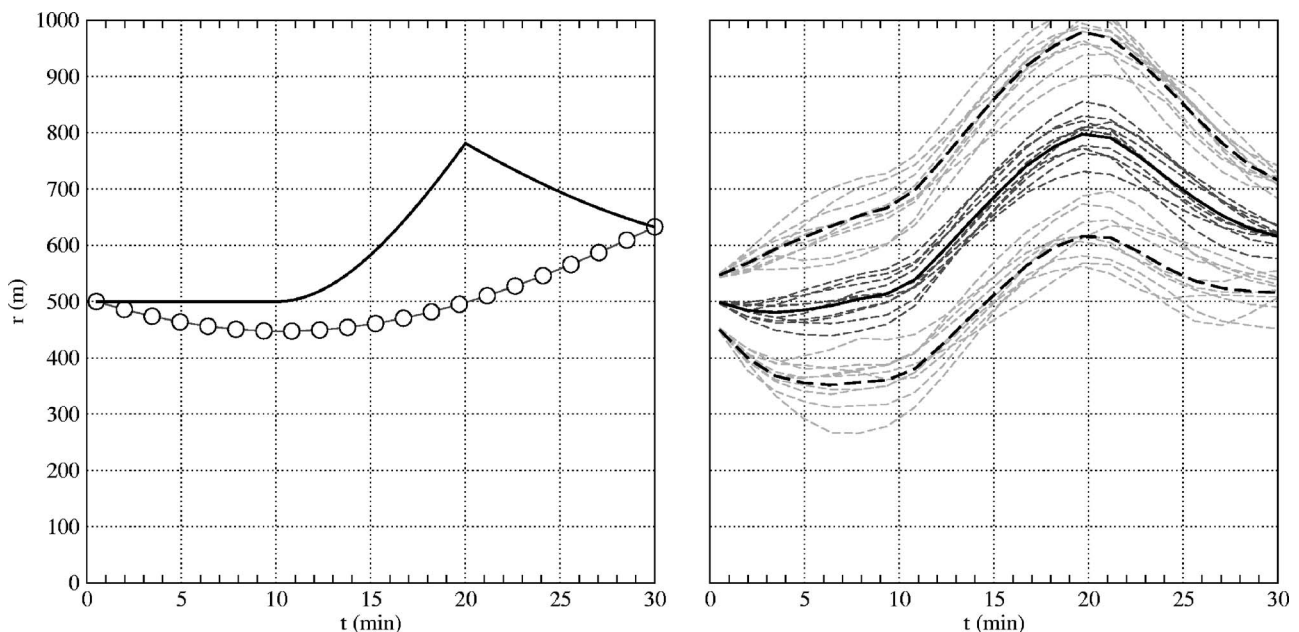


FIG. 4. True, initial, and estimated values of the range r of the sperm whale throughout the dive. See Fig. 3 for detailed comments.

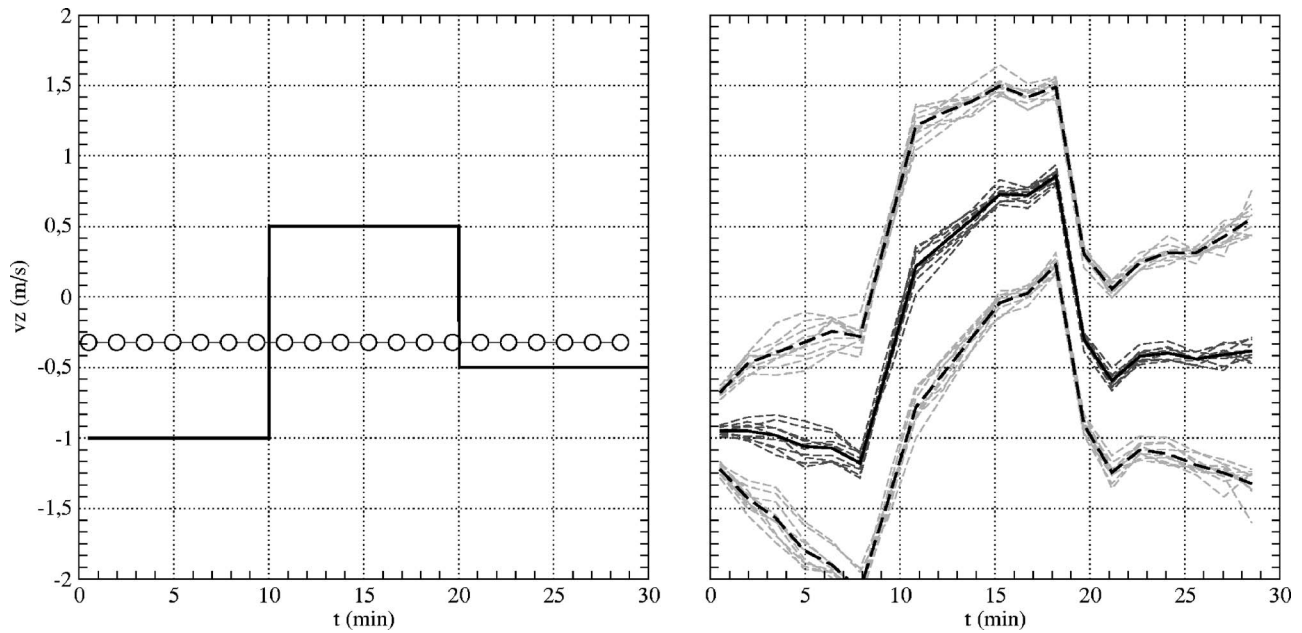


FIG. 5. True, initial, and estimated values of the vertical speed v_z of the sperm whale throughout the dive. See Fig. 3 for detailed comments.

rate is higher for $i < I_0$ and stabilizes for $i \geq I_0$. In the present test, the algorithm accepted 140 (standard deviation: 21.5) new summits every 100 iterations for $i \geq I_0$. Draws of the first ($s=1$) and last ($s=n+1$) summits are more often accepted than the others ($2 \leq s \leq n$), given the different constraints and priors they are bound to.

B. Depth, range, and speeds

The vertical and horizontal speeds of the whale are estimated using Eq. (19). Results are given in Figs. 5 and 6. The estimate of the depth and the range of the whale are calculated using these values. Results are given in Fig. 3 and 4. Results of v_z , v_r , z , and r are plotted plus/minus twice their

standard deviation. The standard deviation $\sigma_{\mathcal{T}}$ of the posterior $p(\mathcal{T}|\mathcal{M})$ is also estimated using the trajectory samples by writing

$$\hat{\sigma}_{\mathcal{T}}^2 = \frac{1}{C(n+1)(I'_2 - I_1 + 1)} \sum_{c=1}^C \sum_{i=I_1}^{I'_2} \sum_{s=1}^{n+1} (\mathcal{T}^{(s)}[i][c] - \hat{\mathcal{T}})^2. \quad (20)$$

Although such an estimator does not take into account any correlation existing between parameters or trajectory samples.

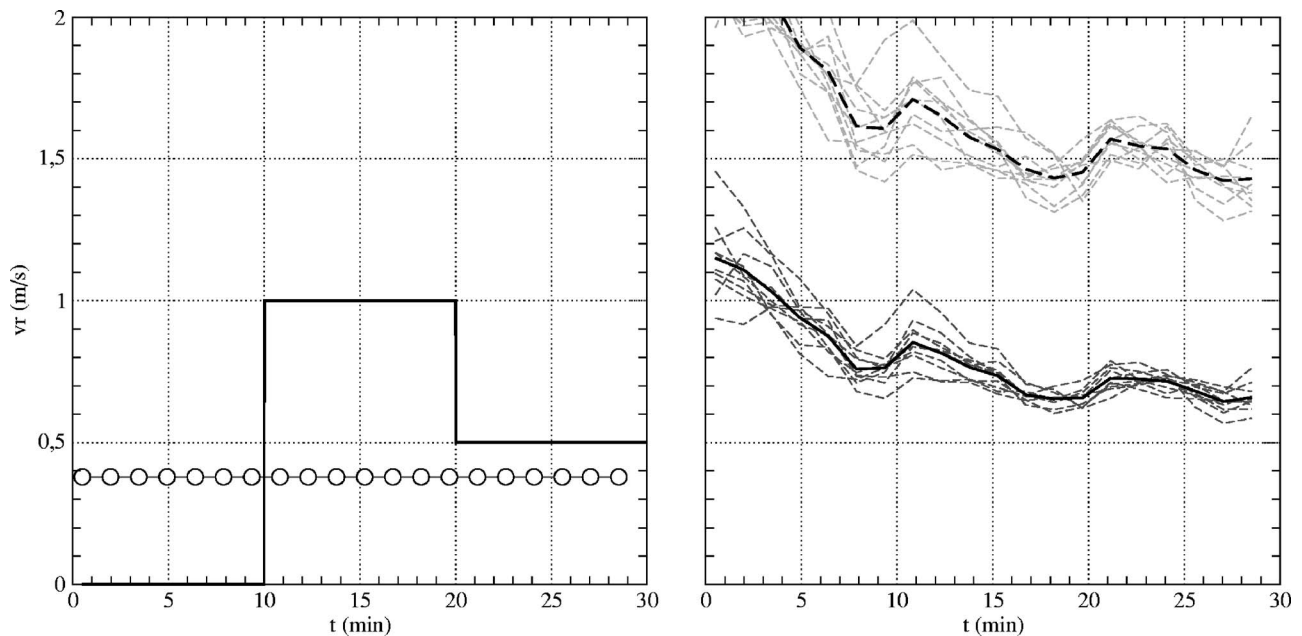


FIG. 6. True, initial, and estimated values of the horizontal speed v_r of the sperm whale throughout the dive. See Fig. 3 for detailed comments.

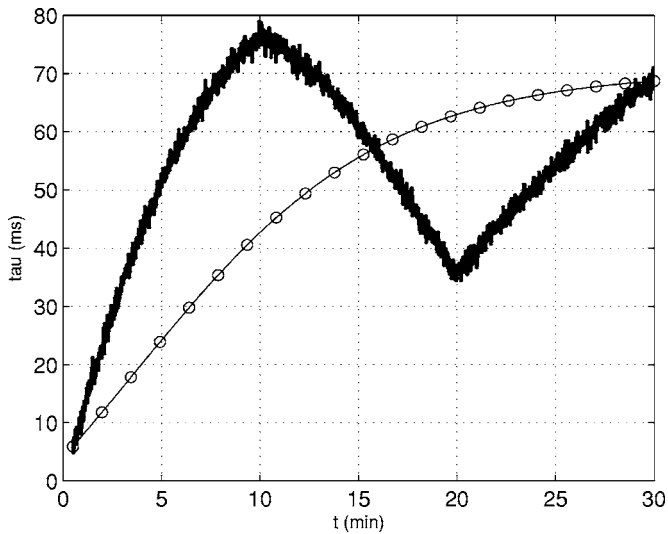


FIG. 7. Surface echo delay τ of the $K=1771$ clicks emitted by the sperm whale during the dive depicted in Figs. 3–6. The solid line describes the data set \mathcal{M} used in the tests. The circles represent the τ values of the trajectory used to initiate the algorithm.

IV. DISCUSSION AND CONCLUSION

The algorithm correctly estimates the depth and the range of the whale throughout the dive (Figs. 3 and 4). The estimates \hat{z} and \hat{r} differ on average from the true values by 24 and 34 m, respectively. The application of this method shows that, indeed, useful information on the depth and the range of the phonating sperm whale is contained in the values of the surface echo delays, and that a proper algorithm can obtain such information. The author has, however, tested the method using only synthetic data. The use of a four-piece trajectory, noisy delay data, uncertain visual cues and priors, and the quality of the results made this first test very promising. Nevertheless, confronting the algorithm with real data, for instance data gathered using far-field acoustic recordings plus digital tagging (Zimmer *et al.*, 2003), is still required for validation.

The quality of the estimation of v_r is however arguable (Figs. 5 and 6). One should find $v_z=1 \text{ m s}^{-1}$ and $v_r=0 \text{ m s}^{-1}$ for $t \in [0.5, 10]$, $v_z=-0.5 \text{ m s}^{-1}$ and $v_r=1 \text{ m s}^{-1}$ for $t \in]10, 20]$, and $v_z=0.5 \text{ m s}^{-1}$ and $v_r=1 \text{ m s}^{-1}$ for $t \in]20, 30]$. The estimate of v_z is correct, because the absolute vertical speed of the whale ($v_z^{(\text{abs})}$, in the terrestrial reference frame) is equal to the apparent vertical speed of the whale ($v_z^{(\text{app})}$, from the hydrophone point of view). The estimate of v_r is wrong, however, because the absolute horizontal speed of the whale ($v_r^{(\text{abs})}$) is not equal to the apparent radial horizontal speed of the whale ($v_r^{(\text{app})}$). One can gather information, using a single hydrophone, on $v_z^{(\text{app})}$ and $v_r^{(\text{app})}$. It is $v_z \equiv v_z^{(\text{abs})}$ and $v_r \equiv v_r^{(\text{abs})}$ however that are used in the trajectory model presented in Eq. (7). The estimate of $v_r^{(\text{abs})}$ is wrong simply because $v_r^{(\text{abs})}$ cannot be estimated using the τ values. Correct values of $v_z^{(\text{abs})}$ and $v_r^{(\text{abs})}$ could be used to reconstruct the trajectory of the whale in three dimensions. It does not seem possible, without making stronger and speculative hypotheses regarding the behavior of the whale (Laplanche *et*

al., 2005), or without using additional information (Tiemann *et al.*, 2006), to reconstruct a three-dimensional trajectory using a single hydrophone.

This Bayesian method has some weaknesses. The algorithm requires, to start, the whole set of values of the surface echo delay and the measurement of the location of the resurfacing point. Monte Carlo sampling is also very time consuming. The algorithm requires more time using a higher number (n) of segments ($3n+5$ parameters to estimate) and a higher number of iterations Δ_I . For such reasons, the algorithm cannot be applied in real-time. Monte Carlo sampling could however be replaced by a faster optimization algorithm of comparable effectiveness. Kalman filtering has been extensively used as a trajectory estimator in other fields of science. It could be interesting to use this method to filter the \mathcal{M} data set to reconstruct the trajectory of the whale.

The algorithm converges, no matter what initial value of the trajectory and what instrumental probability distributions are chosen (Robert and Casella, 2004). There is not, however, a strong theory defining when convergence is reached. The choice of I_1 and I_2' is empirical in this example. The author stopped the simulations (that is to say I_2' had been chosen) when the prior probability and the likelihood of the samples $\mathcal{T}^{(s)}[i][c]$ ($c \in \{1, \dots, C\}$, $s \in \{1, \dots, n+1\}$, $i \in \{1, \dots, I_2'\}$) moved around a stable mean for $\Delta_I=20\,000$ iterations. Samples for $i \in \{I_1, \dots, I_2'\}$ with $I_1=I_2'-\Delta_I'$ were considered in the estimation of the mean and the standard deviation of the posterior. It could be interesting to use a convergence diagnosis, like the criterion proposed by Gelman and Rubin (1992), to make a more rigorous choice of I_1 and I_2' . The author is, however, confident regarding the choice of I_1 for the given data set. A margin ($I_1 > I_0$) has been considered to make the choice of I_1 correct for different values of the parameters of the priors. The choice of I_2' is much more difficult to make. If I_1 is correctly chosen, the samples $\mathcal{T}^{(s)}[i][c]$ ($i \geq I_1$) are drawn according to the posterior of \mathcal{T} . But a given amount of samples $\hat{\Delta}_I$ is required to correctly sample the posterior. The use of $\Delta_I \ll \hat{\Delta}_I$ samples in Eq. (19) and (20) might lead to inaccurate estimates. Such a phenomenon is illustrated in Figs. 3 and 4. The use of $\Delta_I' = 20\,000$ samples is not enough to correctly estimate \hat{T} . The use of $\Delta_I=200\,000$ samples leads to a correct estimate.

Assuming that the chosen set of samples correctly represents the posterior, one can, by applying Eq. (19), find \hat{T} . The value \hat{T} describes the set of parameters which best fits the data in the MMSE sense. Depending on the shape of the posterior, there may be, however, differences between \hat{T} and the true value. This is a shortcoming inherent to Bayesian estimation techniques and this will not be discussed any further. Additional work is required to evaluate this issue regarding the method proposed by the author.

The parameters used in the probability distributions of the priors have been fixed and empirically chosen. The parameters (μ, σ) of the prior truncated normal distributions of $\mathbf{z}^{(1)}$, $\mathbf{r}^{(1)}$, $\mathbf{z}^{(n+1)}$, $\mathbf{r}^{(n+1)}$, and $\boldsymbol{\theta}^{(n+1)}$ have been chosen such that they represent, in the most realistic way, the prior information that one has on the values of these parameters. The standard deviations of such priors have been chosen great

enough so that the choice of their mean is not critical. If the assumption of verticalness of the beginning of the descent and the ascent proves to be inexact, larger standard deviations of these priors should be considered. Using information, accurate or not, on the fluking and resurfacing points is however very helpful. The author did not manage to estimate the depth and the range of the whale or the location of the resurfacing point using only the echo delay data and the constraint of rectilinearity.

The prior vertical and horizontal speeds of the whale for $t \in [t^{(B)}, t^{(C)}]$ follow the same truncated normal distribution $\mathcal{N}_{[0,+\infty)}(0, 0.75)$. The aim of this prior is to bring the information that it is highly unlikely that the speed of the whale is greater than a few m s^{-1} . Choosing a standard deviation of 0.5 m s^{-1} would not be realistic as it would lead to a small (5%) chance of speed greater than 1 m s^{-1} . Choosing a standard deviation of 1 m s^{-1} would not be reasonable either as it would lead to a small but not negligible (5%) chance of speed greater than 2 m s^{-1} . The mean choice of standard deviation (0.75 m s^{-1}) seems reasonable. This choice is not critical since values close to this mean lead to consistent results. However, choosing a too low (0.5 m s^{-1}) or a too high (1 m s^{-1}) standard deviation of the prior speed leads to inaccurate estimates of the depth and the range of the whale throughout the dive. As mentioned previously, one can sample and estimate this standard deviation while sampling and estimating the trajectory parameters. An optimal value of the standard deviation of the prior speed could be found this way.

The correct choice of σ_τ and α is important. The acceptance ratio is calculated using the posterior probability (Robert and Casella, 2004). The posterior probability is a function of the likelihood and the prior probability $p(\Delta_\gamma^{(2)}, \dots, \Delta_\gamma^{(n)})$ [Eq. (16)]. Both terms are exponentials modulo a multiplicative constant. The natural logarithm of the posterior is

$$\log[p(\mathcal{T}|\mathcal{M})] = -\frac{1}{2\sigma_\tau^2} \sum_{k=1}^K [\tau(t_k) - f(\mathcal{T}, t_k)]^2 - \alpha \sum_{s=2}^n |\Delta_\gamma^{(s)}| + \dots \quad (21)$$

The parameters σ_τ and α determine the relative importance (1) of the matching of the data \mathcal{M} and (2) of the constraint of rectilinearity, respectively. The smaller the value of σ_τ the greater the importance of the data. The larger the value of α the greater the importance of the constraint of rectilinearity. The value of σ_τ can be sensitively chosen to representing the similarity of the trajectory model to the truth. The choice of α is purely empirical however. The author here has chosen a value $\alpha=2$ which makes the importance of the constraint of rectilinearity relative to the constraint of the data (1) negligible for $i \in \{1, \dots, I_0\}$ and (2) comparable for $i \in \{I_1, \dots, I_2\}$. This ensures that (1) every trajectory sample fitting to the data (rectilinear or not) may be accepted when starting the simulation, and that (2) trajectories fitting to the data and rectilinear are more likely to be accepted than trajectories fitting to the data but less rectilinear. In the simulations, a lower value of α ($\alpha \leq 1$) made the constraint of rectilinearity negligible even after convergence. A large value of α ($\alpha \geq 5$) made the constraint of rectilinearity too

strong, and did not lead to a sensitive sampling of the posterior. Such statements are still empirical and qualitative, and more work is required to evaluate this issue. The value of α may also vary between sperm whales and dives depending on the click rate. As stated regarding the standard deviation of the prior speed, an optimal value of α could be estimated while sampling and estimating the trajectory parameters.

Nevertheless, having stated the weak points, the method also has many advantages. These first results are very encouraging. Not requiring seafloor echoes, the method could be used to estimate the depth and the range of foraging sperm whales using a single hydrophone in any location no matter what the depth, the relief, or the constitution of the seafloor might be. It could also be interesting to build a similar trajectory model and apply a Bayesian method and a Monte Carlo algorithm to combine TOAD data measured using an array of two (Thode, 2004) or more hydrophones.

ACKNOWLEDGMENTS

An earlier version of this work has been presented at the 150th meeting of the ASA. The author could not have participated in this meeting without the financial help of the Association Dirac. Such participation also greatly encouraged the author to carry out this project. The author thanks A. Alibert for her input on the abstract, both reviewers for the examination of the manuscript, and J. Y. Tournet for providing constructive comments. The author is also very grateful to M. Lonergan for reviewing an earlier version of the manuscript and K. Hunter, R. Allsopp, and M. Adam for refining the English.

- Backus, R., and Schevill, W. (1966). "Physeter clicks," in *Whales, Dolphins and Porpoises*, edited by K. Norris (University of California Press, Berkeley), pp. 510–527.
- Gelman, A., and Rubin, D. (1992). "Inference from iterative simulation using multiple sequences." *Stat. Sci.* **7**, 457–511.
- Laplanche, C., Adam, O., Lopatka, M., and Motsch, J. (2005). "Male sperm whale acoustic behavior observed from multipaths at a single hydrophone." *J. Acoust. Soc. Am.* **118**, 2677–2687.
- Miller, P., Johnson, M., and Tyack, P. (2004a). "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture." *Proc. R. Soc. London, Ser. B* **271**, 2239–2247.
- Miller, P., Johnson, M., Tyack, P., and Terray, E. (2004b). "Swimming gaits, passive drag and buoyancy of diving sperm whales *Physeter macrocephalus*." *J. Exp. Biol.* **207**, 1953–1967.
- Møhl, B., Wahlberg, M., Madsen, P., Heerfordt, A., and Lund, A. (2003). "The monopulsed nature of sperm whale clicks." *J. Acoust. Soc. Am.* **114**, 1143–1154.
- Møhl, B., Wahlberg, M., Madsen, P., Miller, L., and Surlykke, A. (2000). "Sperm whale clicks: Directionality and source level revisited." *J. Acoust. Soc. Am.* **107**, 638–648.
- Perez, P. (1998). "Markov random fields and images." *CWI Q.* **11–4**, 413–437.
- Robert, C., and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. (Springer, New York).
- Spiesberger, J. (2001). "Hyperbolic location errors due to insufficient numbers of receivers." *J. Acoust. Soc. Am.* **109**, 3076–3079.
- Spiesberger, J. (2005). "Probability distributions for locations of calling animals, receivers, sound speeds, winds, and data from travel time differences." *J. Acoust. Soc. Am.* **118**, 1790–1800.
- Thode, A. (2004). "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array." *J. Acoust. Soc. Am.* **116**, 245–253.
- Thode, A., Mellinger, D., and Stienessen, S. (2002). "Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico." *J. Acoust. Soc. Am.* **112**, 308–321.

- Tiemann, C., Thode, A., Straley, J., O'Connell, V., and Folkert, K. (2006). "Three-dimensional localization of sperm whales using a single hydrophone," *J. Acoust. Soc. Am.* **120**, 2355–2365.
- Urick, R. (1983). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
- Villadsgaard, A., Wahlberg, M., and Tougaard, J. (2007). "Echolocation signals of wild harbour porpoises, *Phocoena phocoena*" *J. Exp. Biol.* **210**, 56–64.
- Wahlberg, M., Møhl, B., and Madsen, P. (2001). "Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics," *J. Acoust. Soc. Am.* **109**, 397–406.
- Watkins, W., and Schevill, W. (1972). "Sound source location with a three-dimensional hydrophone array," *Deep-Sea Res.* **19**, 691–706.
- Whitehead, H., and Weilgart, L. (1990). "Click rates from sperm whales," *J. Acoust. Soc. Am.* **87**, 1798–1806.
- Zimmer, W., Johnson, M., D'Amico, A., and Tyack, P. (2003). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*)," *IEEE J. Ocean. Eng.* **28-1**, 13–28.

Toward bias minimization in acoustic feedback cancellation systems

Christos Boukis

Autonomic and Grid Computing Group, Athens Information Technology, Athens 19002, Greece

Danilo P. Mandic and Anthony G. Constantinides

Communications and Signal Processing Group, Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2BT United Kingdom

(Received 19 June 2006; revised 11 December 2006; accepted 12 December 2006)

A novel technique for bias suppression within acoustic feedback cancellation systems is proposed. This is achieved based on the use of all-pass filters in the forward part of the hearing aid. The poles of these filters are made time-varying, which results in a frequency response with constant magnitude and varying phase. This is a desired feature of the proposed approach, since the results from human psychoacoustics show that the human ear is not sensitive to moderate phase perturbations. The derivation of the proposed algorithms for the time variation of the location of the poles of all pass filters is based on a rigorous analysis of the phenomenon of bias in acoustic systems. Practical issues, such as the dependence of the steady-state error on the order of the all-pass filter, the number of varying poles, and their standard deviation are examined and strategies for the variation of the poles are introduced. Results obtained from a simulated hearing aid are provided to support the analysis. The quality of the processed audio signals is evaluated through subjective tests. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431341]

PACS number(s): 43.60.Mn, 43.66.Ts, 43.60.Ac [EJS]

Pages: 1529–1537

I. INTRODUCTION

A major drawback of hearing aids is the acoustic feedback from their receiver to their microphone, whereby a part of the acoustic signal emitted from the receiver propagates through the ventilation duct and is recaptured by the microphone.¹ This audio signal is then again processed and transmitted, which causes an acoustic feedback, a phenomenon detrimental to the performance which results in *echoes* and *howling*.² The suppression of this feedback is therefore crucial. In order to improve the quality of the emitted audio signal and to increase the maximum allowable gain (MAG).³

The most efficient current approach for feedback cancellation is based on the modeling of the feedback path with a finite impulse response (FIR) filter in order to produce feedback estimates which are subsequently subtracted from the microphone signal.^{4,5} To cope with the variability of the acoustic feedback path, least squares (LS) stochastic gradient descent (SGD) algorithms have been used^{6,7} for the training of this FIR filter. Despite their widespread use, these algorithms provide biased feedback estimates^{8–10} since

- (1) the data used for the optimization of the filter coefficients are collected in closed loop, which renders the standard independence assumptions ineffective,¹¹
- (2) the actual estimation error, that is, the difference between the feedback signal and its estimate, is not readily available.

Alternatively, we can use techniques based on the decorrelation between the input and the output of a hearing aid. These result in more accurate estimates of the feedback path. This decorrelation can be achieved either by disconnecting the forward path of the hearing aid when the coefficients of the digital filter are being adapted,¹² or by introducing delays

in the forward and/or the cancellation path of a hearing aid.¹³ The former technique implies noncontinuous adaptation of the coefficients. Its major drawback is the need to estimate instances when the values of the coefficients of the filter need to be updated. Moreover, there is a trade-off between the adaptation time and the achieved perceptual quality. The introduction of delays, on the other hand, implies a continuous adaptation of the coefficients of the adaptive filter. The degree of bias suppression critically depends on both, the statistics of the input signal and the properties of the feedback path.⁸

To that cause we propose a novel bias suppression technique based on the use of an all-pass filter with time-varying coefficients in the forward path of an acoustic feedback cancellation system.¹⁴ We show, both analytically and through simulations, that this technique has the potential to achieve greater bias reduction than the existing approaches, and that it is more robust to the changes in the statistics of the input signal and the characteristics of the feedback path.¹⁵ Practical issues, such as the effect of the parameters of the all-pass filter, like its order and the number of the varying poles, on the performance of the acoustic feedback cancellation are evaluated and several approaches for the variation of the filter poles are examined. Simulation results on both recorded and synthetic signals are provided to support the analysis.

II. HEARING AID DESCRIPTION

A hearing aid, whose block diagram is shown in Fig. 1, typically consists of the forward path $A(z)$, whose objective is to amplify appropriately the captured audio signals, the acoustic feedback path $G(z)$ which propagates the receiver's signal to the microphone and the feedback cancellation path

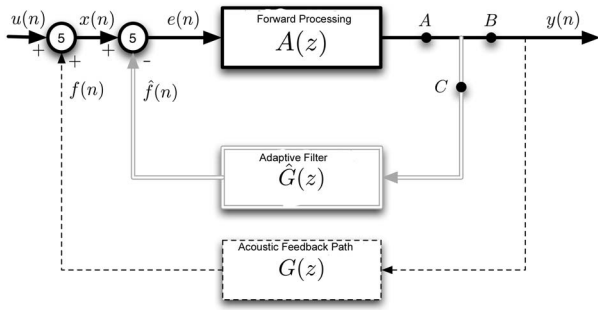


FIG. 1. Block diagram of a hearing aid with an integrated feedback cancelling adaptive filter. The dashed line indicates the inaccessible part of the system, i.e., the acoustic path.

$\hat{G}(z)$ which estimates the parameters of the feedback path.¹³ For convenience, in our analysis the forward acoustic channel¹⁶ is neglected. The involved signals are the external input denoted by $u(n)$, the feedback and its estimate $f(n)$ and $\hat{f}(n)$, respectively, the microphone signal $x(n)$, the signal after feedback removal $e(n)$, and the output of the receiver $y(n)$.

The most significant components of the processing unit $A(z)$ of the forward path are the automatic gain control (AGC) unit, which is adjusted according to the audiogram of the user, and the amplifier. A delay z^{-d} , where $d \geq 1$, is also introduced so as to avoid having a closed loop without a delay, a mathematically ill-posed problem. Since the AGC and the transfer functions of the microphone and the receiver have known and fixed values,¹⁷ our focus is solely on the identification of the acoustic feedback path.¹⁶ Thus in our simulations we assume that $A(z) = A_o z^{-d}$.

III. ADAPTIVE FEEDBACK CANCELLATION

Since a part of the signal produced by the receiver leaks back to the microphone, the actual transfer function of a hearing aid,

$$T(z) \triangleq \frac{y(z)}{u(z)} = \frac{A(z)}{1 - A(z)G(z)}, \quad (1)$$

differs substantially from its intended transfer function $A(z)$. A common approach for the suppression of the undesired feedback signal is to include an FIR filter in the hearing aid, placed in parallel to the feedback path as illustrated in Fig. 1. The order of this feedback cancelling filter should be large enough to avoid the undermodeling of the feedback path, and its output is given by

$$\hat{f}(n) = \sum_{i=0}^{N-1} \hat{g}_i(n)y(n-i) = \hat{\mathbf{g}}^t(n)\mathbf{y}(n), \quad (2)$$

where $\hat{\mathbf{g}}(n) = [\hat{g}_0(n), \hat{g}_1(n), \dots, \hat{g}_{N-1}(n)]^t$ are its coefficients, $\mathbf{y}(n) = [y(n), y(n-1), \dots, y(n-N+1)]^t$ is the output regressor vector and $(\cdot)^t$ the matrix transpose operator. The transfer function of the hearing aid now becomes

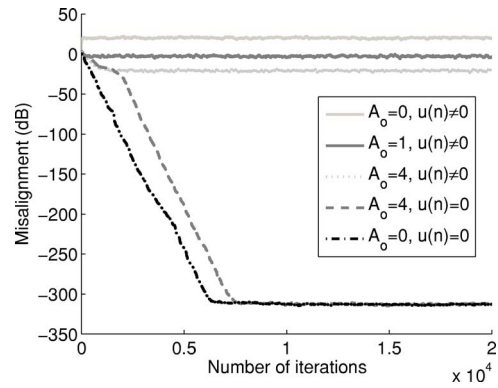


FIG. 2. The performance of the adaptive feedback cancelling filter for five distinct cases.

$$T(z) = \frac{A(z)}{1 - A(z)[G(z) - \hat{G}(z)]}. \quad (3)$$

To cope with the variability of the acoustic feedback path, the coefficients of the feedback cancelling filter are adapted with a SGD algorithm¹⁸ that aims at the minimization of the square of the error $e(n) = u(n) - \hat{f}(n)$, resulting in¹²

$$\hat{\mathbf{g}}(n+1) = \hat{\mathbf{g}}(n) + \mu e(n)\mathbf{y}(n), \quad (4)$$

where μ is a constant learning rate. Notice that for the derivation of (4) the dependence of the output $y(n)$ on the coefficients of the adaptive filter is deliberately neglected. This is, strictly speaking, not correct, especially for large learning rate values.¹⁹ In order to facilitate adaptation a low-power noisy signal $w(n)$ can be added to the output $y(n)$, yielding

$$y(n) = A_o e(n-1) + w(n). \quad (5)$$

Thus, when the data are collected in a closed loop, the optimal solution in the LS sense is given by¹³

$$\hat{\mathbf{g}}_{cl}^* = E\{\mathbf{y}(n)\mathbf{y}'(n)\}^{-1}[E\{u(n)\mathbf{y}(n)\} + E\{f(n)\mathbf{y}(n)\}] \quad (6)$$

and by

$$\hat{\mathbf{g}}_{ol}^* = E\{\mathbf{w}(n)\mathbf{w}'(n)\}^{-1}E\{f(n)\mathbf{w}(n)\} \quad (7)$$

when the data are collected in open loop, that is for interrupted forward path $A(z) = 0$. Comparing (6) with (7) and noting that SGD algorithms converge in the mean to unbiased solutions when they perform in open loop,²⁰ the bias of the feedback cancelling filter in the steady state becomes

$$\mathbf{q}^* = \mathbf{g}_{cl}^* - \mathbf{g}_{ol}^*. \quad (8)$$

During the derivation of (6) and (7) it was assumed that the input $u(n)$ and the injected noise $w(n)$ are uncorrelated. The dependence of the output $y(n)$ on the coefficients of the feedback cancelling filter was also neglected in the derivation of (6).

The misalignment curves of the adaptive feedback cancelling filter for several values of the amplification A_o and both in the presence and absence of the input signal $u(n)$ are shown in Fig. 2. Observe that the steady-state error depends primarily on the external input $u(n)$ and to a lesser extent on the gain A_o [it is a function of A_o only when $u(n) \neq 0$]. The value of gain A_o , however, affects the settling time. The

acoustic feedback path was approximated by the first 100 samples of the impulse response of ear canal measured on a KEMAR mannequin,²¹ and for its identification an FIR filter of the same order was employed. The normalized LMS (NLMS) algorithm was employed for the adaptation of the filter coefficients (with learning rate $\mu=1$). The input $u(n)$ was a zero mean and unit variance random signal with Gaussian distribution. To facilitate adaptation 60 dB of white noise were added to the output $y(n)$.

IV. BIAS REDUCTION WITH TIME-VARYING ALL-PASS FILTERS

From the analysis of the previous section it is concluded that when operating in closed loop the adaptive filter produces biased estimates of the acoustic feedback path. As a consequence the acoustic feedback is not removed completely; a residual feedback signal is always present. Reducing the bias in the estimates of the acoustic channel results in reduced residual feedback.

A straightforward approach toward bias suppression, and thus more efficient feedback cancellation, is to interrupt the forward path ($A_o=0$) when adapting the filter coefficients. Moreover, if the coefficients of the filter are adapted in the absence of input then unbiased feedback estimates are produced.⁵ However, this method is not preferable in acoustic feedback cancellation systems and especially in hearing aids, since the procedure of interrupting the forward path and feeding the receiver with noise might be very irritating to the user.

Bias suppression in continuously adapting feedback cancelling systems can be accomplished by introducing delays in the forward or the cancellation path of the hearing aid.^{13,22} The fundamental idea of this approach is to decorrelate the signal $u(n)$ from the output $y(n)$. These decorrelating delays can be placed in the points A, B , or C of the hearing aid depicted in Fig. 1 and they can be either constant or time varying.¹⁹ Delays in the forward path (point A or B in Fig. 1) suppress the bias only for colored inputs $u(n)$. Delays in the cancellation path (point C in Fig. 1) compensate for the inherent delay in the acoustic feedback path. Thus bias suppression is successful, irrespective of the characteristics of the input signal. If the introduced delay is larger than the actual delay within the feedback path, this results in an increase instead of reduction of the steady-state error,²³ and pre-echo or “comb”-filter effects can be observed.²⁴

To this end we propose an approach based on all-pass filters with time-varying poles in the forward path, for which the transfer function is given by

$$H_M(z, \alpha(n)) = \frac{z^{-M} + \sum_{i=1}^M r_i(n)z^{M-i}}{1 + \sum_{i=1}^M r_i(n)z^{-i}} = \prod_{i=1}^M \frac{z^{-1} - \alpha_i(n)}{1 - \alpha_i(n)z^{-1}}, \quad (9)$$

where $\alpha(n)=[\alpha_1(n), \alpha_2(n), \dots, \alpha_M(n)]^t$ are the values of the varying poles at time instant n , $r_i(n)$ the coefficient values ($i=1, 2, \dots, M$) and M the order of the filter as shown in Fig. 3.

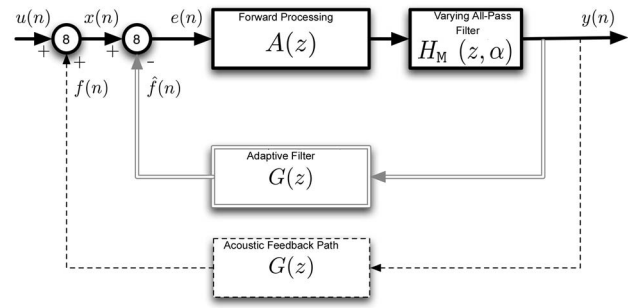


FIG. 3. Block diagram of a hearing aid with a feedback cancelling adaptive filter and a varying all-pass filter in the forward path.

The transfer function of the proposed system from Fig. 3 in the Z domain becomes

$$T(z, n) = \frac{A(z)H_M(z, \alpha(n))}{1 - A(z)H_M(z, \alpha(n))[G(z) - \hat{G}(z)]}. \quad (10)$$

Since the all-pass filter $H_M(z, \alpha(n))$ is time varying the transfer function of the system will be varying even when the adaptive filter reaches steady state, introducing some form of “controlled” nonstationarity to the system.

The output of the system is no longer given by

$$y(n) = A_o e(n - 1),$$

but instead by the recursive equation

$$y(n) = - \sum_{i=1}^M r_i(n)y(n - i) + A_o e(n - M - 1) + A_o \sum_{i=1}^M r_i(n)y(n - M - 1 + i) \quad (11)$$

which results in an increase of the computational complexity by $2M$ multiplications and $2M$ additions. For example, when a first-order all-pass filter is used, we have

$$y(n) = \alpha(n)y(n - 1) + A_o e(n - 2) - A_o \alpha(n)e(n - 1). \quad (12)$$

Equation (12) can be written in a vector-matrix form as follows:

$$y(n) = \mathbf{A}(n)y(n - 1) + A_o e(n - 2) - A_o \mathbf{A}(n)e(n - 1), \quad (13)$$

where $\mathbf{A}(n)$ is an $(N \times N)$ diagonal matrix given by

$$\mathbf{A}(n) = \begin{bmatrix} \alpha(n) & 0 & \dots & 0 \\ 0 & \alpha(n - 1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha(n - N + 1) \end{bmatrix}. \quad (14)$$

This method is in agreement with an earlier result⁹ where it is stated that “The system can be made identifiable with $w(n)=0$ if the signal processing unit $A(z)$ is nonlinear or time-variant in such a way that the autocorrelation matrix of the output $y(n)$ becomes nonsingular.” Namely, since the magnitude of the frequency response of this all-pass filter is flat and also remains constant, pole variations result in phase variations. This method can therefore be seen as an attempt to decorrelate the input from the output of a hearing aid by

adding noise to their phase. An advantage of this approach is the preservation of the perceptual quality, since the human ear is insensitive to low-scale phase changes.²⁵

A. Steady-state analysis

The introduction of the all-pass filter $H_M(z, \alpha(n))$ in the forward path of the system, modifies the output signal $y(n)$, rendering it more “random” and the corresponding correlation matrices nonsingular. Since the transfer function of the system is time varying, even when the adaptive filter has reached steady state, its output $y(n)$ will be nonstationary and the steady-state analysis provided in Sec. III only applies

$$\begin{aligned} E\{y(n)y^t(n)\} &= E\{A(n)y(n-1)y^t(n-1)A(n)\} + A_o E\{e(n-2)y^t(n-1)A(n)\} - A_o E\{A(n)e(n-1)y^t(n-1)A(n)\} \\ &\quad + A_o E\{A(n)y(n-1)e(n-2)\} + A_o^2 E\{e(n-2)e^t(n-2)\} - A_o^2 E\{A(n)e(n-1)e^t(n-2)\} - A_o E\{A(n)y(n-1)e(n-1)A(n)\} \\ &\quad - A_o^2 E\{e(n-2)e(n-1)A(n)\} + A_o^2 E\{A(n)e(n-1)e(n-1)A(n)\}. \end{aligned} \quad (17)$$

From these equations no direct conclusion about the effect of the introduced all-pass filter, or the variation of its poles, on the steady-state error can be drawn. Intuitively, we expect a decrease in the value of the terms that contain the random pole, which is proportional to its variance.

B. Pole variation

Three approaches for the variation of the poles of the all-pass filter within the forward path are proposed. A simple approach would be to make poles $\alpha_i(n)$ vary randomly around a fixed value $\alpha_{o,i}$ according to¹⁴

$$\alpha_i(n) = \alpha_{o,i} + \sigma_i v_i(n), \quad (18)$$

where $v_i(n)$ is a stochastic process with zero mean, unit variance and Gaussian or uniform distribution, and σ_i is a constant that specifies the variance of the i th pole.

Alternatively, a recursive formula can be employed, based on a convex combination of the previous pole value and its random displacement, which is given by

$$\alpha_i(n+1) = \lambda \alpha_i(n) + (1-\lambda) \sigma_i v_i(n), \quad (19)$$

where σ_i and $v_i(n)$ have the same meaning as in (18) and $\lambda \in (0, 1)$ is a convex parameter that controls the “randomness” within the update of the pole location.

Finally, in order to minimize the square of the cross correlation $\xi(n)$ between the input $u(n)$ and the output $y(n)$ of the hearing aid we may apply a stochastic gradient descent approach, that updates the poles $\alpha_i(n)$ toward the direction (Appendix A)

approximately due to the time-varying nature of the optimal coefficient. For example, when a first-order all-pass filter is introduced in the forward path the terms $E\{u(n)y(n)\}$, $E\{f(n)y(n)\}$, and $E\{y(n)y^t(n)\}$ become, respectively,

$$\begin{aligned} E\{u(n)y(n)\} &= E\{u(n)A(n)y(n-1)\} + A_o E\{u(n)e(n-2)\} \\ &\quad - A_o E\{u(n)A(n)e(n-1)\}, \end{aligned} \quad (15)$$

$$\begin{aligned} E\{f(n)y(n)\} &= E\{f(n)A(n)y(n-1)\} + A_o E\{f(n)e(n-2)\} \\ &\quad - A_o E\{f(n)A(n)e(n-1)\}, \end{aligned} \quad (16)$$

and

$$\begin{aligned} \frac{\partial \xi(n)}{\partial \alpha_i(n)} &= \left(- \sum_{k=1}^M \frac{\partial r_k(n)}{\partial \alpha_i(n)} y(n-k) + A_o e(n-M) \right. \\ &\quad \left. + A_o \sum_{k=1}^M \frac{\partial r_k(n)}{\partial \alpha_i(n)} e(n+M-k) \right) e^2(n)y(n) \end{aligned} \quad (20)$$

for $i=1, 2, \dots, N$, where $r_i(n)$ are the coefficients and $\alpha_i(n)$ the poles of the introduced all-pass filter (Eq. (9)). Since the magnitude of the poles of the all-pass filter should be less than unity (for stability reasons), the following hard-bounding formula is applied:

$$\alpha_i(n+1) = \begin{cases} 0.9 & \alpha_i(n+1) > 0.9, \\ \alpha_i(n+1) & 0.9 \geq \alpha_i(n+1) \geq -0.9, \\ -0.9 & \alpha_i(n+1) < -0.9. \end{cases} \quad (21)$$

A comparison of the effect of the pole updating formula on the performance of the adaptive feedback cancelling filter is given in Fig. 4. Observe that the use of adaptive poles

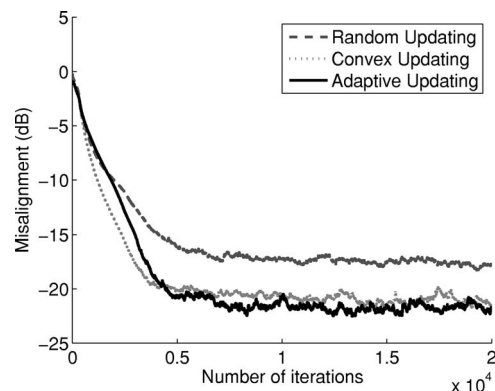


FIG. 4. Effect of the pole updating rule on the convergence behavior of the adaptive filter.

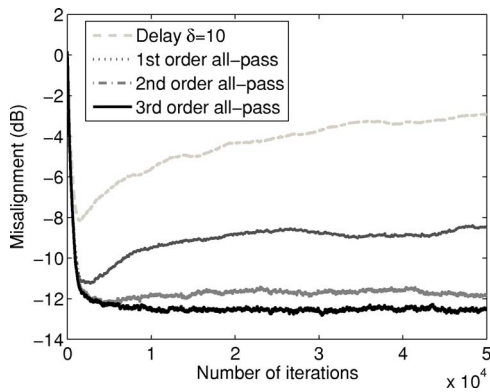


FIG. 5. Misalignment curves for speech-shaped input signals for a delay and an all-pass filter in the forward path.

results in lower steady-state error. Notice that all-pass filters with adaptive poles do not require any *a priori* information since their coefficients converge provided a small step size. Moreover, they have the ability to cope with nonstationary environments more effectively. For the derivation of this graph a first-order all-pass filter with the following parameter values was used: $\alpha(0)=\alpha_o=-0.2$, $\lambda=0.2$, $\eta=10^{-8}$, $\sigma=0.2$, and $v(n)\sim\mathcal{N}(0,1)$.

V. SIMULATIONS AND DISCUSSION

Two sets of simulations were conducted with the aim to (i) illustrate the benefits of the proposed bias reduction approach over existing methods and (ii) examine the effect of the parameters of the introduced all-pass filter on the performance of the feedback cancelling filter.

A. Experimental setup

The forward path processing unit had a transfer function of the form $A(z)=A_o z^{-1}$, with $A_o=4$. The acoustic feedback path was approximated by the first 100 samples of the impulse response of an ear canal measured on a KEMAR mannequin and sampled at 22 kHz. To avoid the situation of undermodeling, the feedback cancelling filter was an FIR filter of the same order. For the adaptation of its coefficients the NLMS algorithm was employed, with $\mu=0.1$. The misalignment was used as a metric for the evaluation of the performance of the adaptive filter, defined as

$$v(n) = \sqrt{[\hat{\mathbf{g}}(n) - \mathbf{g}_o]^T [\hat{\mathbf{g}}(n) - \mathbf{g}_o]}, \quad (22)$$

where $\hat{\mathbf{g}}(n)$ is the coefficient vector of the adaptive filter and \mathbf{g}_o are the samples of the impulse response of the acoustic feedback path. Notice that since it is assumed that the acoustic feedback path can be adequately modeled by this adaptive filter, both vectors had the same length.

B. Varying all-pass filter vs delay in the forward path

The convergence behavior of the adaptive filter was assessed for the cases of a delay and an all-pass filter in the forward path for colored noise (Fig. 5), white noise (Fig. 6), and speech (Fig. 7) input. The transfer function of the all-pass filter is given by (9). Unless stated otherwise, its poles

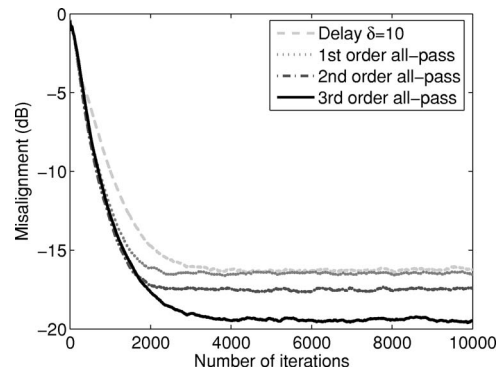


FIG. 6. Misalignment curves for white noise input for the cases of forward path delay and all-pass filter.

were varying according to (20), with $\eta=10^{-8}$, $\alpha_i(0)=-0.2$, $\sigma_i=0.2$, and $v_i(n)$ were zero mean and unit variance random processes with Gaussian distribution.

From Fig. 5 it is observed that a reduction in the bias of 10 dB is achieved when in the forward path a third-order all-pass filter is used instead of a delay of z^{-10} . The input was colored noise derived by passing white noise of zero mean and unit variance through a stable autoregressive model of the form

$$K(Z) = \frac{1}{1 - 1.79z^{-1} + 1.85z^{-2} - 1.27z^{-3} + 0.41z^{-4}}. \quad (23)$$

Time-varying all-pass filters hold another strong advantage over forward delays: they reduce the steady-state bias even for white noise input $u(n)$. This is illustrated in Fig. 6 where it is shown that a third-order all-pass filter achieves approximately 5 dB reduction of the steady-state error compared to a forward path delay.

In Fig. 7 the performance improvement achieved by the proposed use of varying all-pass filters over the standard method with delays is illustrated in terms of bias reduction for the case of speech input signal. This speech signal was recorded in a typical office room with a DAT machine at a sampling rate of 48 kHz and it was downsampled to 24 kHz. From this plot it is observed that the introduction of a third-order all-pass filter with a varying pole in the forward path

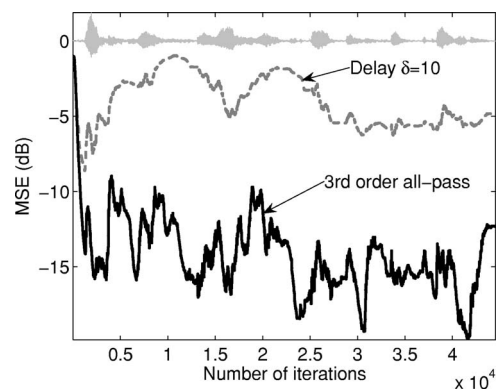


FIG. 7. Misalignment curves of a system with a delay and a system with an all-pass filter in the forward path, for speech input signal. The input signal is scaled and given within an offset in order to assess the performance of the feedback cancelling adaptive filter.

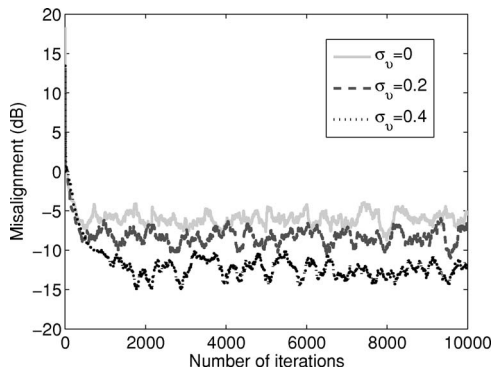


FIG. 8. Dependence of the steady-state bias on the variance of the poles of the all-pass filter.

can reduce the steady-state bias up to 10 dB. Notice that significant reduction in the bias was achieved mainly during silent intervals of the speech, indicating that the filter coefficient adaptation can be suspended or relaxed during the voiced intervals.

Notice also that the introduction of time-varying all-pass filters does not exclude the existence of delays in the forward path; on the contrary these can be combined to achieve even lower steady-state misalignment and thus more accurate acoustic feedback path estimates.

C. Effect of the parameters of the all-pass filter on the bias

The amount of bias in a hearing aid supplied with an adaptive feedback cancelling filter and an all-pass filter with varying poles is mainly a function of the variance of the poles of the all-pass filter. This can be verified from Fig. 8 where it is shown that the higher the variance of the poles the lower the bias value. Stochastic poles with uniform instead of Gaussian distribution were also tested and the results were found to be similar to those of Fig. 8.

From Fig. 5 and Fig. 6 it is concluded that the steady-state error is inversely proportional to the order of the decorrelating all-pass filter. The mean value of the poles did not have significant impact, since the steady-state misalignment was not sensitive to $\alpha_{o,i}$.

Finally in Fig. 9 it is illustrated that even a single varying pole can achieve a bias reduction that is comparable to

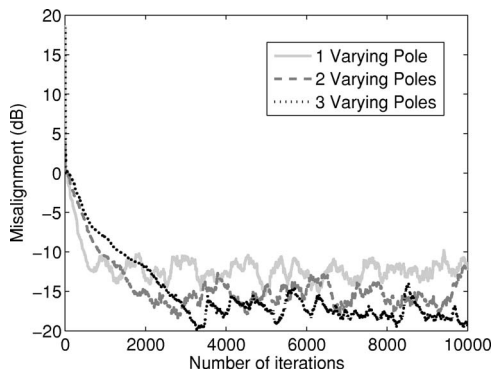


FIG. 9. Convergence behavior of the adaptive filter as a function of the number of varying poles for a third order decorrelating all-pass filter.

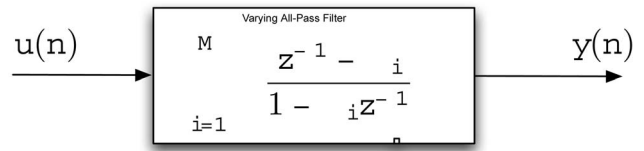


FIG. 10. The setting used for the direct effect of an all-pass filter on an audio signal $u(n)$.

the case were all the poles are varying. This can be explained from Viéta's formulas,²⁶ which imply that a single time-varying pole forces all the coefficients of the all-pass filter to vary, and thus it decorrelates (up to a satisfying degree) the input $u(n)$ from the output $y(n)$ of the hearing aid. For the derivation of the curves of Figs. 8 and 9, randomly updated poles according to (18) were employed with mean value $\alpha_o = -0.2$.

VI. SIGNAL QUALITY

A major task of audio processing systems is the preservation and/or enhancement of the quality of the processed signals. In other words, the elimination of the undesired echoes and ringing effects, which is accomplished by suppressing the acoustic feedback, should not be accompanied by noticeable distortion or degradation of the amplified audio signals.

To gain insight into the effect of the proposed all-pass-filters-based processing on the quality of the audio signals, their performance was further evaluated via subjective tests with 17 participants. The quality of the audio signals was evaluated using a slightly modified version of the ITU-R.BS.1116-1 impairment scale (Fig. 11), to measure the extent of the distortion.²⁷ The audio signals were reproduced with the use of headphones in order to simulate the hearing aid conditions more accurately.

5.0	Imperceptible
4.0	Perceptible but not annoying
3.0	Slightly annoying
2.0	Acceptable but very annoying
1.0	Unacceptable

FIG. 11. The modified version of the ITU-R.BS.1116-1 impairment scale that was used in our subjective tests.

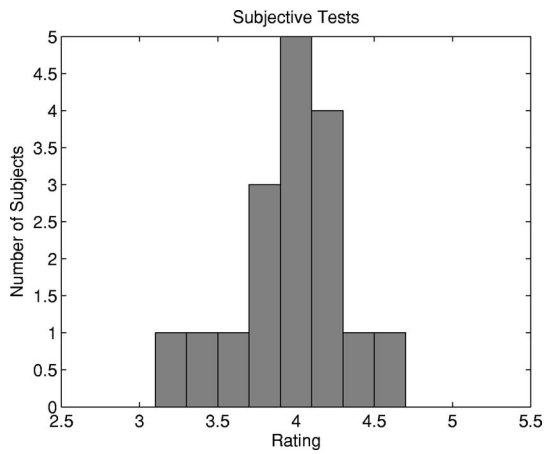


FIG. 12. Histogram of the results of subjective tests for the evaluation of the direct impact of a varying all-pass filter on an audio signal.

A. Direct impact

To assess the direct impact of a varying all-pass filter on an audio signal, a recorded speech signal sampled at 22 KHz, was filtered with a third-order all-pass filter (Fig. 10) whose coefficients were varying randomly according to (19), where $\lambda=0.2$ and $\sigma_i=0.5$ for every pole ($i=1,2,3$). Each subject listened to the input $u(n)$ and the output $y(n)$ separately and was asked to assign a number from 1 to 5 to the latter, indicating whether the artifacts added by the processing were noticeable or not. The results are illustrated in the histogram in Fig. 12. This clearly shows that our proposed varying filters add some noticeable artifacts to the output signal, but not perceptually annoying. Listeners described the perceived distortion as a barely noticeable hiss. From the spectrograms of the input and the output signal (Fig. 13) observe the close match between the two; this is due to the introduced distortion having significantly smaller power than the original signal as it appears only during speech intervals, rendering it imperceptible during silence periods.

B. Impact on a hearing aid

For the evaluation of the effect of a varying all-pass filter on the quality of the output signal of a hearing aid,

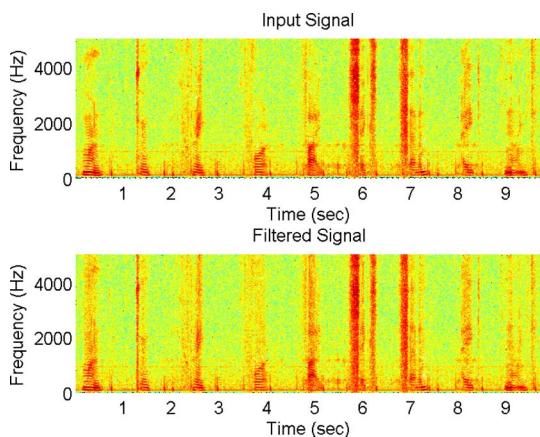


FIG. 13. (Color online) Spectrogram of the input and the output of the system of Fig. 10.

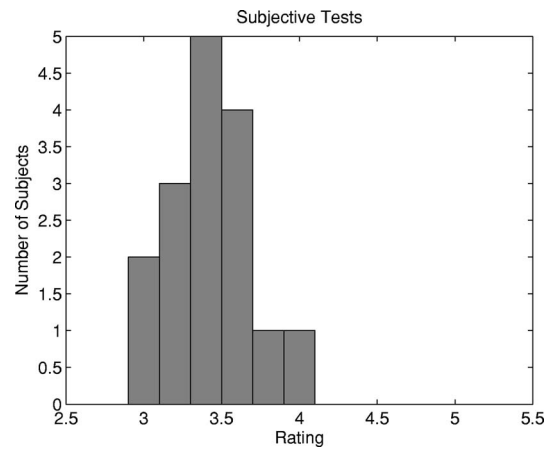


FIG. 14. Subjective tests for the evaluation of the output of the hearing aid whose block diagram is depicted in Fig. 1.

listeners were asked to compare the output of the system of Fig. 1 and the output of the system depicted in Fig. 3 to the signal $\mathcal{F}^{-1}\{A(z)\} \otimes u(n)$ which is the desired output. The input signal $u(n)$ was a recorded speech signal sampled at 22 KHz and the forward processing unit was a simple amplifier described by $A(z)=A_0z^{-1}$, $A_0=6$. The varying all-pass filter was of third-order and its poles were adapted according to the scheme presented in Appendix A.

From the results of the subjective tests, illustrated in Figs. 14 and 15, it is concluded that the participating subjects felt that the introduction of a varying all-pass filter improved the quality of the output of the hearing aid. Listeners claimed that the output of the hearing aid that included an all-pass filter was more clear than the output of the system of Fig. 1 since the metallic timbre and the echoes that are inherent in the output of the latter did not appear in the output of the system of Fig. 3. This is also supported by the spectrograms from Fig. 16, which show that the output of the hearing aid is closer to the desired output $A_0u(n)$ when a varying all-pass filter is introduced in the forward path of the system.

Therefore although all-pass filters introduce a noticeable, but not annoying hiss to the audio signals, especially during voiced intervals, they improve the overall performance of a hearing aid. More specifically they allow for

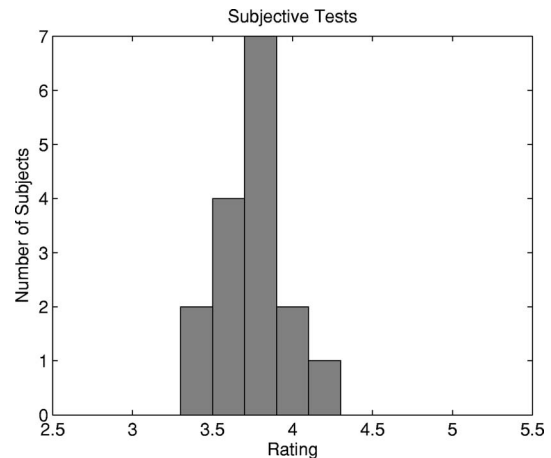


FIG. 15. Subjective tests for the evaluation of the output of the hearing aid whose block diagram is depicted in Fig. 3.

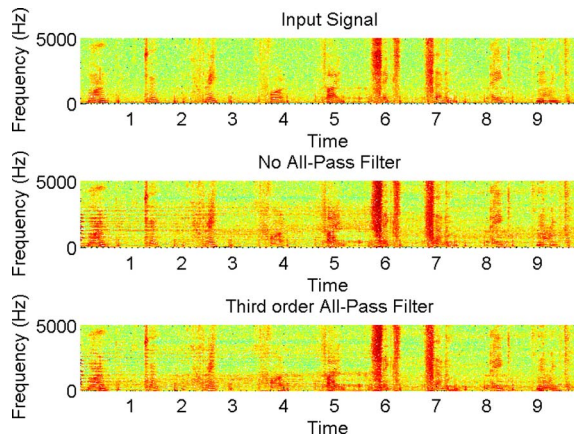


FIG. 16. (Color online). Spectrograms of the output of a hearing aid with and without an all-pass filter in its forward path.

more precise identification of the feedback path and thus for a remarkable suppression of the echoes and the metallic timbre of the output of a hearing aid.

VII. CONCLUSIONS

A novel technique for the reduction of bias in acoustic feedback cancellation systems has been proposed. This approach, which is based on the use of an all-pass filter in the forward path of a hearing aid (or any other system with acoustic feedback), has been shown to out-perform the existing ones in terms of steady-state error. Moreover, its behavior does not depend on the characteristics of the input signal. Several policies for the variation of the poles have been examined, and a method for the adaptation of the pole values has been proposed that aims at the minimization of the input-output cross correlation. The effect of the choice of the parameters of this all-pass filter on the steady-state error was studied. Experimental results from a simulated hearing aid support the analysis. Subjective tests were also conducted to assess the quality of the processed audio signals.

APPENDIX A: ADAPTING THE POLE OF THE ALL-PASS FILTER

Since the purpose of the introduced all-pass filter is to decorrelate the output of the hearing aid $y(n)$ from its input $u(n)$ an adaptive pole can be employed which aims at the minimization of the input-output cross correlation given by

$$\xi(n) = \frac{1}{2} r_{uy}^2(n) = \frac{1}{2} E\{u(n)y(n)\}^2, \quad (\text{A1})$$

and is updated at every time instant n according to

$$\alpha_i(n+1) = \alpha_i(n) + \eta_i \frac{\partial \xi(n)}{\partial \alpha_i(n)}. \quad (\text{A2})$$

Approximating the expectation of the product $u(n)y(n)$ with its instantaneous value and differentiating both sides of (A1) yields

$$\frac{\partial \xi(n)}{\partial \alpha_k(n)} = \left(\sum_{i=1}^M \frac{\partial r_i(n)}{\partial \alpha_k(n)} y(n-i) - A_o \sum_{i=1}^M \frac{\partial r_i(n)}{\partial \alpha_k(n)} e(n-M+i) \right) u^2(n)y(n). \quad (\text{A3})$$

Since $u(n)$ is not available its estimate $e(n)$ can be used instead resulting in

$$\frac{\partial \xi(n)}{\partial \alpha_k(n)} = \left(- \sum_{i=1}^M \frac{\partial r_i(n)}{\partial \alpha_k(n)} y(n-i) + A_o e(n-M) + A_o \sum_{i=1}^M \frac{\partial r_i(n)}{\partial \alpha_k(n)} e(n+M-i) \right) e^2(n)y(n) \quad (\text{A4})$$

when a varying all-pass filter of order M is introduced in the forward path with a transfer function given by (9). The coefficients $r_i(n)$ and the poles $\alpha_k(n)$ ($i=1, 2, \dots, M$) are related according to Viéta's formulas.²⁶ For $M=1$ (A4) becomes

$$\frac{\partial \xi(n)}{\partial \alpha(n)} \approx [y(n-1) - A_o e(n-1)] e^2(n)y(n) \quad (\text{A5})$$

and the pole adapts according to

$$\alpha(n+1) = \alpha(n) + \eta(y(n-1) - A_o e(n-1)) e^2(n)y(n) + \sigma v(n), \quad (\text{A6})$$

where η is the learning rate, $v(n)$ is a random variable of zero mean and unit variance, and σ is the standard deviation. This stochastic term is added to guarantee that the pole will vary even when steady state is reached. The derivation of recursive equations for the adaptation of the poles of higher-order all-pass filters is a straightforward procedure.

¹W. Knecht, "Some notes on feedback suppression with adaptive filters in hearing aids," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (1997).

²T. L. J. Hellgren and S. Arlinger, "System identification of feedback in hearing aids," *J. Acoust. Soc. Am.* **105**, 3481–3496 (1999).

³Defined as the maximum amplification value for which the hearing aid is stable.

⁴D. Bustamante, T. Worrall, and M. Williamson, "Measurement and adaptive suppression of acoustic feedback in hearing aids," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3 (1989), pp. 2017–2020.

⁵J. Maxwell and P. Zurek, "Reducing feedback in hearing aids," *IEEE Trans. Speech Audio Process.* **3**, 304–313 (1995).

⁶B. Rafaely, N. A. Shusina, and J. L. Hayes, "Robust compensation with adaptive feedback cancellation in hearing aids," *Speech Commun.* **39**, 163–170 (2003).

⁷B. Rafaely and M. Roccasalva-Firenze, "Control of feedback in hearing aids—A robust filter design approach," *IEEE Trans. Speech Audio Process.* **8**, 754–756 (2000).

⁸J. Hellgren and F. Urban, "Bias of feedback cancellation in hearing aids based on direct closed loop identification," *IEEE Trans. Speech Audio Process.* **9**, 906–913 (2001).

⁹J. Hellgren and U. Forssell, "Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification," *IEEE Trans. Speech Audio Process.* **9**, 906–913 (2001).

¹⁰J. Hellgren, "Analysis of feedback cancellation in hearing aids with filtered-X LMS and the direct method of closed loop identification," *IEEE Trans. Speech Audio Process.* **10**, 119–131 (2002).

¹¹L. Ljung, *System Identification, Theory for the User* (Prentice-Hall, Englewood Cliffs, NJ, 1987).

¹²J. Kates, "Feedback cancellation in hearing aids: Results from a computer simulation," *IEEE Trans. Signal Process.* **39**, 553–562 (1991).

- ¹³M. Siqueira and A. Alwan, "Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing aids," *IEEE Trans. Speech Audio Process.* **8**, 443–453 (2000).
- ¹⁴M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (1998) Vol. **6**, 3689–3692.
- ¹⁵C. Boukis, D. Mandic, and A. Constantinides, "Bias reduction in acoustic feedback cancellation systems with varying all-pass filters," *IEE Elec. Lett.* **42**, 556–558 (2006).
- ¹⁶B. Rafaely and J. Hayes, "On the modeling of the vent path in hearing aid systems," *J. Acoust. Soc. Am.* **109**, 1747–1749 (2001).
- ¹⁷It is assumed that these parameters have their values within the linear region of operation.
- ¹⁸B. Widrow and S. Stearns, *Adaptive Signal Processing* (Prentice Hall, New Jersey, 1985).
- ¹⁹C. Boukis, "Adaptive digital signal processing structures and identification algorithms for feedback control," Ph.D. thesis, Imperial College of Science, Technology and Medicine, University of London (2004).
- ²⁰Provided that they are of sufficient order.
- ²¹B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head Microphone," Technical Report 280, MIT Media Lab Perceptual Computing (1994).
- ²²P. Estermann and A. Kaelin, "Feedback cancellation in hearing aids: Results from using frequency domain adaptive filters," in *Proceedings of the IEEE International Symposium on Circuits and Systems* **2**, 257–260 (1994).
- ²³J. Hellgren and U. Forssell, "Bias of feedback cancellation algorithm as based on direct closed loop identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (2000) 869–872.
- ²⁴J. Greenberg, P. Zurek, and M. Brantley, "Evaluation of feedback-reduction algorithms for hearing aids," *J. Acoust. Soc. Am.* **108**, 2366–2376 (2000).
- ²⁵S. Lipshitz, M. Pocock, and J. Vanderkooy, "On the audibility of midrange phase distortions in audio systems," *J. Audio Eng. Soc.* **30**, 580–595 (1982).
- ²⁶E. W. Weisstein, "Viéta's formulas," <http://mathworld.wolfram.com/VietasFormulas.html> (2006).
- ²⁷G. S. G. Norcross and M. Lavoie, "Subjective investigations of inverse filtering," *J. Acoust. Soc. Am.* **52**, 1003–1027 (2004).

A full vectorial contrast source inversion scheme for three-dimensional acoustic imaging of both compressibility and density profiles

Koen W. A. van Dongen^{a)} and William M. D. Wright^{b)}

Ultrasonics Research Group, Department of Electrical and Electronic Engineering, University College Cork, College Road, Cork, Ireland

(Received 3 August 2006; revised 6 December 2006; accepted 6 December 2006)

Imaging the two acoustic medium parameters density and compressibility requires the use of both the acoustic pressure and velocity wave fields, described via integral equations. Imaging is based on solving for the unknown medium parameters using known measured scattered wave fields, and it is difficult to solve this ill-posed inverse problem directly using a conjugate gradient inversion scheme. Here, a contrast source inversion method is used in which the contrast sources, defined via the product of changes in compressibility and density with the pressure and velocity wave fields, respectively, are computed iteratively. After each update of the contrast sources, an update of the medium parameters is obtained. Total variation as multiplicative regularization is used to minimize blurring in the reconstructed contrasts. The method successfully reconstructed three-dimensional contrast profiles based on changes in both density and compressibility, using synthetic data both with and without 50% white noise. The results were compared with imaging based only on the pressure wave field, where speed of sound profiles were solely based on changes in compressibility. It was found that the results improved significantly by using the full vectorial method when changes in speed of sound depended on changes in both compressibility and density. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431333]

PACS number(s): 43.60.Pt, 43.20.El, 43.20.Fn, 43.60.Uv, 43.80.Qf [TDM] Pages: 1538–1549

I. INTRODUCTION

Ultrasonic tomographic imaging^{1–4} is a well-established technique for medical diagnosis and other applications. One of the medical applications is to use ultrasound for imaging temperature distributions in the human body (e.g., Johnson *et al.*⁵) and in particular during high-temperature hyperthermia cancer treatments with high-intensity focused ultrasound (HIFU).^{6,7} During this treatment, the tumor temperature is increased up to 50–55 °C for a duration of 1 or 2 min or for a shorter period of time at temperatures over 60 °C.^{8,9} For efficient treatment, it is important to monitor the process by verifying (i) that the tumor temperature has become sufficiently high to induce cell necrosis, and (ii) that the healthy tissue surrounding the tumor remains at or near normal body temperature to prevent it from being damaged. The advantages of using ultrasound as a guidance method are that it is noninvasive, relatively cheap, and compatible with the HIFU apparatus.

Several techniques exist to estimate the tissue temperature from ultrasound. Some are based on thermally induced strain^{10–12} while others are based on the observation that biological tissues can be described by semiregular lattices of which the scattering properties change as a function of temperature.^{13–17}

Using acoustics for monitoring the process is a difficult task due to the fact that the changes in acoustic medium parameters between various soft tissues or between regions of the same tissue at different temperatures are relatively small. In liver tissue for example, a change in temperature from 37 °C to 50 °C typically produces a contrast in sound speed of only 6 m/s.^{10,11,18} Consequently, a more sensitive imaging method utilizing through-transmission may be more appropriate,¹⁹ which has also been used for temperature estimation.^{20,21} These techniques typically operate with transient signals of which the frequency range is a compromise between using high enough frequencies to obtain adequate spatial resolution, and a low enough frequency content for adequate tissue penetration. The drawback of using low frequencies is that the corresponding wavelengths of the probing wave field are of the same order of magnitude as the spatial dimensions of the contrasts of interest, and effects like diffraction and multiple scattering between various regions with different medium parameters will occur. These effects are well described by utilizing the acoustic wave equations and using these wave equations for imaging is therefore an appropriate approach.²² So far, the above methods do not employ the knowledge that these changes in speed of sound are not caused solely by changes in compressibility but by changes in both compressibility and volume density of mass. Hence, by doing so it is expected that the imaging results will improve, and this will be demonstrated in the work described here. While the motivation was to develop an imaging method for HIFU temperature moni-

^{a)}Current address: Laboratory of Acoustical Imaging and Sound Control, Faculty of Applied Sciences, Delft University of Technology, The Netherlands. Electronic mail: k.w.a.vandongen@tudelft.nl

^{b)}Electronic mail: bill.wright@ucc.ie

toring, the techniques described could easily be applied to any acoustic imaging problem containing small, weakly scattering contrasts.

The propagation and scattering of acoustic wave fields in the presence of acoustic contrasts is described via integral equations. These are obtained by combining the acoustic wave equations with reciprocity.²³ We refer to the forward problem where we solve the integral equation for the unknown total wave field in the case where we know the incident wave field and the contrast function. The inverse problem refers to the situation where the integral equation is solved for an unknown contrast function and known incident and scattered wave fields. Inspired by the success of conjugate gradient iterative solution methods for the forward problem,^{24,25} a modified gradient inversion method has already been developed. With this method, the unknown contrast function was reconstructed from measured scattered wave fields with a conjugate gradient inversion scheme.²⁶ In order to tackle the blurring caused by noise in the data, the technique was extended by taking the total variation (TV) of the reconstructed contrast function into account.^{27–30} The method was further improved by combining the contrast function with the total wave field leading to the introduction of source type integral equations (STIE),³¹ and later on in combination with the modified gradient approach^{26,32,33} to the contrast source (CS) method.^{34,35} This has proven to be very successful, especially in combination with TV as a multiplicative regularizer for the electromagnetic,³⁶ elastodynamic,³⁷ and finally the acoustic scalar cases³⁸ where changes in density were neglected. Problems arise in situations where changes in speed of sound are not only due to changes in compressibility but also due to changes in density. In these situations, the acoustic scalar method is not sufficient and must be extended to a full vectorial method, which takes both the velocity and pressure wave fields into account. A full vectorial solution was presented in Refs. 39 and 40 for electromagnetics; here, we have developed a full vectorial scheme for the acoustic counterpart which meets the demands of reconstructing both density and compressibility profiles⁴¹ and included an additional multiplicative regularization method based on minimization of the TV in the reconstructed contrast profiles. Contrary to the electromagnetic counterpart,^{39,40} we normalize the error functionals using (i) a norm based on a combination of the pressure and velocity wave fields, and (ii) a norm which combines the presence of contrasts in both compressibility and density. Finally, we investigate the effect of the standard scalar approach on the reconstructed speed of sound profiles where changes in density are neglected.

II. FORWARD PROBLEM

The forward and inverse problems are formulated in the temporal Laplace domain with Laplace parameter \hat{s} . Frequency domain results are obtained by taking the limit $\hat{s} \rightarrow -i\omega$, with $i^2 = -1$ and where $\omega \in \Omega$, with Ω the temporal angular frequency domain. The symbol “ \wedge ” on top of a pa-

rameter is used to show its frequency dependency, whereas the vectors \mathbf{r} and \mathbf{r}' are used to notate a position in the three-dimensional spatial domain \mathbb{D} .

The propagation of acoustic wave fields is described via integral equations of the second type.²³ In the presence of acoustic contrasts, the total pressure and velocity wave fields $\hat{p}^{\text{tot}}(\mathbf{r})$ and $\hat{v}_i^{\text{tot}}(\mathbf{r})$ for $\{i, j\} = 1, 2, \text{ or } 3$ equals

$$\hat{p}^{\text{tot}}(\mathbf{r}) = \hat{p}^{\text{inc}}(\mathbf{r}) + \hat{p}^{\text{sct}}(\mathbf{r}), \quad (1)$$

$$\hat{v}_i^{\text{tot}}(\mathbf{r}) = \hat{v}_i^{\text{inc}}(\mathbf{r}) + \hat{v}_i^{\text{sct}}(\mathbf{r}), \quad (2)$$

where $\hat{p}^{\text{inc}}(\mathbf{r})$ and $\hat{v}_i^{\text{inc}}(\mathbf{r})$ are the incident pressure and velocity wave fields and where $\hat{p}^{\text{sct}}(\mathbf{r})$ and $\hat{v}_i^{\text{sct}}(\mathbf{r})$ refer to the scattered pressure and velocity fields, respectively.

The scattered wave fields occur due to the presence of acoustic contrasts defined via the contrast functions $\chi^{\kappa}(\mathbf{r}')$ and $\chi^{\rho}(\mathbf{r}')$, which read

$$\chi^{\kappa}(\mathbf{r}') = \kappa^{\text{bg}} - \kappa^{\text{obj}}(\mathbf{r}'), \quad (3)$$

$$\chi^{\rho}(\mathbf{r}') = \rho^{\text{bg}} - \rho^{\text{obj}}(\mathbf{r}'). \quad (4)$$

The superscripts “bg” and “obj” are used to refer to the background and object medium parameters, respectively, which we assume to be nondispersive. In addition, we define the following Green’s functions \hat{G} applied on a volume density of injection rate source $\hat{q}(\mathbf{r}')$ or a volume density of force $\hat{f}_j(\mathbf{r}')$ as follows:

$$\{\hat{G}^{pq}\hat{q}\}(\mathbf{r}) = \hat{s}\rho^{\text{bg}} \int_{\mathbf{r}' \in \mathbb{D}} \hat{G}(\mathbf{r} - \mathbf{r}') \hat{q}(\mathbf{r}') dV(\mathbf{r}'), \quad (5)$$

$$\{\hat{G}^{pf}\hat{f}_j\}(\mathbf{r}) = - \sum_j \partial_j \int_{\mathbf{r}' \in \mathbb{D}} \hat{G}(\mathbf{r} - \mathbf{r}') \hat{f}_j(\mathbf{r}') dV(\mathbf{r}'), \quad (6)$$

$$\{\hat{G}^{vq}\hat{q}\}_i(\mathbf{r}) = - \partial_i \int_{\mathbf{r}' \in \mathbb{D}} \hat{G}(\mathbf{r} - \mathbf{r}') \hat{q}(\mathbf{r}') dV(\mathbf{r}'), \quad (7)$$

$$\{\hat{G}^{vf}\hat{f}_j\}_i(\mathbf{r}) = \frac{1}{\hat{s}\rho^{\text{bg}}} \left[\partial_i \sum_j \partial_j \int_{\mathbf{r}' \in \mathbb{D}} \hat{G}(\mathbf{r} - \mathbf{r}') \hat{f}_j(\mathbf{r}') dV(\mathbf{r}') + \sum_j \int_{\mathbf{r}' \in \mathbb{D}} \delta(\mathbf{r} - \mathbf{r}') \delta_{i,j} \hat{f}_j(\mathbf{r}') dV(\mathbf{r}') \right], \quad (8)$$

with ∂_i the spatial derivative in the x_i direction, $\delta_{i,j}$ Kronecker’s delta function, $\delta(\mathbf{r} - \mathbf{r}')$ the impulse response function, $\hat{G}(\mathbf{r} - \mathbf{r}')$ the scalar form of Green’s function.²³

It is well known that the scattered wave fields $\hat{p}^{\text{sct}}(\mathbf{r})$ and $\hat{v}_i^{\text{sct}}(\mathbf{r})$ read

$$\hat{p}^{\text{sct}}(\mathbf{r}) = \{\hat{G}^{pq}\hat{s}\chi^{\kappa}\hat{p}^{\text{tot}}\}(\mathbf{r}) + \{\hat{G}^{pf}\hat{s}\chi^{\rho}\hat{v}_j^{\text{tot}}\}(\mathbf{r}), \quad (9)$$

$$\hat{v}_i^{\text{sct}}(\mathbf{r}) = \{\hat{G}^{vq}\hat{s}\chi^{\kappa}\hat{p}^{\text{tot}}\}_i(\mathbf{r}) + \{\hat{G}^{vf}\hat{s}\chi^{\rho}\hat{v}_j^{\text{tot}}\}_i(\mathbf{r}), \quad (10)$$

which can be proved by using reciprocity.²³ Equations (9) and (10) form an integral equation which can be solved numerically for known contrast functions and unknown total wave fields.²²

The above vectorial formulation can be reduced to a scalar formulation by assuming that there is only a contrast in compressibility. This leads to the following expression for the total pressure wave field:

$$\hat{p}^{\text{tot}}(\mathbf{r}) = \hat{p}^{\text{inc}}(\mathbf{r}) + \delta^2 \int_{\mathbf{r}' \in \text{D}} \hat{G}(\mathbf{r} - \mathbf{r}') \times \left(\frac{1}{[c^{\text{bg}}]^2} - \frac{1}{[c^{\text{obj}}(\mathbf{r}')]^2} \right) \hat{p}^{\text{tot}}(\mathbf{r}') dV(\mathbf{r}'), \quad (11)$$

with $[c^{\text{bg}}]^2 = 1/[\kappa^{\text{bg}}\rho^{\text{bg}}]$ and $[c^{\text{obj}}(\mathbf{r}')]^2 = 1/[\kappa^{\text{obj}}(\mathbf{r}')\rho^{\text{bg}}]$. This very same result is obtained if we use the scalar wave equation as the starting point. Note that the situation where there is only a contrast in density will lead to a scattered velocity wave field only, which is independent of the pressure wave field.

III. INVERSE PROBLEM AND CONTRAST SOURCE FORMULATION

The inverse problem refers to the situation where the unknown contrast function is computed. Here, we make use of the knowledge we have about the incident wave field everywhere in the region of interest and the measured scattered wave field at a limited number of receiver positions. For the acoustic case this results in the reconstruction of the acoustic medium parameters compressibility κ and density ρ by solving the set of integral equations stated in Eqs. (1)–(8).

Various methods exist to solve this problem within the Born or Rytov approximation.^{42–45} However, this approximation only works as long as multiple scattering can be neglected and the scatterers themselves are weak. Improvements can be expected by employing the distorted Born iteration method;⁴⁶ however, this still involves utilizing an approximation to the formal solution of the inverse problem. Hence, research has been conducted towards methods which go beyond the Born approximation and which do not require solution of the forward problem after each update of the contrast function.

Due to the ill-posed nature of the inverse problem, it cannot be solved with a standard conjugate gradient (CG) inversion scheme, and regularization is required. This regularization is obtained by using a contrast source (CS) formulation as applied to, amongst others, electromagnetic scattering problems.^{34,35} This formulation leads to the introduction of two, instead of one, error functionals which are minimized iteratively with an adapted CG inversion scheme. Although the problem remains ill-posed,⁴⁷ the approach has proven to be very effective in many inversion procedures.³¹ Additional regularization is obtained by minimizing the TV present in the reconstructed contrast functions.^{36–39} The full electromagnetic scattering problem, where contrasts in both complex permittivity and permeability are present, has been investigated previously;³⁹ here, we investigate the full vectorial acoustic counterpart. We first formulate the inverse problem based on the full vectorial acoustic problem. We then investigate the simplified scalar wave field problem by neglecting changes in mass density which are related to the velocity wave field.

Examining Eqs. (9) and (10) reveals that there are two source types that generate the scattered wave fields: sources defined via the product of changes in compressibility with the total pressure wave field, and sources defined via the product of changes in density with the velocity wave fields. This leads to the introduction of the contrast sources $\hat{w}_k^p(\mathbf{r})$ and $\hat{w}_{i,k}^v(\mathbf{r})$, which are defined as

$$\hat{w}_k^p(\mathbf{r}) = \hat{\chi}^\kappa(\mathbf{r}) \hat{p}_k^{\text{tot}}(\mathbf{r}), \quad (12)$$

$$\hat{w}_{i,k}^v(\mathbf{r}) = \hat{\chi}^\rho(\mathbf{r}) \hat{v}_{i,k}^{\text{tot}}(\mathbf{r}), \quad (13)$$

where the subscript $k=1, \dots, K$ refers to the discrete spatial positions of the transmitter which generates the incident wave fields. Consequently, the scattered wave fields formulated in Eqs. (9) and (10) are a function of the contrast sources $\hat{w}_k^p(\mathbf{r})$ and $\hat{w}_{i,k}^v(\mathbf{r})$, and equal

$$\hat{p}_k^{\text{sct}}(\mathbf{r}) = \{\hat{\mathcal{G}}^{pq} \hat{w}_k^p(\mathbf{r})\} + \{\hat{\mathcal{G}}^{pf} \hat{w}_{j,k}^v(\mathbf{r})\}, \quad (14)$$

$$\hat{v}_{i,k}^{\text{sct}}(\mathbf{r}) = \{\hat{\mathcal{G}}^{vq} \hat{w}_{j,i}^p(\mathbf{r})\} + \{\hat{\mathcal{G}}^{vf} \hat{w}_{j,k}^v(\mathbf{r})\}. \quad (15)$$

Combining Eqs. (12)–(15) leads to an expression of the contrast sources which reads

$$\hat{w}_k^p(\mathbf{r}) = \hat{\chi}^\kappa(\mathbf{r}) \hat{p}_k^{\text{inc}}(\mathbf{r}) + \hat{\chi}^\kappa(\mathbf{r}) \{\hat{\mathcal{G}}^{pq} \hat{w}_k^p(\mathbf{r})\} + \hat{\chi}^\kappa(\mathbf{r}) \times \{\hat{\mathcal{G}}^{pf} \hat{w}_{j,k}^v(\mathbf{r})\}, \quad (16)$$

$$\hat{w}_{i,k}^v(\mathbf{r}) = \hat{\chi}^\rho(\mathbf{r}) \hat{v}_{i,k}^{\text{inc}}(\mathbf{r}) + \hat{\chi}^\rho(\mathbf{r}) \{\hat{\mathcal{G}}^{vq} \hat{w}_{j,i}^p(\mathbf{r})\} + \hat{\chi}^\rho(\mathbf{r}) \times \{\hat{\mathcal{G}}^{vf} \hat{w}_{j,k}^v(\mathbf{r})\}. \quad (17)$$

Next, Eqs. (12)–(17) are used to formulate two sets of error functionals. The first one is referred to as the error in the data equation and is based on the residual of the measured scattered wave fields minus the computed scattered wave fields based on the approximated contrast sources. Hence, at the N th iteration step these error functionals read

$$\hat{r}_{S;N;k}^p(\mathbf{r}) = \hat{p}_k^{\text{sct}}(\mathbf{r}) - \{\hat{\mathcal{G}}_S^{pq} \hat{w}_{N;k}^p(\mathbf{r})\} - \{\hat{\mathcal{G}}_S^{pf} \hat{w}_{N;j;k}^v(\mathbf{r})\} \quad \forall \mathbf{r} \in S, \quad (18)$$

$$\hat{r}_{S;N;i,k}^v(\mathbf{r}) = \hat{v}_{i,k}^{\text{sct}}(\mathbf{r}) - \{\hat{\mathcal{G}}_S^{vq} \hat{w}_{N;j,i}^p(\mathbf{r})\} - \{\hat{\mathcal{G}}_S^{vf} \hat{w}_{N;j;k}^v(\mathbf{r})\} \quad \forall \mathbf{r} \in S, \quad (19)$$

with S the spatial domain containing the transducers. The second one is referred to as the error in the object equation and is based on the error in the computed contrast sources. Consequently, these error functionals equal

$$\begin{aligned} \hat{r}_{D;N;k}^p(\mathbf{r}) &= \hat{s}\chi_N^k(\mathbf{r})\hat{\rho}_k^{\text{inc}}(\mathbf{r}) + \hat{s}\chi_N^k(\mathbf{r})\{\hat{\mathcal{G}}^{pq}w_k^p\}(\mathbf{r}) \\ &+ \hat{s}\chi_N^k(\mathbf{r})\{\hat{\mathcal{G}}^{pf}\hat{w}_{N;j;k}^v\}(\mathbf{r}) - \hat{w}_{N;k}^p(\mathbf{r}) \quad \forall \mathbf{r} \in \mathbb{D}, \end{aligned} \quad (20)$$

$$\begin{aligned} \hat{r}_{D;N;i,k}^v(\mathbf{r}) &= \hat{s}\chi_N^i(\mathbf{r})\hat{v}_{i,k}^{\text{inc}}(\mathbf{r}) + \hat{s}\chi_N^i(\mathbf{r})\{\hat{\mathcal{G}}^{vq}w_k^p\}(\mathbf{r}) \\ &+ \hat{s}\chi_N^i(\mathbf{r})\{\hat{\mathcal{G}}^{vf}\hat{w}_{N;j;k}^v\}(\mathbf{r}) - \hat{w}_{N;i,k}^v(\mathbf{r}) \quad \forall \mathbf{r} \in \mathbb{D}, \end{aligned} \quad (21)$$

with \mathbb{D} the spatial domain containing the contrasts and contrast sources.

A. UPDATE OF THE CONTRAST SOURCES

For given contrast sources $w_{N;k}^p(\mathbf{r})$ and $w_{N;j,k}^v(\mathbf{r})$ and contrasts $\chi_N^p(\mathbf{r})$ and $\chi_N^i(\mathbf{r})$, in each iteration, a single step in a CG scheme with Polak-Ribière update directions is employed to update the contrast sources^{39,40} where updating of the contrast sources and the contrasts occurs in an alternating fashion. The update direction for the contrast sources in the inversion scheme depends on the definition of the norm of a vector. Contrary to the method described in Refs. 39 and 40, which uses two separate norms for each wave field quantity based solely on each wave field quantity, we propose to use a combined L_2 norm of the four-dimensional vector representing the acoustic wave fields in the spatial domain \mathbb{S} . This improves the weighting and normalizing of the field quantities and consequently our method will perform better, especially in situations where the contrast in density is considerably smaller (or larger) than the contrast in the compressibility. Hence, the L_2 norm of the combined pressure and velocity wave fields equals

$$\left\| \begin{array}{c} \hat{p}_k(\mathbf{r}) \\ \hat{v}_{i,k}(\mathbf{r}) \end{array} \right\|_{\mathbb{S}}^2 = \sum_{\mathbf{r} \in \mathbb{S}, k, \Omega} \left[Z\hat{p}_k(\mathbf{r})\overline{\hat{p}_k(\mathbf{r})} + \sum_i \hat{v}_{i,k}(\mathbf{r})\overline{\hat{v}_{i,k}(\mathbf{r})} \right], \quad (22)$$

where the lines over the quantities $\hat{p}_k(\mathbf{r})$ and $\hat{v}_{i,k}(\mathbf{r})$ are used to denote that the complex conjugate is taken, whereas the L_2 norm of a four-dimensional vector representing the contrast sources in the spatial domain \mathbb{D} reads

$$\left\| \begin{array}{c} \hat{w}_{N;k}^p(\mathbf{r}) \\ \hat{w}_{N;i,k}^v(\mathbf{r}) \end{array} \right\|_{\mathbb{D}}^2 = \sum_{\mathbf{r} \in \mathbb{D}, k, \Omega} \left[\hat{w}_{N;k}^p(\mathbf{r})\overline{\hat{w}_{N;k}^p(\mathbf{r})} + \sum_i Z\hat{w}_{N;i,k}^v(\mathbf{r})\overline{\hat{w}_{N;i,k}^v(\mathbf{r})} \right]. \quad (23)$$

Note that we use the constant Z to correct for differences in the dimensions between the pressure and velocity wave field; consequently, Z is defined as

$$Z = \frac{\kappa^{\text{bg}}}{\rho^{\text{bg}}}. \quad (24)$$

Finally, we define the normalized error functional Err_N^{SD} at the N th iteration step as

$$\text{Err}_N^{\text{SD}} = \frac{\left\| \hat{r}_{S;N;k}^p(\mathbf{r}) \right\|_{\mathbb{S}}^2 + \left\| \hat{r}_{D;N;k}^p(\mathbf{r}) \right\|_{\mathbb{D}}^2}{\left\| \hat{p}_k^{\text{sct}}(\mathbf{r}) \right\|_{\mathbb{S}}^2} + \frac{\left\| \hat{r}_{D;N;i,k}^v(\mathbf{r}) \right\|_{\mathbb{D}}^2}{\left\| \hat{s}\chi_{N-1}^i(\mathbf{r})\hat{v}_{i,k}^{\text{inc}}(\mathbf{r}) \right\|_{\mathbb{D}}^2}, \quad (25)$$

with $\chi_{N-1}^k(\mathbf{r})$ and $\chi_{N-1}^i(\mathbf{r})$ based on the approximated medium parameters of the objects. These medium parameters are obtained from the contrast sources as discussed in the next section.

B. UPDATE OF DENSITY AND COMPRESSIBILITY PROFILES FROM CONTRAST SOURCES

Profiles of the contrasts in compressibility and density are obtained from the contrast sources in an alternating fashion; after each update of the contrast sources, the contrasts $\chi_N^k(\mathbf{r})$ and $\chi_N^i(\mathbf{r})$ are computed by minimizing the error in the contrast sources defined in Eqs. (12) and (13). Direct minimization⁴⁸ of the L_2 norm of the error functionals based on these definitions leads to the following expressions for the medium parameters of the objects which read

$$\kappa_N^{\text{obj}}(\mathbf{r}) = \kappa^{\text{bg}} - \frac{\sum_{k, \Omega} \Re[\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})\overline{\hat{w}_{N;k}^p(\mathbf{r})}]}{\sum_{k, \Omega} |\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})|^2}, \quad (26)$$

$$\rho_N^{\text{obj}}(\mathbf{r}) = \rho^{\text{bg}} - \frac{\sum_{i, k, \Omega} \Re[\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})\overline{\hat{w}_{N;i,k}^v(\mathbf{r})}]}{\sum_{i, k, \Omega} |\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})|^2}. \quad (27)$$

However, if we explicitly assume the medium parameters to be real and non-negative, and multiply both sides of Eqs. (12) and (13) with $[\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})]/|\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})|^M$ and $[\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})]/|\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})|^M$, respectively, the following expressions for $\kappa_N^{\text{obj}}(\mathbf{r})$ and $\rho_N^{\text{obj}}(\mathbf{r})$ are obtained:

$$\begin{aligned} \kappa_N^{\text{obj}}(\mathbf{r}) &= \left(\frac{\sum_{k, \omega} \left\{ \Re \left[(\hat{w}_{N;k}^p(\mathbf{r}) - \hat{s}\kappa^{\text{bg}}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})) \frac{\overline{\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})}}{|\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})|^M} \right] \right\}^2 \right)^{1/2}}{\sum_{k, \omega} \left\{ \Re \left[\frac{\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})\overline{\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})}}{|\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})|^M} \right] \right\}^2}, \end{aligned} \quad (28)$$

$$\begin{aligned} \rho_N^{\text{obj}}(\mathbf{r}) &= \left(\frac{\sum_{i, k, \omega} \left\{ \Re \left[(\hat{w}_{N;i,k}^v(\mathbf{r}) - \hat{s}\rho^{\text{bg}}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})) \frac{\overline{\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})}}{|\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})|^M} \right] \right\}^2 \right)^{1/2}}{\sum_{i, k, \omega} \left\{ \Re \left[\frac{\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})\overline{\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})}}{|\hat{s}\hat{v}_{N;i,k}^{\text{tot}}(\mathbf{r})|^M} \right] \right\}^2}, \end{aligned} \quad (29)$$

where M can be chosen to be 1 or 2.

C. REGULARIZATION WITH MULTIPLICATIVE TOTAL VARIATION

In situations where the data contain noise, additional regularization is preferable to stabilize the inversion process and to diminish the blurring in the reconstructed image. Results obtained with the CG-CS method are improved by applying regularization based on minimization of the TV in the reconstructed contrast profiles.^{36–39} This is based on the knowledge that the objects under investigation are embedded in a homogeneous background medium, and that the amount of transient fluctuations is small. Implementation of this regularization is done by multiplying the error functional Err_N^{SD} as defined in Eq. (25) with an additional term that depends solely on the TV in the reconstructed contrasts. Consequently, we obtain a new error function Err_N which equals

$$\text{Err}_N = \text{Err}_N^{\text{SD}} \text{Err}_N^{\text{TV}}, \quad (30)$$

where Err_N^{TV} is a weighted L_2 norm of the gradient of the contrast functions and equals

$$\text{Err}_N^{\text{TV}} = \int_{r \in \text{D}} \left[\frac{|\nabla \chi_N^{\kappa}(\mathbf{r})|^2 + \delta_N^{\kappa}}{b_N^{\kappa}} + \frac{|\nabla \chi_N^{\rho}(\mathbf{r})|^2 + \delta_N^{\rho}}{b_N^{\rho}} \right] dV(\mathbf{r}), \quad (31)$$

with

$$b_N^{\kappa}(\mathbf{r}) = V(|\nabla \chi_{N-1}^{\kappa}(\mathbf{r})|^2 + \delta_N^{\kappa}), \quad (32)$$

$$b_N^{\rho}(\mathbf{r}) = V(|\nabla \chi_{N-1}^{\rho}(\mathbf{r})|^2 + \delta_N^{\rho}), \quad (33)$$

where V equals the volume of the spatial domain D , $\nabla \chi_N^{\kappa}(\mathbf{r})$ and $\nabla \chi_N^{\rho}(\mathbf{r})$ equal the gradients of the reconstructed compressibility and density, respectively, and where δ_N^{κ} and δ_N^{ρ} are used to restore the differentiability of the TV factor. Good results are obtained with the following choices for δ_N^{κ} and δ_N^{ρ} :

$$\delta_N^{\kappa} = \frac{\theta \text{Err}_N^{\text{SD}}(\kappa^{\text{bg}})^2}{(\Delta V)^{2/3}}, \quad (34)$$

$$\delta_N^{\rho} = \frac{\theta \text{Err}_N^{\text{SD}}(\rho^{\text{bg}})^2}{(\Delta V)^{2/3}}, \quad (35)$$

with steering parameter θ controlling the reduction of the transient fluctuations in the reconstructed image; the lower the value, the larger the influence of the TV factor in the reconstruction process and the smoother the image becomes. This choice is based on the knowledge that $|\nabla \chi_{N-1}^{\text{obj}}(\mathbf{r})|^2$ and $|\nabla \rho_{N-1}^{\text{obj}}(\mathbf{r})|^2$ are of the order of $(\kappa^{\text{bg}})^2(\Delta V)^{-2/3}$ and $(\rho^{\text{bg}})^2(\Delta V)^{-2/3}$. The influence of the TV factor should be increased during the inversion process when the system becomes increasingly sensitive to noise present in the data; hence, the factor Err_N^{SD} is added to obtain this effect; the overall desired effect with the TV factor depends on the amount of transient fluctuations permitted in the image, which is based for example on the noise present in the data. Consequently, a separate steering parameter θ is needed which takes this into account. Typical values for θ are in the range $\theta = [10^{-3}, 10^0]$. Note that the choice of our δ_N^{κ} and δ_N^{ρ} is

TABLE I. The medium parameters compressibility, κ , volume density of mass, ρ , and speed of sound, c , for human liver tissues at $T=37, 45$, and 50 °C.

	Compressibility $\kappa[10^{-9}(\text{Pa})^{-1}]$	Density $\rho[\text{kg}/\text{m}^3]$	Speed of sound $c[\text{m}/\text{s}]$
Liver at 37 °C	0.366 48	1056.6	1607
Liver at 45 °C	0.365 36	1053.3	1612
Liver at 50 °C	0.365 71	1051.0	1613

different from the one chosen in Refs. 39 and 49. Here, contrast source inversion is investigated for the electromagnetic case where changes in electric permittivity and conductivity, and magnetic permeability are present. Based on their formulations, δ_N^{κ} and δ_N^{ρ} would read

$$\delta_N^{\kappa} = \frac{\left(\frac{\|\hat{p}_{D:N;k}^{\rho}(\mathbf{r})\|_{\text{D}}^2}{\|\hat{s}\chi_{N-1}^{\kappa}(\mathbf{r})\hat{p}_k^{\text{inc}}(\mathbf{r})\|_{\text{D}}^2} \right)}{[\max(\Delta x_1, \Delta x_2, \Delta x_3)]^2}, \quad (36)$$

$$\delta_N^{\rho} = \frac{\left(\frac{\|\hat{v}_{D:N;i;k}^{\rho}(\mathbf{r})\|_{\text{D}}^2}{\|\hat{s}\chi_{N-1}^{\rho}(\mathbf{r})\hat{v}_{i,k}^{\text{inc}}(\mathbf{r})\|_{\text{D}}^2} \right)}{[\max(\Delta x_1, \Delta x_2, \Delta x_3)]^2}, \quad (37)$$

which is disadvantageous in the situations where the error and/or changes in density are considerably larger or smaller than the error and/or changes in compressibility. In addition, the parameters $(\kappa^{\text{bg}})^2$ and $(\rho^{\text{bg}})^2$ are omitted in their formulations since they work with normalized contrasts.

Minimization of the error functional Err_N^{TV} leads to the following updating scheme for the reconstructed acoustic medium parameters:

$$\chi_N^{\kappa, \text{TV}}(\mathbf{r}) = \chi_N^{\kappa}(\mathbf{r}) + \beta_N^{\kappa} g_N^{\kappa}(\mathbf{r}), \quad (38)$$

$$\chi_N^{\rho, \text{TV}}(\mathbf{r}) = \chi_N^{\rho}(\mathbf{r}) + \beta_N^{\rho} g_N^{\rho}(\mathbf{r}), \quad (39)$$

with update directions $g_N^{\kappa}(\mathbf{r})$ and $g_N^{\rho}(\mathbf{r})$ equal to

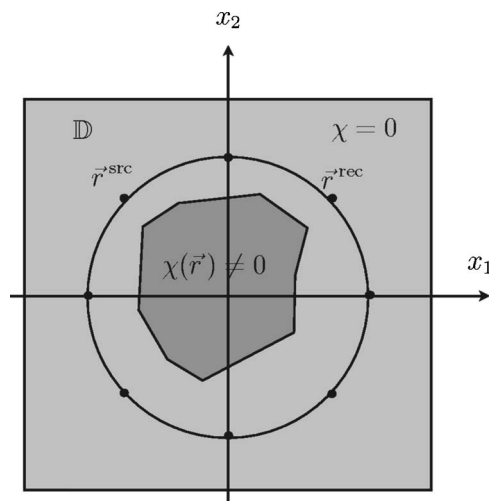


FIG. 1. The setup used for measurements. The transmitter and receiver positions are denoted by \vec{r}^{src} and \vec{r}^{rec} , whereas $\chi(\mathbf{r})$ are the contrast functions.

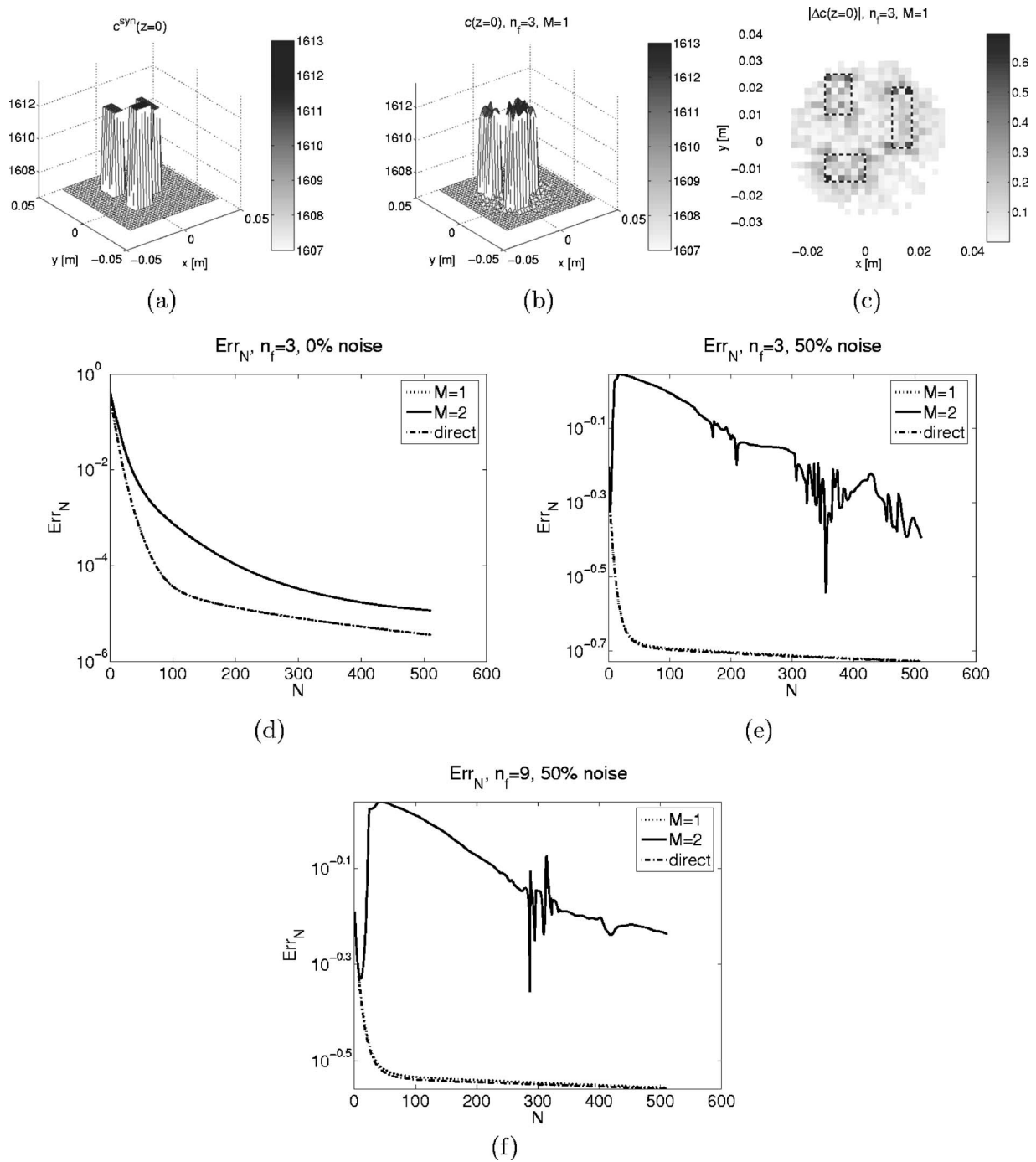


FIG. 2. Cross sections of the (a) synthetic and (b) reconstructed speed of sound profiles $c(\mathbf{r})$ at $z=0$ m and the absolute error $|\Delta c(\mathbf{r})|$ in the computed speed of sound profile; (d)–(f) the error functionals Err_N for $n_f=$ three and nine frequency components, without and with noise. The synthetic pressure wave field and the imaging are based on changes in the compressibility only.

$$g_N^\kappa(\mathbf{r}) = -2 \nabla \cdot \left[\frac{\nabla \chi_{N-1}^\kappa(\mathbf{r})}{b_N^\kappa} \right], \quad (40)$$

$$g_N^\rho(\mathbf{r}) = -2 \nabla \cdot \left[\frac{\nabla \chi_{N-1}^\rho(\mathbf{r})}{b_N^\rho} \right]. \quad (41)$$

The amplitudes of the update directions are chosen such that a maximum minimization of the error functional Err_N is obtained; hence,

$$\beta_N^\kappa = \arg \min_{\text{real } \beta^\kappa} \{ \text{Err}_M[\chi_N^\kappa(\mathbf{r}) + \beta_N^\kappa g_N^\kappa(\mathbf{r}), \chi_N^\rho(\mathbf{r})] \}, \quad (42)$$

$$\beta_N^\rho = \arg \min_{\text{real } \beta^\rho} \{ \text{Err}_M[\chi_N^\kappa(\mathbf{r}), \chi_N^\rho(\mathbf{r}) + \beta_N^\rho g_N^\rho(\mathbf{r})] \}. \quad (43)$$

This results in a fourth-degree polynomial in β_N^κ and a second one in β_N^ρ which can both be solved analytically.⁴⁹ Note that, after we have obtained the contrast functions $\chi_N^{\kappa;\text{TV}}(\mathbf{r})$ and $\chi_N^{\rho;\text{TV}}(\mathbf{r})$, we update the contrasts sources with $\chi_N^\kappa(\mathbf{r}) = \chi_N^{\kappa;\text{TV}}(\mathbf{r})$ and $\chi_N^\rho(\mathbf{r}) = \chi_N^{\rho;\text{TV}}(\mathbf{r})$. Note that the effect of multiplicative regularization on the inversion scheme was investigated earlier by Van den Berg *et al.*³⁵

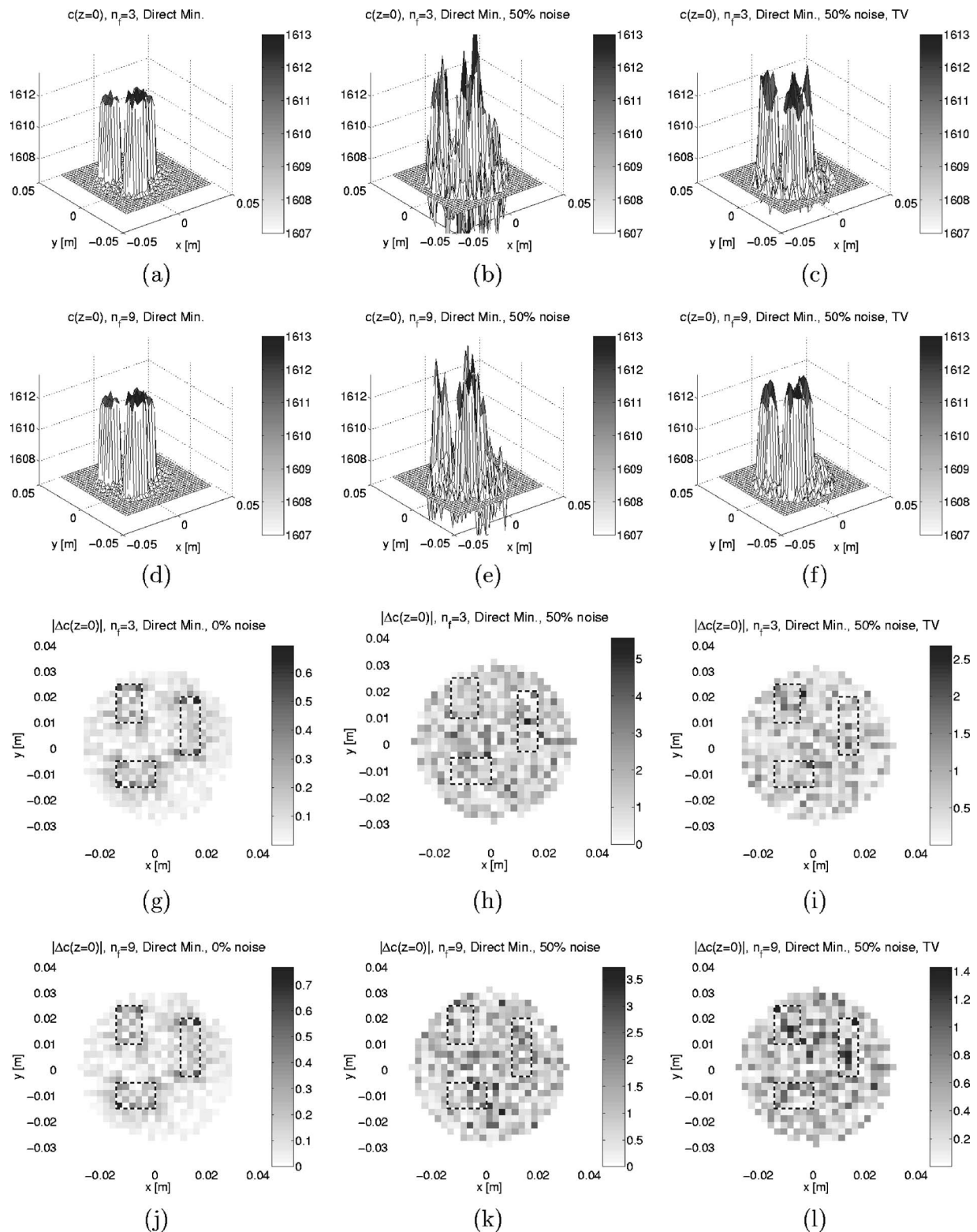


FIG. 3. Cross sections of the reconstructed speed of sound profiles $c(r)$ at $z=0$ m based on direct minimization and the absolute errors $|\Delta c(r)|$. The imaging is based on direct minimization (direct min.), direct minimization when noise is present, and direct minimization with total variation (TV) when noise is present. For all cases the situation is shown when either three or nine frequencies are used. The synthetic pressure wave field and the imaging are based on changes in the compressibility only.

IV. RESULTS AND DISCUSSION

Synthetic data were obtained by solving the forward problem^{22,50} for medium parameters representing homogeneous human liver tissue at 37 °C containing three heated lesions, two at 45 °C and one at 50 °C. Note that the tech-

nique could be applied to any acoustical configuration in principle. The corresponding medium parameters are given in Table I and were obtained by combining results presented in the literature.^{18,10,11} Note that the change in speed of sound due to heating is relatively small, *viz.*, 5 m/s between 37 °C

and 45 °C and 1 m/s between 45 °C and 50 °C. All media were assumed to be nondispersive and lossless. These losses could be included by taking complex instead of real values for the contrasts, such that the imaginary parts refer to the losses. The three-dimensional cylindrical volume contained $32 \times 32 \times 4$ cubic elements of size $2.5 \times 2.5 \times 2.5$ mm³. By choosing the number of elements to be a power of two in each direction, the usage of fast Fourier transform (FFTs) becomes feasible to compute the spatial convolutions efficiently. The frequency range of the transient signal was from 5 up to 156 kHz, which has corresponding wavelengths in the range of 10–320 mm. It will be shown that reliable imaging of small contrasts of approximately 2.5 mm in size is possible at these frequencies.

The synthetic data set is based on a tomographic setup as shown in Fig. 1, where 36 transducers are positioned circumferentially around the circular region of interest, in the plane $z=0$ m. During the imaging, it was assumed that there was no contrast outside the circular cylindrical volume enclosed by the transducers. Increasing the number of transducers did little to improve the results, while it resulted in a large increase in the computational requirements. For complicated situations with larger volumes and real objects, the results would be expected to improve with an increase in the number of transducers.

In most applications only the pressure wave field is measured, which makes it interesting to start with imaging based on the scalar wave equation as formulated in Eq. (11). This equation is obtained by neglecting the presence of a velocity wave field. Consequently, the corresponding imaging algorithms are obtained by neglecting the presence and influence of the components related to the velocity wave field in Eqs. (12)–(43). During imaging, this will result in additional changes in the compressibility, in order to compensate for changes present in the density profile. Consequently, synthetic acoustic data were computed for a speed of sound profile, a cross section of which in the (x, y) plane is shown in Fig. 2(a). Note that changes in the density were neglected and compensated for by additional changes in the compressibility. A cross section of the result of the reconstruction based on the CS formulation after 512 iterations without taking the TV into account is shown in Fig. 2(b). Here, we used the expression given in Eq. (28) with $M=1$ to obtain an estimate of the compressibility, which in combination with the known density of the background medium resulted in a speed of sound profile. Note that only three frequencies, $n_f=3$, equally spread over the full frequency range, were used. The error in the reconstructed speed of sound profile [see Fig. 2(c)] is small; $|\Delta c(\mathbf{r})| \leq 0.5$ m/s, and is distributed evenly throughout the image, except for a few points at the edges of the contrasts.

In this paper, three methods were given which can be used to reconstruct speed of sound profiles from contrast sources; direct minimization via Eq. (26), and with positivity constraint via Eq. (28) for $M=1$ and $M=2$. We tested the three methods on the noise-free synthetic data without taking TV into account and plotted the normalized error functionals for all three methods in Fig. 2(d). The results show that the fastest convergence is obtained with direct minimization and

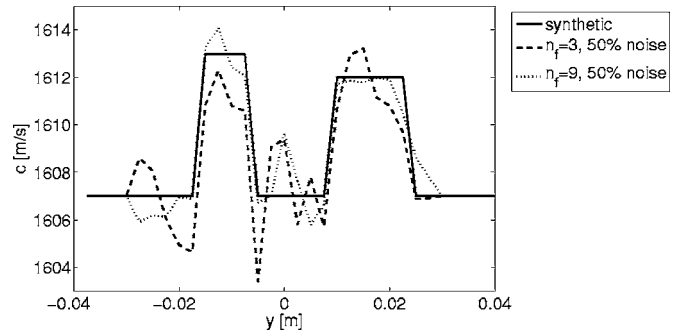


FIG. 4. A cross section of the reconstructed speed of sound profile $c(\mathbf{r})$ along the line $x=0$ m and $z=0$ m. The synthetic pressure wave field and the imaging are based on changes in the compressibility only.

for $M=1$; note that both lines coincide. In order to test the sensitivity of all methods to noise we added 50% complex white noise to the data, *viz.*,

$$\hat{p}^{\text{sct};\text{noise}}(\mathbf{r}) = \hat{p}^{\text{sct}}(\mathbf{r}) + \nu \max_r(|\hat{p}^{\text{sct}}(\mathbf{r})|), \quad (44)$$

where the complex random number ν reads $\nu = \alpha + i\beta$, with real numbers based on white noise in the range $-(1/2\sqrt{2}) \leq \alpha \leq (1/2\sqrt{2})$ and $-(1/2\sqrt{2}) \leq \beta \leq (1/2\sqrt{2})$. Hence, the absolute value of ν equals $0 \leq |\nu| \leq \frac{1}{2}$. The results for all three methods are shown in Fig. 2(e). It clearly shows that minimization with $M=2$ has become unstable, while the remaining two methods are stable and behave comparably. Finally, we increased the number of frequencies from $n_f=3$ to $n_f=9$. This stabilized the process slightly as can be observed from Fig. 2(f). The results can be explained if we replace in Eqs. (26) and (28) the reconstructed contrast source $\hat{w}_{N;k}^p(\mathbf{r})$ with its corresponding value $\hat{s}(\kappa_{N;k}^{\text{bg}} - \kappa_{N;k}^{\text{obj}}(\mathbf{r}))\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})$ based on the computed compressibility and total pressure wave field. This leads to the following expressions for $\kappa_N^{\text{obj}}(\mathbf{r})$; based on direct minimization:

$$\kappa_N^{\text{obj}}(\mathbf{r}) = \frac{\sum_{k,\Omega} \kappa_{N;k}^{\text{obj}}(\mathbf{r}) |\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})|^2}{\sum_{k,\Omega} |\hat{s}\hat{p}_{N;k}^{\text{tot}}(\mathbf{r})|^2}, \quad (45)$$

based on Eq. (28) for $M=1$

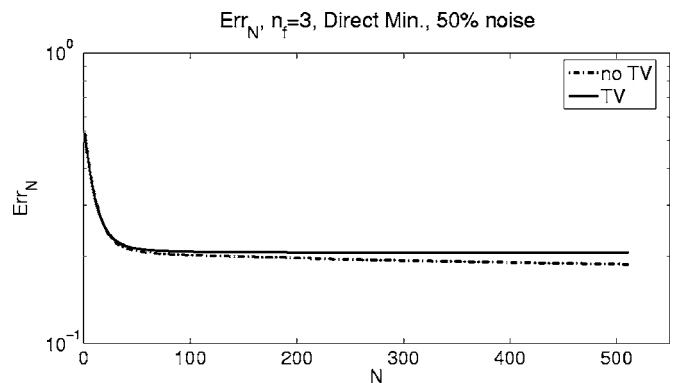


FIG. 5. The effect of total variation (TV) on the error functional Err_N .

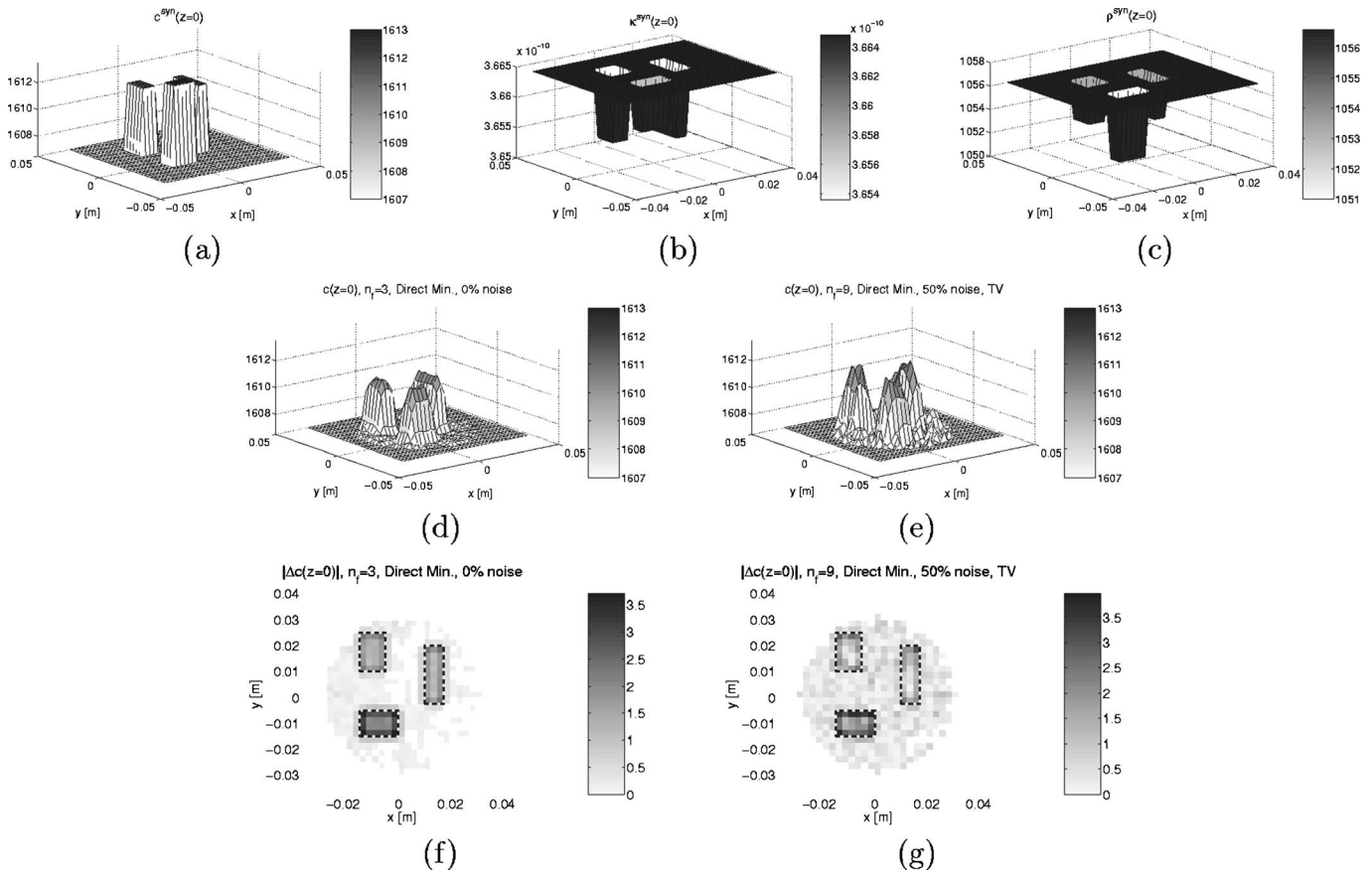


FIG. 6. Cross sections of the reconstructed speed of sound profiles $c(\mathbf{r})$ at $z=0$ m and the absolute errors $|\Delta c(\mathbf{r})|$ in the reconstructions. The imaging is based on direct minimization (direct min.) using three frequencies only and direct minimization with total variation (TV) using nine frequencies when noise is present. The synthetic pressure wave field is based on changes in both compressibility and volume density of mass while the imaging is based on changes in compressibility only.

$$\kappa_N^{\text{obj}}(\mathbf{r}) = \left(\frac{\sum_{k,\Omega} [\kappa_{N;k}^{\text{obj}}(\mathbf{r})]^2 |\hat{s}_{N;k}^{\text{tot}}(\mathbf{r})|^2}{\sum_{k,\Omega} |\hat{s}_{N;k}^{\text{tot}}(\mathbf{r})|^2} \right)^{1/2}, \quad (46)$$

and based on Eq. (28) for $M=2$

$$\kappa_N^{\text{obj}}(\mathbf{r}) = \left(\frac{\sum_{k,\Omega} [\kappa_{N;k}^{\text{obj}}(\mathbf{r})]^2}{\sum_{k,\Omega} 1} \right)^{1/2}. \quad (47)$$

Based on the large similarity between Eqs. (45) and (46), similar results are expected for these two minimization methods, while Eq. (47) is similar to taking the root-mean-square of all reconstructed values and therefore very sensitive to noise in the data.

Total variation is taken into account in order to suppress the noise present in the data. The results are shown in Fig. 3. Prior to investigating the effect of TV, we first investigated the effect of the number of frequencies used during the inversion process. For noise-free data this effect is minimal as can be observed from Figs. 3(a), 3(d), 3(g), and 3(j) and Fig. 4. However, in the presence of noise, increasing the number of frequencies results in a clear improvement in the reconstructed contrast profiles; see Figs. 3(b), 3(e), 3(h), and 3(k). Finally, TV is applied in order to regularize the problem.

This proves to be a very effective method to handle noise, as can be observed from Figs. 3(c), 3(f), 3(i), and 3(l). In addition, it is observed that the usage of TV does not always have a positive effect on the convergence speed, as can be concluded from Fig. 5. Here, a decrease in convergence speed is observed in the case data with noise is used.

In most cases, changes in speed of sound occur due to the presence of changes in both the compressibility and the density. This results in both a scattered scalar pressure and a scattered vectorial velocity wave field. Consequently, it is expected that a scalar model is not sufficient to obtain accurate speed of sound profiles. The effect of neglecting the vectorial component of the full acoustic wave equation in the reconstruction process is tested by computing synthetic data based on the same speed of sound profile as used for the previous results, but now taking changes in both compressibility and density into account; see Figs. 6(a)–6(c). Using three frequency components in the absence of noise enables the localization of all three objects, but with amplitudes which are lower than the actual synthetic values. The error in the reconstructed speed of sound profile is approximately five times bigger than the results obtained with the full scalar model. Increasing the number of frequencies corrects partly for this effect. In the case where 50% noise is present in the data and TV is applied, the error increases approximately by a factor of $2\frac{1}{2}$ compared with the same situation where syn-

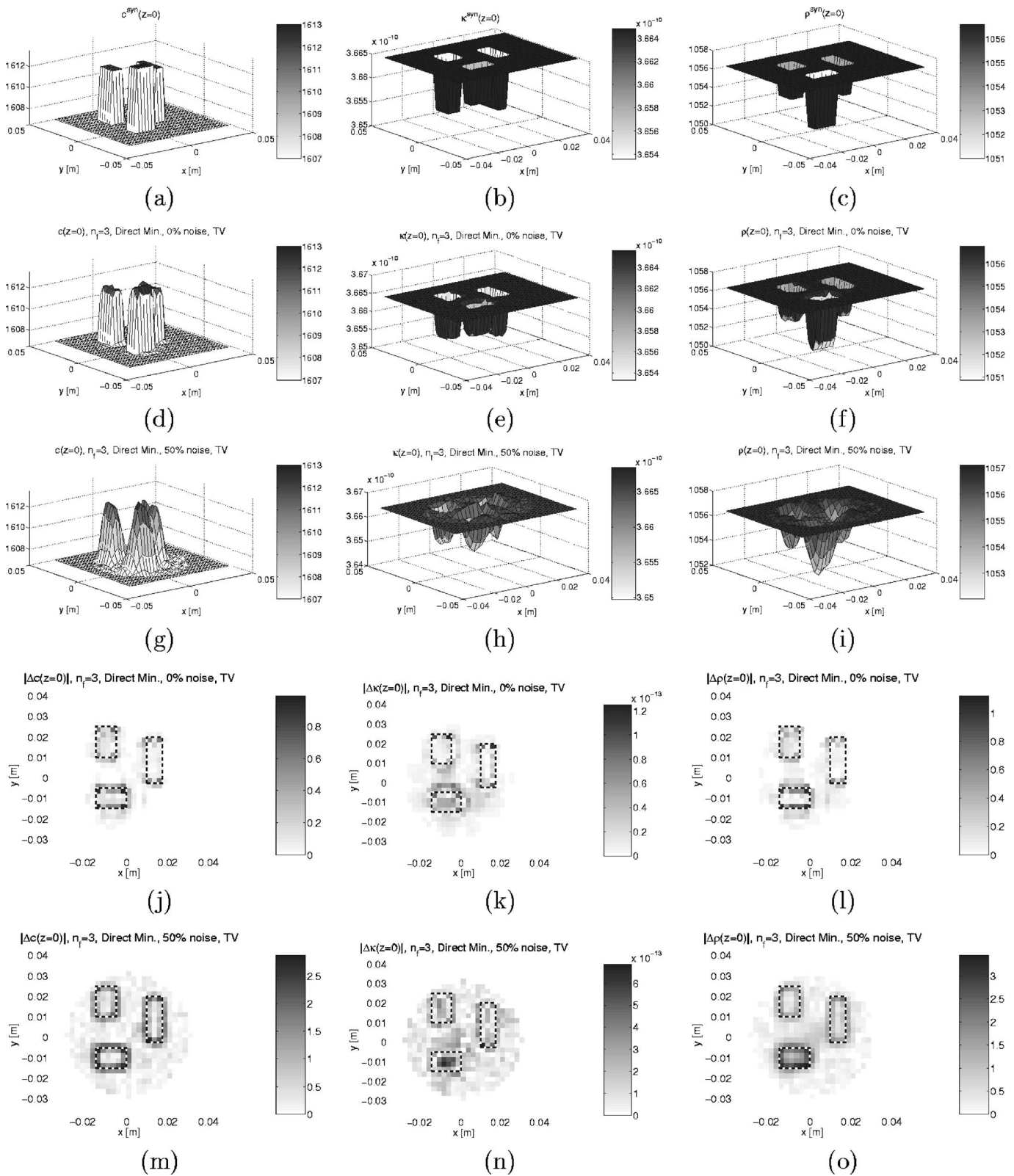


FIG. 7. Cross sections of the reconstructed speed of sound profiles $c(r)$ at $z=0$ m based on the compressibility $\kappa(r)$ and density $\rho(r)$ profiles, and their absolute errors. The imaging is based on direct minimization (direct min.) with total variation (TV) using three frequencies. The synthetic pressure and velocity wave field and the imaging are based on changes in both compressibility and volume density of mass.

thetic data are obtained from the scalar model. Hence, the scalar reconstruction alone is insufficient in situations where there are changes in both density and compressibility.

Finally, the full vectorial inversion scheme is tested on the same synthetic data set as discussed previously. The syn-

thetic profile was based on both changes in compressibility and density, and during the imaging both the pressure and velocity wave fields were taken into account. The results are shown in Figs. 7 and 8. In the absence of noise, a reconstructed speed of sound profile based on three frequencies is

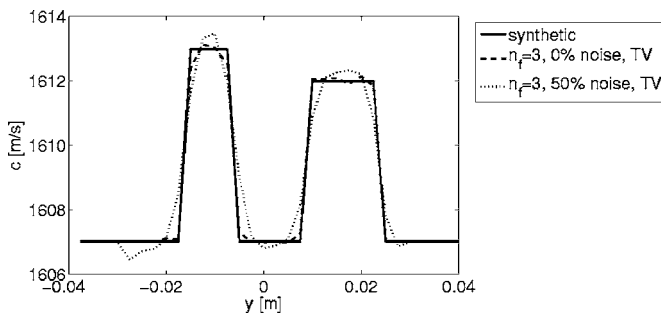


FIG. 8. A cross section of the reconstructed speed of sound profile $c(r)$ along the line $x=0$ m and $z=0$ m. The synthetic pressure and velocity wave field and the imaging are based on changes in both compressibility and volume density of mass.

almost $2\frac{1}{2}$ times as accurate as the reconstructed profile using the pressure field only; compare Fig. 6(f) with Fig. 7(j). In the presence of noise the error in the reconstruction with the full model in combination with TV based on three frequencies is still approximately $1\frac{1}{2}$ times smaller than the error in the image based on the scalar inversion model using nine frequencies; compare Fig. 6(g) with Fig. 7(m). More importantly, the full vectorial imaging method allows the reconstruction of both density and compressibility profiles; see Figs. 7(e), 7(f), 7(h), and 7(i). Note that the obtained accuracy is excellent, even when using frequencies with corresponding wavelengths larger than the contrasts to be detected. The images could be improved further by increasing the number of nodes and frequencies, but at the expense of computational efficiency.

V. CONCLUSION

In this paper we described a full vectorial three-dimensional inversion method based on the acoustic wave equations. The combined scalar pressure and vectorial velocity wave field description allows the reconstruction of the acoustic medium parameters compressibility and volume density of mass with a conjugate gradient contrast source (CG-CS) inversion scheme. This minimizes two error functionals in an iterative scheme. The first error functional is referred to as the data equation and relates the measured scattered acoustic wave fields to contrast sources in the volume of interest, while the second error functional is referred to as the object equation and relates the contrast sources to the contrasts and the total wave fields. The contrast profiles are obtained by applying direct minimization on the second error functional for computed total wave fields based on reconstructed contrast sources. Testing this method on synthetic measured data resulted in the reconstruction of profiles of the acoustic medium parameters compressibility and mass density in the region of interest. From these reconstructed medium parameters, speed of sound profiles are obtained in a straightforward manner. In the presence of noise, additional regularization becomes essential; it stabilizes the minimization procedure and reduces the blurring in the reconstructed contrast profiles. This regularization is obtained by taking the total variation (TV) of the reconstructed contrast profiles into account and has proven to be effective for handling data after adding 50% complex-valued white noise.

In many situations, only the pressure wave field is measured and the scalar wave equation is used as a starting point for computing speed of sound profiles. Therefore, we also developed a contrast source inversion scheme for pressure wave fields only. We tested the algorithms for two situations based on the same speed of sound profile: (i) a profile based solely on changes in the compressibility, and (ii) a profile based on changes in both the compressibility and density. In both situations the imaging method using only the pressure wave field was successful in localizing the contrasts; however, the amplitude of the reconstructed speed of sound profile in the second case, with the combined contrast, was lower than the amplitude of the synthetic profile or the reconstructed profile in the case where there was only a contrast in compressibility.

To conclude, in situations where changes in speed of sound are caused by changes in both compressibility and volume density of mass, it is still possible to obtain speed of sound profiles by using pressure wave fields only. The results improve significantly by using the full vectorial acoustic wave equation, which enables the use of both the velocity and the pressure wave field, and allows the reconstruction of both acoustic medium parameters density and compressibility. The three-dimensional CG-CS method with TV as an additional regularization tool is an effective method to obtain these results with good resolution.

ACKNOWLEDGMENTS

This work was supported by Marie Curie Intra-European Fellowship Number MEIF-CT-2003-501333. We also gratefully acknowledge the Boole Centre for Research in Informatics, University College Cork, Ireland, for the usage of their computer facility.

- ¹S. J. Norton and M. Linzer, "Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical, and spherical apertures," *IEEE Trans. Biomed. Eng.* **28**(2), 202–220 (1981).
- ²J. F. Greenleaf and R. C. Bahn, "Clinical imaging with transmissive ultrasonic computerized-tomography," *IEEE Trans. Biomed. Eng.* **28**(2), 177–185 (1981).
- ³A. J. Devaney, "A filtered back-propagation algorithm for diffraction tomography," *Ultrason. Imaging* **4**(4), 336–350 (1982).
- ⁴H. W. Jones, "Recent activity in ultrasonic tomography," *Ultrasonics* **31.5**, 353–360 (1993).
- ⁵S. A. Johnson, D. A. Christensen, C. C. Johnson, J. F. Greenleaf, and B. Rajagopalan, "Noninvasive measurement of microwave or ultrasound induced hyperthermia by acoustic temperature tomography," *Proceedings 1977 IEEE Ultrasonics Symposium*, 977–982 (1977).
- ⁶N. T. Sanghvi, F. J. Fry, R. Bihrlé, R. S. Foster, M. H. Phillips, J. Syrus, A. V. Zaitsev, and C. W. Hennige, "Noninvasive surgery of prostate tissue by high-intensity focused ultrasound," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **43**(6), 1099–1110 (1996).
- ⁷J. E. Kennedy, G. R. ter Haar, and D. Cranston, "High intensity focused ultrasound: Surgery of the future?," *Br. J. Audiol.* **76**, 590–599 (2003).
- ⁸G. Ter Haar, "Ultrasound focal beam surgery," *Ultrasound Med. Biol.* **21**, 1089–1100 (1995).
- ⁹C. J. Diederich and K. Hynynen, "Ultrasound technology for hyperthermia," *Ultrasound Med. Biol.* **25**, 871–887 (1999).
- ¹⁰N. R. Miller, J. C. Bamber, and G. R. ter Haar, "Imaging of temperature-induced echo strain: Preliminary *in vitro* study to assess feasibility for guiding focused ultrasound surgery," *Ultrasound Med. Biol.* **30**(3), 345–356 (2004).
- ¹¹R. Souchon, G. Bouchoux, E. Maciejko, C. Lafon, D. Cathignol, M. Bertrand, and J. Y. Chapelon, "Monitoring the formation of thermal lesions with heat-induced echo-strain imaging: A feasibility study," *Ultrasound*

- Med. Biol. **31**(2), 251–259 (2005).
- ¹²N. R. Miller, J. C. Bamber, and P. M. Meaney, “Fundamental limitations of noninvasive temperature imaging by means of ultrasound echo strain estimation,” *Ultrasound Med. Biol.* **28**(10), 1319–1333 (2002).
- ¹³K. A. Wear, R. F. Wagner, M. F. Insana, and T. J. Hall, “Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **40**(1), 50–58 (1993).
- ¹⁴N. T. Sanghvi, R. S. Foster, F. J. Fry, R. Bihrl, C. Hennige, and L. V. Hennige, “Ultrasound intracavity system for imaging, therapy planning and treatment of focal disease,” *Proceedings IEEE Ultrasonics Symposium 1992*, **2**, 1249–1253 (October 1992).
- ¹⁵R. Seip, P. VanBaren, and E. S. Ebbini, “Dynamic focusing in ultrasound hyperthermia treatments using implantable hydrophone arrays,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **41**(5), 706–713 (1994).
- ¹⁶R. Seip and E. S. Ebbini, “Noninvasive estimation of tissue temperature response to heating fields using diagnostic ultrasound,” *IEEE Trans. Biomed. Eng.* **42**(8), 828–839 (1995).
- ¹⁷A. N. Amini, E. S. Ebbini, and T. T. Georgiou, “Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques,” *IEEE Trans. Biomed. Eng.* **52**(2), 221–228 (2005).
- ¹⁸T. D. Mast, “Empirical relationships between acoustic parameters in human soft tissues,” *ARLO* **1**(2), 37–42 (2000) and references herein.
- ¹⁹J.-W. Jeong, T.-S. Kim, D. C. Shin, S. Do, M. Singh, and V. Z. Marmarelis, “Soft tissue differentiation using multiband signatures of high resolution ultrasonic transmission tomography,” *IEEE Trans. Med. Imaging* **24**(3), 399–408 (2005).
- ²⁰Y. Nawata, and K. Kaneko, “Measurement of temperature distribution in phantom body by an ultrasonic CT method,” *Proceedings of the ASME/JSME Joint Thermal Engineering Conference* **3**, 469–474 (1999).
- ²¹D. Kourtiche, M. Nadim, G. Kontaxakis, C. Marchal, and G. Prieur, “Temperature measurements using US tomography: Theoretical aspects,” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society Proc. IEEE Eng. Med. Biol. Soc.* **13**(1), 325–326 (1991).
- ²²K. W. A. van Dongen and W. M. D. Wright, “A forward model and conjugate gradient inversion technique for low-frequency ultrasonic imaging,” *J. Acoust. Soc. Am.* **120**, 2086 (2006); **120**(4), 2086–2095 (2006).
- ²³A. T. de Hoop, *Handbook of Radiation and Scattering of Waves: Acoustic Waves in Fluids, Elastic Waves in Solids, Electromagnetic Waves* (Academic, London, 1995).
- ²⁴P. M. van den Berg, “Iterative schemes based on minimization of a uniform error criterion,” *PIERS 5, Application of Conjugate Gradient Method to Electromagnetics and Signal Analysis*, edited by T. K. Sarkar (Elsevier, New York, 1991), pp. 27–66.
- ²⁵R. E. Kleinman and P. M. van den Berg, “Iterative methods for solving integral equations,” *PIERS 5, Application of Conjugate Gradient Method to Electromagnetics and Signal Analysis*, edited by T. K. Sarkar (Elsevier, New York, 1991), pp. 67–102.
- ²⁶R. E. Kleinman and P. M. van den Berg, “A modified gradient method for two-dimensional problems in tomography,” *J. Comput. Appl. Math.* **42**, 17–35 (1992).
- ²⁷P. M. van den Berg and R. E. Kleinman, “A total variation enhanced modified gradient algorithm for profile reconstruction,” *Inverse Probl.* **11**, L5–L10 (1995).
- ²⁸R. Acar and C. R. Vogel, “Analysis of bounded variation penalty methods for ill-posed problems,” *Inverse Probl.* **10**, 1217–1229 (1994).
- ²⁹S. Osher and L. Rudin, “Feature-oriented image enhancement using shock filters,” *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **27**, 919–940 (1990).
- ³⁰L. Rudin, S. Oshet, and C. Fatemi, “Nonlinear total variation based noise removal algorithm,” *Physica D* **60D**, 259–268 (1992).
- ³¹T. M. Habashy, M. L. Oristaglio, and A. T. de Hoop, “Simultaneous nonlinear reconstruction of two-dimensional permittivity and conductivity,” *Radio Sci.* **29**, 1101–1118 (1994).
- ³²R. E. Kleinman and P. M. van den Berg, “An extended range modified gradient technique for profile inversion,” *Radio Sci.* **28**, 877–884 (1993).
- ³³R. E. Kleinman and P. M. van den Berg, “Two-dimensional location and shape reconstruction,” *Radio Sci.* **29**, 1157–1169 (1994).
- ³⁴P. M. van den Berg and R. E. Kleinman, “A contrast source inversion method,” *Inverse Probl.* **13**, 1607–1620 (1997).
- ³⁵P. M. van den Berg and A. Abubakar, “Inverse scattering algorithms based on contrast source integral representations,” *Inverse Probl. Eng.* **10**, 559–576 (2002).
- ³⁶P. M. van den Berg and A. Abubakar, “Contrast Source Inversion Method: State of Art,” in *PIER 34, Electromagnetic Waves*, edited by J. A. Kong (EMW, Cambridge, MA, 2001), pp. 189–218.
- ³⁷G. Pelekanos, A. Abubakar, and P. M. van den Berg, “Contrast source inversion methods in elastodynamics,” *J. Acoust. Soc. Am.* **114**(5), 2825–2834 (2003).
- ³⁸A. Abubakar, T. M. Habashy, P. M. van den Berg, and D. Gisolf, “The diagonalized contrast source approach: An inversion method beyond the Born approximation,” *Inverse Probl.* **21**(2), 685–702 (2005).
- ³⁹A. Abubakar and P. M. van den Berg, “Iterative forward and inverse algorithms based on domain integral equations for three-dimensional electric and magnetic objects,” *J. Comput. Phys.* **195**, 236–262 (2004).
- ⁴⁰P. M. van den Berg and A. Abubakar, “Contrast source inversion of 3D objects with electric and magnetic contrasts,” *2004 URSI/EMTS Proceedings of the International Symposium on Electromagnetic Theory*, 700–702 (2004).
- ⁴¹K. W. A. van Dongen and W. M. D. Wright, “A contrast source inversion scheme for imaging acoustic contrast,” *Proceedings IEE Irish Signals and Systems Conference*, Dublin, Ireland, 1–2 September, 278–283 (2005).
- ⁴²A. J. Devaney, “Variable density acoustic tomography,” *J. Acoust. Soc. Am.* **78**(1), 120–130 (1985).
- ⁴³M. Moghaddam and W. C. Chew, “Simultaneous inversion of compressibility and density in the acoustic inverse problem,” *Inverse Probl.* **9**, 715–730 (1993).
- ⁴⁴S. Mensah and J. P. Lefebvre, “Enhanced compressibility tomography,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **44**(6), 1245–1252 (1997).
- ⁴⁵K. Fisher, S. Lehman, and D. Chambers, “Development of a multiview time-domain imaging algorithm with a Fermat correction,” *J. Acoust. Soc. Am.* **118**(5), 3122–3128 (2005).
- ⁴⁶C. Lu, W. C. Chew, and G. P. Otto, “Image reconstruction with acoustic measurement using distorted Born iterative method,” *Ultrason. Imaging* **18**, 140–156 (1996).
- ⁴⁷N. Bleistein and J. K. Cohen, “Nonuniqueness in the inverse source problem in acoustics and electromagnetics,” *J. Math. Phys.* **18**, 194–201 (1997).
- ⁴⁸J. T. Fokkema and P. M. van den Berg, *Seismic Applications of Acoustic Reciprocity* (Elsevier Science, Amsterdam, 1993).
- ⁴⁹P. M. van den Berg, A. Abubakar, and J. T. Fokkema, “Multiplicative regularization for contrast profile inversion,” *Radio Sci.* **38**(2), VIC 23-1-10 (2003).
- ⁵⁰K. W. A. van Dongen, C. Brennan, and W. M. D. Wright, “A reduced forward operator for acoustic scattering problems,” *Proceedings IEE Irish Signals and Systems Conference*, Dublin, Ireland, 1–2 September 2005, pp. 294–299 (2005).

Statistically optimized near field acoustic holography using an array of pressure-velocity probes^{a)}

Finn Jacobsen^{b)} and Virginie Jaud^{c)}

Acoustic Technology, Ørsted-DTU, Technical University of Denmark, Building 352, Ørsted's Plads, DK-2800 Kgs. Lyngby, Denmark

(Received 1 September 2006; revised 15 December 2006; accepted 19 December 2006)

Statistically optimized near field acoustic holography (SONAH) differs from conventional near field acoustic holography (NAH) by avoiding spatial Fourier transforms; the processing is done directly in the spatial domain. The main advantage of SONAH compared with NAH is that the usual requirement of a measurement aperture that extends well beyond the source can be relaxed. Both NAH and SONAH are based on the assumption that all sources are on one side of the measurement plane whereas the other side is source free. An extension of the SONAH procedure based on measurement with a double layer array of pressure microphones has been suggested. The double layer technique makes it possible to distinguish between sources on the two sides of the array and thus suppress the influence of extraneous noise coming from the “wrong” side. It has also recently been demonstrated that there are significant advantages in NAH based on an array of acoustic particle velocity transducers (in a single layer) compared with NAH based on an array of pressure microphones. This investigation combines the two ideas and examines SONAH based on an array of pressure-velocity intensity probes through computer simulations as well as experimentally. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434245]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.20.Rz [EJS]

Pages: 1550–1558

I. INTRODUCTION

Near field acoustic holography (NAH) is an experimental technique that makes it possible to reconstruct three-dimensional sound fields from measurements on two-dimensional surfaces. This can be extremely useful, and NAH is a well-established tool for visualizing and analyzing sound fields near sources of noise.^{1,2} Conventional planar NAH is based on discrete spatial Fourier transforms of sound pressure data measured with a microphone array. However, to avoid serious truncation errors caused by the finite two-dimensional spatial transform (“leakage” in the wave number domain) the array must extend well beyond the source so that the sound pressure has decayed to an insignificant level near the edges of the array.²

Statistically optimized near field acoustic holography (SONAH) is an interesting variant of NAH developed a few years ago by Steiner and Hald.³ It has the great advantage of avoiding spatial transforms and thus the mentioned truncation effects; therefore the measurement array can be smaller than the source.^{3–5}

Both NAH and SONAH are usually based on measurement of the sound pressure. However, an acoustic particle velocity transducer has been available for some years,⁶ and it

has recently been demonstrated that NAH based on measurement of the normal component of the particle velocity is more accurate than pressure-based NAH.⁷ The purpose of this paper is to examine whether there is a similar advantage in measuring the particle velocity with the SONAH procedure and, in particular, to examine whether one can combine predictions based on measurement of the sound pressure with predictions based on measurement of the normal component of the particle velocity and thereby be able to distinguish between sound coming from the two sides of the measurement plane in the same way as one can do that with a double layer array of pressure transducers.^{8–10}

II. OUTLINE OF THEORY

A. SONAH based on measurement of the sound pressure

The following derivation is based on Hald.⁴ In planar SONAH the “propagator” that transforms data from one plane to another is a transfer matrix that works directly on the measured data. Thus the sound pressure at an arbitrary position above the source, $\mathbf{r}=(x,y,z)$ (where $z>0$), is expressed as a weighted sum of sound pressures measured at N positions ($\mathbf{r}_{h,n}$) in the hologram plane ($z=z_h$),

$$p(\mathbf{r}) \approx \sum_{n=1}^N c_n(\mathbf{r})p(\mathbf{r}_{h,n}) = \mathbf{p}^T(\mathbf{r}_h)\mathbf{c}(\mathbf{r}), \quad (1)$$

where T indicates that the column vector $\mathbf{p}(\mathbf{r}_h)$ is transposed. The transfer vector $\mathbf{c}(\mathbf{r})$ does not depend on the sound field but only on positions. It is determined by requiring that an infinite set of propagating and evanescent elementary waves of the form

^{a)}Portions of this work were presented in “Statistically optimised near field acoustic holography based on particle velocity measurements,” Proceedings of the 13th International Congress on Sound and Vibration, Vienna, Austria, July 2006, and in “Statistically optimised near field acoustic holography with an array of pressure-velocity intensity probes,” Proceedings of Inter-Noise 2006, Honolulu, HI, December 2006.

^{b)}Author to whom correspondence should be addressed; electronic mail: fja@oersted.dtu.dk

^{c)}Electronic mail: virginie.jaud@gmail.com

$$\Phi_m(\mathbf{r}) = e^{-j(k_{x,m}x + k_{y,m}y + k_{z,m}z)}, \quad m = 1, 2, \dots, M, \quad M \rightarrow \infty, \quad (2)$$

where

$$k_{z,m} = \begin{cases} \sqrt{k^2 - k_{x,m}^2 - k_{y,m}^2} & \text{for } \sqrt{k_{x,m}^2 + k_{y,m}^2} \leq k \\ -j\sqrt{k_{x,m}^2 + k_{y,m}^2 - k^2} & \text{for } \sqrt{k_{x,m}^2 + k_{y,m}^2} > k \end{cases} \quad (3)$$

(with the $e^{j\omega t}$ sign convention) are projected from the measurement plane to the prediction plane [in other words, satisfy Eq. (1)] with optimal accuracy. As can be seen from Eq. (2) all the elementary waves have an amplitude of unity in the source plane, from which it follows that the transfer vector $\mathbf{c}(\mathbf{r})$ is optimized for a “white” wave number spectrum in this plane; hence the term “statistically optimal” near field acoustic holography.³⁻⁵ In vector and matrix form,

$$\boldsymbol{\alpha}(\mathbf{r}) \approx \mathbf{A}\mathbf{c}(\mathbf{r}), \quad (4)$$

where $\boldsymbol{\alpha}(\mathbf{r})$ is a column vector with M elements, $[\boldsymbol{\alpha}(\mathbf{r})]_m = \Phi_m(\mathbf{r})$, and \mathbf{A} is an M by N matrix, $[\mathbf{A}]_{mn} = \Phi_m(\mathbf{r}_{h,n})$. Since $M > N$ Eq. (4) is overdetermined. The least-squares solution is

$$\mathbf{c}(\mathbf{r}) = (\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r}), \quad (5)$$

where \mathbf{I} is the identity matrix, θ is a regularization parameter, and the superscript H indicates the Hermitian transpose.³⁻⁵ Note that $\mathbf{A}^H\mathbf{A}$ is an N by N matrix and $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$ is a column vector with N elements. It now follows that the sound pressure in the prediction plane is

$$p(\mathbf{r}) = \mathbf{p}^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r}). \quad (6)$$

The normal component of the particle velocity in the prediction plane is obtained from Eq. (6) using Euler’s equation of motion,

$$u_z(\mathbf{r}) = \frac{-1}{j\omega\rho}\mathbf{p}^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\frac{\partial\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})}{\partial z} \\ = \mathbf{p}^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\beta}(\mathbf{r}), \quad (7)$$

where the only quantity that is differentiated is the vector $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$ since this is the only quantity that depends on z .

In practice it is desirable to estimate the sound pressure and the particle velocity not just at a single position \mathbf{r} but in a grid of N points in a plane; thus the column vectors $\mathbf{A}^H\boldsymbol{\alpha}$ and $\mathbf{A}^H\boldsymbol{\beta}$ become N by N matrices, and the left-hand sides of Eqs. (6) and (7) become column vectors. When $M \rightarrow \infty$ the elements of $\mathbf{A}^H\mathbf{A}$, $\mathbf{A}^H\boldsymbol{\alpha}$, and $\mathbf{A}^H\boldsymbol{\beta}$ turn into integrals over k_x and k_y that can be evaluated numerically.³⁻⁵ The infinite column vector $\boldsymbol{\alpha}(\mathbf{r})$ and the infinite matrix \mathbf{A} are not determined at all.

B. SONAH based on measurement of the particle velocity

It is a fairly simple matter to extend the foregoing considerations to the case where the normal component of the particle velocity rather than the sound pressure is measured. For any of the elementary waves the pressure ratio $p(\mathbf{r})/p(\mathbf{r}_{h,n})$ is identical to the particle velocity ratio $u_z(\mathbf{r})/u_z(\mathbf{r}_{h,n})$, which leads to the conclusion that the transfer

vector given by Eq. (5) also projects the normal component of the particle velocity from the measurement plane to the prediction plane. Thus

$$u_z(\mathbf{r}) = \mathbf{u}_z^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r}). \quad (8)$$

One can also predict the sound pressure from the measured normal component of the particle velocity, as follows:

$$p(\mathbf{r}) = \mathbf{u}_z^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}(-j\omega\rho) \int \mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})dz \\ = \mathbf{u}_z^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\gamma}(\mathbf{r}), \quad (9)$$

where the only quantity that is integrated is the vector $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$ since this is the only quantity that depends on z .

Equations (8) and (9) can, of course, be extended from a single position to N positions in the prediction plane in the same way as Eqs. (6) and (7); and in the limit of $M \rightarrow \infty$ the elements of $\mathbf{A}^H\boldsymbol{\gamma}$ become integrals that can be evaluated numerically in the same way as the elements of $\mathbf{A}^H\mathbf{A}$, $\mathbf{A}^H\boldsymbol{\alpha}$, and $\mathbf{A}^H\boldsymbol{\beta}$. See the Appendix.

C. SONAH based on measurement of the pressure and the particle velocity

One cannot distinguish between sound coming from the two sides of the measurement plane if one measures only the sound pressure or the normal component of the particle velocity. However, if both quantities are measured then the fact that the particle velocity unlike the pressure is a vector component that changes its sign if the source is moved to a symmetrical position on the other side of the measurement plane makes it possible to separate the two contributions from each other. Thus the pressure generated by the primary source can be estimated as the average of a pressure- and a particle velocity-based estimate,

$$p(\mathbf{r}) = \frac{1}{2}(\mathbf{p}^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r}) \\ + \mathbf{u}_z^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\gamma}(\mathbf{r})), \quad (10)$$

and the particle velocity can be estimated in a similar way,

$$u_z(\mathbf{r}) = \frac{1}{2}(\mathbf{u}_z^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r}) \\ + \mathbf{p}^T(\mathbf{r}_h)(\mathbf{A}^H\mathbf{A} + \theta^2\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\beta}(\mathbf{r})). \quad (11)$$

However, since the transfer matrices have been optimized for elementary waves coming from the source plane one should perhaps not expect the same accuracy in the general case where the disturbing noise is not coming from a source placed symmetrically. A future investigation will examine the influence of the wave number weighting.

In what follows this combination of estimates based on pressure and velocity is referred to as “the p - u method.” In principle one could also determine the pressure and the particle velocity generated by the disturbing sources on the other side of the measurement plane by subtracting instead of adding the pressure- and particle velocity-based estimates.

III. A SIMULATION STUDY

To examine the performance of the methods described in the foregoing a simulation study has been carried out. Sev-

eral test cases were examined, including combinations of point sources and point driven simply supported vibrating panels mounted in infinite baffles without and with disturbing background noise from a monopole placed on the “wrong” side of the measurement plane. The structural sources were modeled by modal sums, and the radiated sound fields were calculated from numerical approximations to Rayleigh’s first integral.² In all cases there were 8×8 sound pressure and particle velocity transducers in the simulated measurement array; the measurement plane was 6 cm from the source, the prediction plane was 3 cm from the source, and both planes had dimensions 21×21 cm corresponding to 3 cm between adjacent positions.

The recommended value of the regularization parameter θ is related to the signal-to-noise ratio,^{3,4} which, however, can be difficult to estimate.¹¹ The results presented in what follows are based on an alternative regularization technique, the generalized cross validation method.^{11,12} This method (which is similar to the technique described in Ref. 5) has the advantage that no *a priori* knowledge about the signal-to-noise ratio of the data is needed.

A. Simple SONAH with ideal transducers

Figure 1 shows a typical set of results from the first test case, a baffled 5 mm steel panel of size 21×21 cm, that is, of the same size as the measurement plane, driven at 500 Hz by a point force near a corner. (Rayleigh’s integral was approximated by 32×32 point sources on the baffle.) Figure 1(a) compares the pressure predicted from the pressure and from the velocity with the “true” pressure along a diagonal, and Fig. 1(b) compares the velocity predicted from the velocity and from the pressure with the “true” velocity (note that the velocity reference is 50 nm/s). Inspection shows that the best results are obtained if the pressure is predicted from the pressure and the velocity is predicted from the velocity.

B. The influence of amplitude and phase mismatch

Figure 2 demonstrates the influence of amplitude mismatch on SONAH predictions of the sound field generated by a baffled 5 mm steel plate of size 1×1 m, driven by a point force near a corner. (In this case Rayleigh’s integral was approximated by 64×64 point sources on the baffle.) Figure 2 shows five outcomes of a stochastic experiment where independent random amplitude mismatch evenly distributed between ± 0.5 dB has been introduced in each of the “measured” pressure and velocity signals. Not surprisingly, such amplitude mismatch has a negative influence on the performance of the SONAH procedure. By far the most serious effect occurs when the particle velocity is predicted from pressure data, and pressure-to-pressure and particle velocity-to-particle velocity predictions are the least affected. These observations agree with observations from conventional NAH simulations.⁷

Very similar results have been obtained with phase mismatch (not shown).

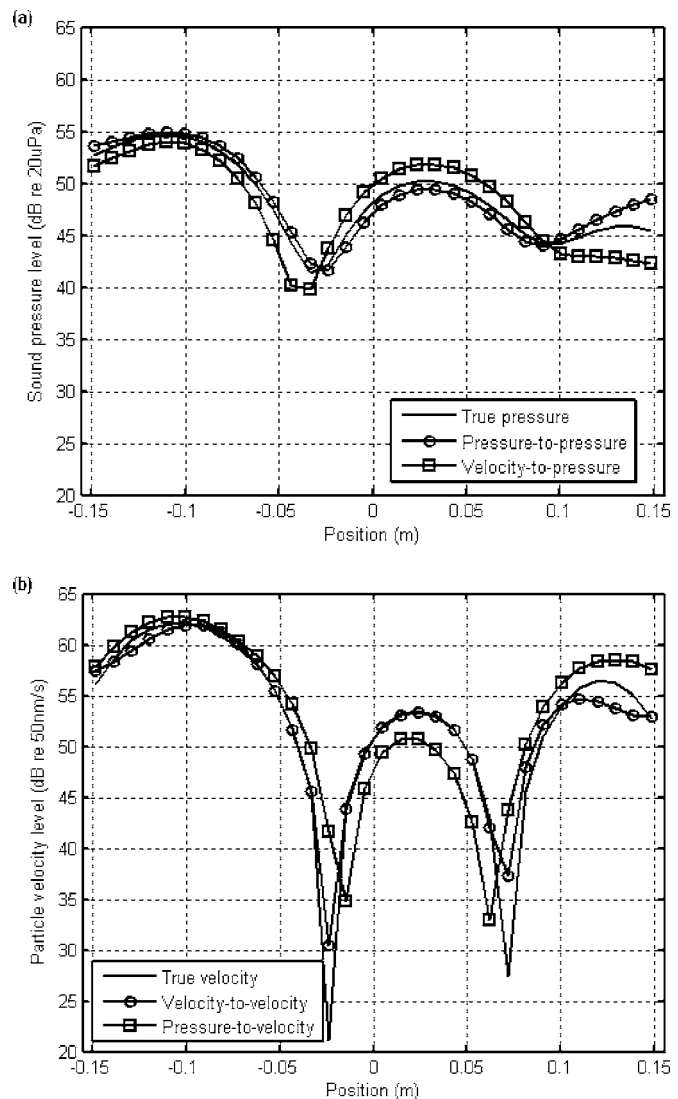


FIG. 1. Small panel in a baffle driven at 500 Hz. (a) “True” and predicted sound pressure and (b) true and predicted particle velocity in a diagonal across the prediction plane.

C. The pressure-velocity method

The *p-u* method based on Eqs. (10) and (11) has been examined with combinations of simple sources disturbed by other simple sources, with vibrating panels disturbed by simple sources, and with an infinite vibrating plate disturbed by simple sources.

Figures 3 and 4 show the results of a test case with two interfering plane waves propagating in the *yz* plane at angles of $\pm 40^\circ$ from the normal to the *z* plane at 3 kHz disturbed by two monopoles on the “wrong” side of the measurement plane, located at $(0, 0, 16)$ and $(-12, -2, 16)$, respectively [coordinates in centimeters, and $(0, 0, 6)$ being the coordinate of the center of the measurement plane]. The two interfering plane waves might conceivably be generated by two supersonic bending waves traveling in opposite directions on an infinite plate at $z=0$ (or in any other parallel plane on the “right” side of the measurement plane). This is an extreme test case since (i) the primary source is infinite, so the measurement array is infinitely smaller than the source, and (ii) the wave number spectrum of the primary source consists of

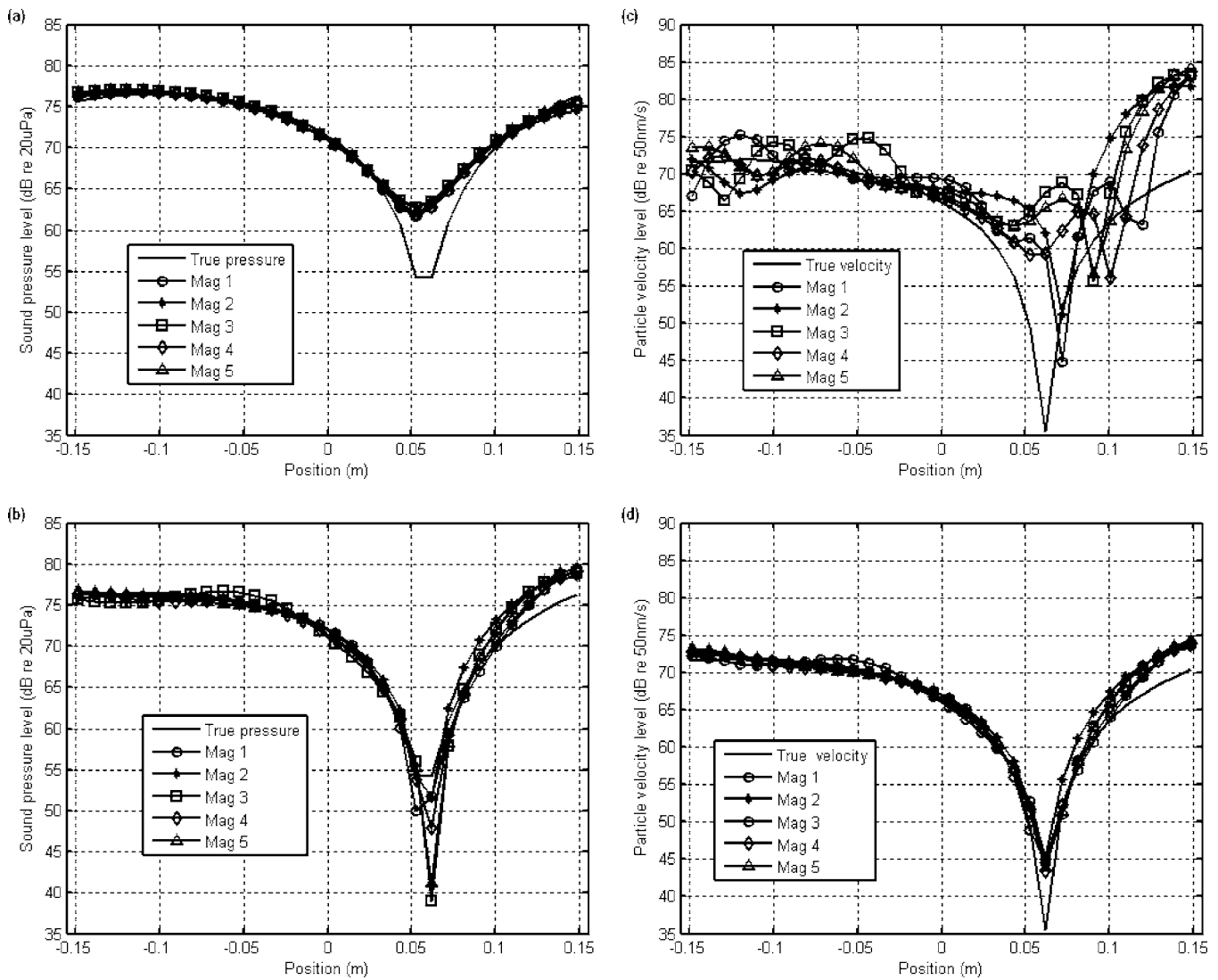


FIG. 2. The influence of transducer amplitude mismatch on predictions near a large panel driven at 600 Hz. (a) Pressure predicted from pressure; (b) pressure predicted from particle velocity; (c) particle velocity predicted from pressure; (d) particle velocity predicted from particle velocity. All quantities are shown in a diagonal across the prediction plane.

two delta functions and is thus very far from “white.” As can be seen in Fig. 3 the two monopoles generate a sound pressure level comparable to the level of the two plane waves in the prediction plane. Figure 4 demonstrates that predictions based only on pressure or velocity are severely contaminated by the disturbing sources whereas the p - u method performs reasonably well. The reason for the poor performance of predictions based on either the pressure or the particle velocity is, of course, that the predicted sound field generated by the disturbing monopoles will be added to the predicted primary sound field (in pressure-based predictions) or subtracted from it (in velocity-based predictions), since these predictions are based on the assumption of a source-free region on the “wrong” side of the measurement plane. Similar simulations (not shown here, but presented in Ref. 13) have demonstrated that pressure-to-pressure and particle velocity-to-particle velocity predictions in the absence of extraneous noise are more accurate than the p - u method. All in all it can be concluded that (i) SONAH can indeed cope with sources that are larger than the measurement array—even infinitely

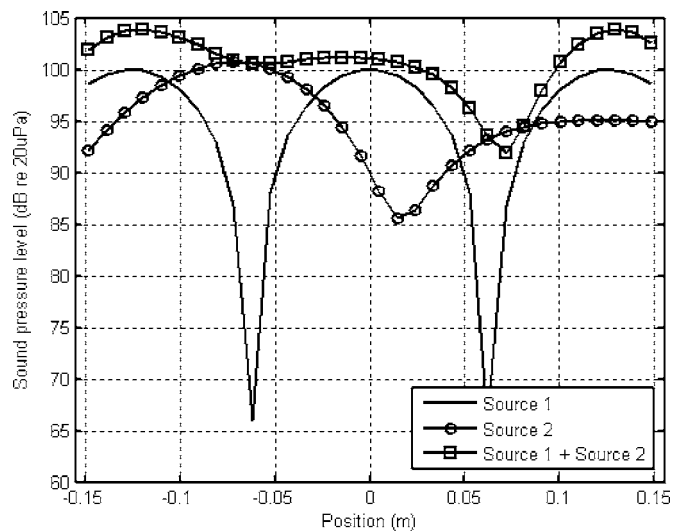


FIG. 3. Sound pressure level in a diagonal across the prediction plane generated by two interfering plane waves and two monopoles on the “wrong” side of the measurement plane.

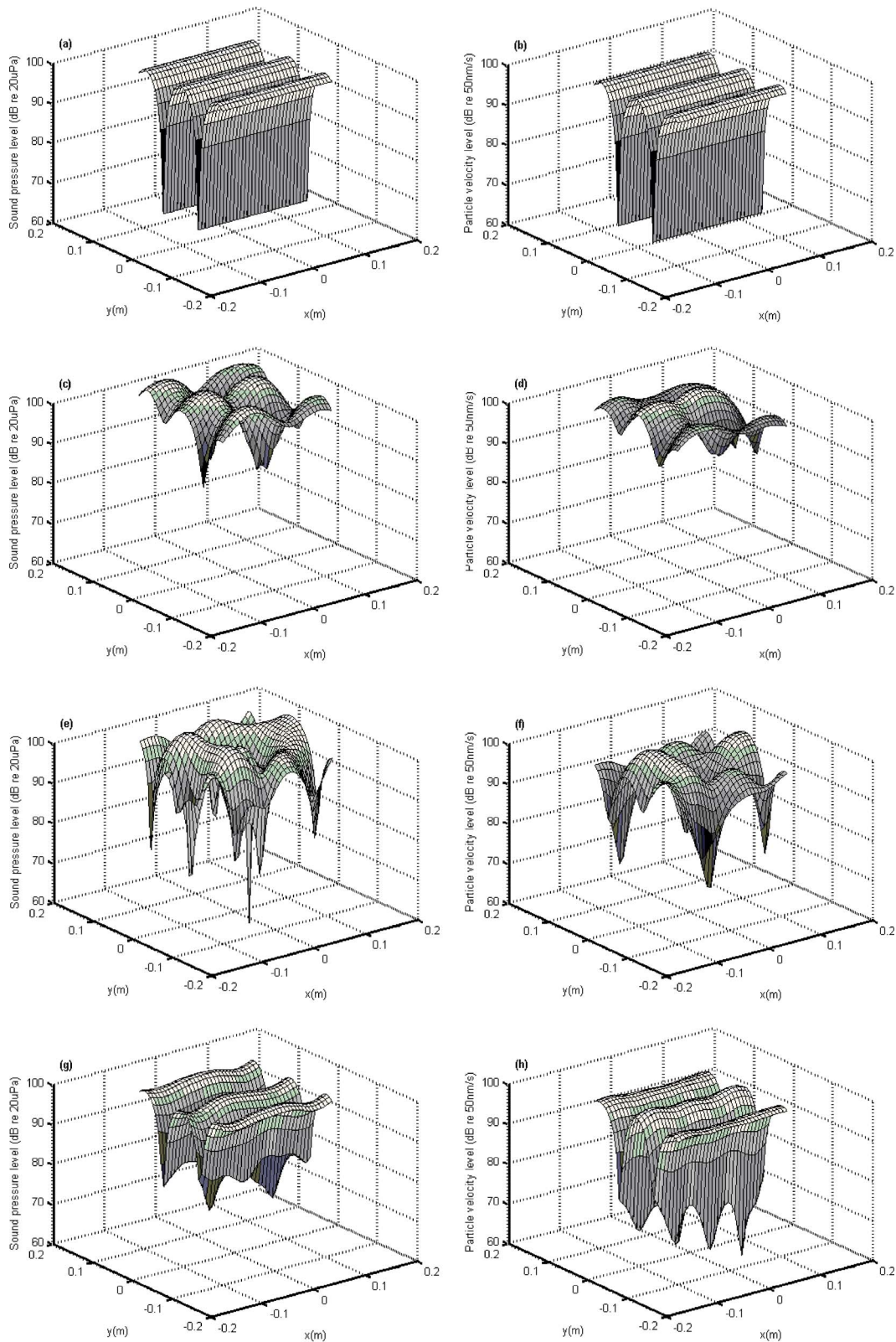


FIG. 4. Two interfering plane waves disturbed by two monopoles on the wrong side of the measurement plane. Left column (a, c, e, g), sound pressure; right column (b, d, f, h), particle velocity; top row, true (undisturbed) sound field; second row, predictions from the sound pressure; third row, predictions from the particle velocity; bottom row, predictions with the p - u method.

larger; (ii) SONAH gives good predictions even when the wave number spectrum in the source plane is extremely far from “white”; (iii) in the absence of disturbing noise, pressure-to-pressure and particle velocity-to-particle velocity

predictions work better than predicting one quantity from the other; and (iv) the p - u method copes successfully with disturbing noise.

Figure 5 shows a typical example of the results of the

SONAH procedures based on Eqs. (6)–(11) when the sound field generated by the large baffled panel is disturbed by sound from a monopole on the other side of the measurement plane. The monopole, located at (0, 0, 16), generates a sound pressure level in the prediction plane that is comparable to the level generated by the primary source. It is apparent that p - u method works fairly well whereas predictions based either on the pressure or on the particle velocity are seriously affected by the disturbing sources. The same source configuration has been examined with a lower level of extraneous noise, with a higher level, and with a much higher level. Not surprisingly the results (not shown) demonstrate that the advantage of the p - u method vanishes if the disturbing signal is low enough. They also show that the method can, of course, only cope with disturbing noise within limits; when the level of the noise is more than 20 dB higher than the primary noise the method no longer works.

It should be mentioned that the image source resulting from reflections in the source plane has been ignored in determining the results shown in Fig. 5. Adding such a source increases the error of the p - u method. Unfortunately there is no way of eliminating the sound field produced by image sources caused by reflections in the primary source since they are placed on the “right” side of the measurement plane; this is simply a fundamental limitation of any p - u or double layer technique.

IV. EXPERIMENTAL RESULTS

Some experiments have been carried out in DTU’s large anechoic room. Three different sources were used, a “monopole” (a Brüel & Kjær (B&K) 4299 “Volume Velocity Source,” which is a tube with a built-in sound intensity probe connected to a B&K 4295 “OmniSource” loudspeaker¹⁴), a “monopole on a sphere” (a device consisting of a hollow rigid sphere with a small hole driven by an internal loudspeaker, developed for calibrating p - u sound intensity probes¹⁵), and a model of a helicopter gearbox driven by an internal B&K 4809 electrodynamic exciter. The monopole on a sphere and the gearbox are shown in Figs. 6(a) and 6(b). The sound pressure and the particle velocity were measured at 6×6 points in two planes of 25×25 cm using a single $\frac{1}{2}$ in. p - u sound intensity probe produced by Microflown.⁶ The transducer, which was calibrated as described in Ref. 15, was moved manually over the two measurement planes using a “manual robot,” shown in Fig. 6(b). The two measurement planes were 6 and 3 cm from the nearest part of the source (all of which had a curved surface); the measurement plane near the source served the purpose of providing the true sound pressure and particle velocity in the prediction plane. A B&K “PULSE” analyzer in 1/12 octave mode was used for measuring the frequency responses between each of the two output signals from the p - u probe and the signal driving the source under test.

Figure 7 shows an example of results obtained with the gearbox. In this case there was no noise from the “wrong” side of the measurement plane, and therefore there is no

advantage in the p - u method; accordingly pressure-to-pressure and velocity-to-velocity predictions are slightly better than the p - u method.

Figure 8 shows results determined with the monopole as the primary source and the monopole on a sphere as disturbing source on the other side of the measurement plane. [The monopole was located at (-5.5, 9, 0), and the monopole on a sphere was located with the hole at (3.5, 1.5, 23) and the sphere behind.] Figure 8(a) shows the sound pressure levels generated by the two sources in the prediction plane, Fig. 8(b) shows the global relative error of the sound pressure, Fig. 8(c) shows the global relative error of the particle velocity, and Fig. 8(d) shows the global relative error of the sound intensity. (The global relative error of, say, the sound pressure is defined as the ratio of the sum of the absolute values of all local deviations between the predicted and the true pressure to the sum of the absolute values of the local true pressure.⁵) The sound intensity has been determined in four different ways, from the pressure, from the particle ve-

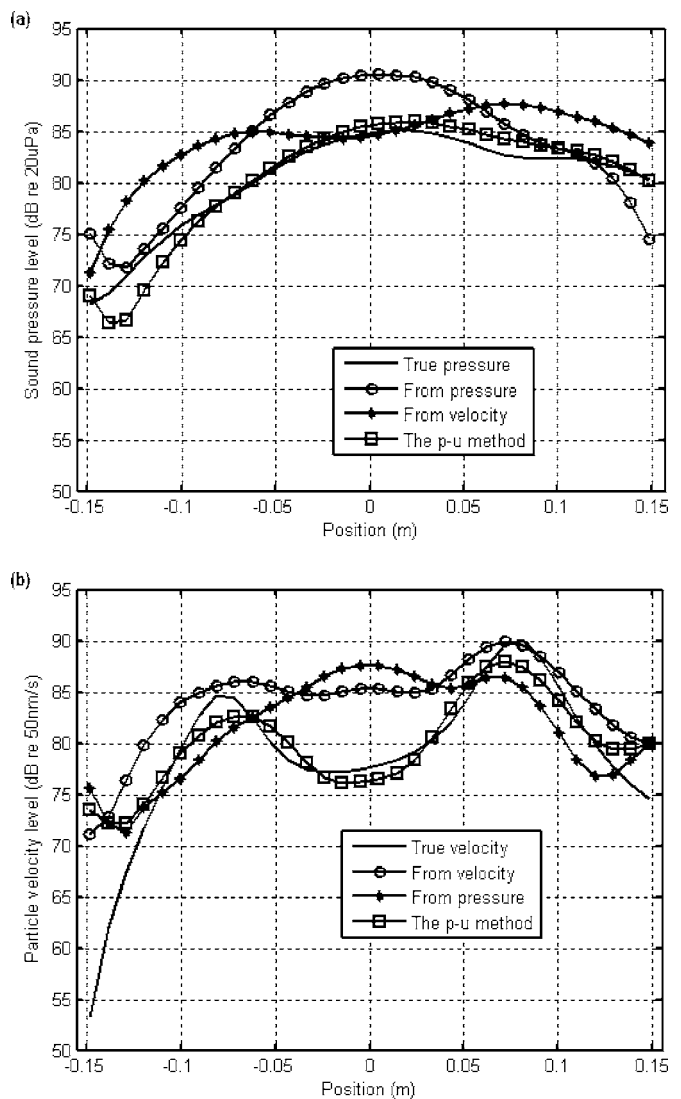
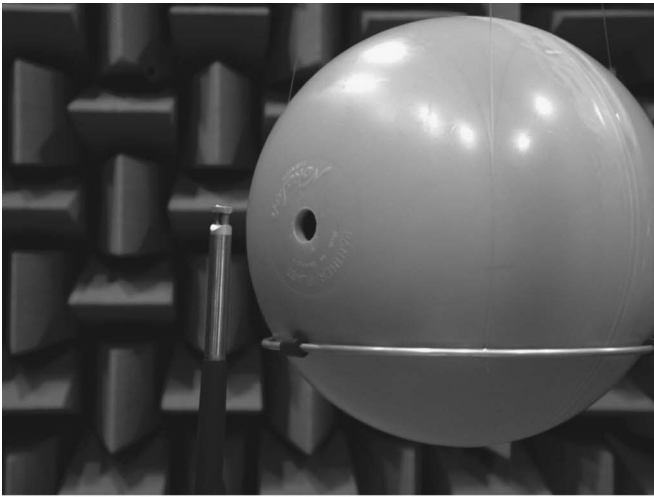
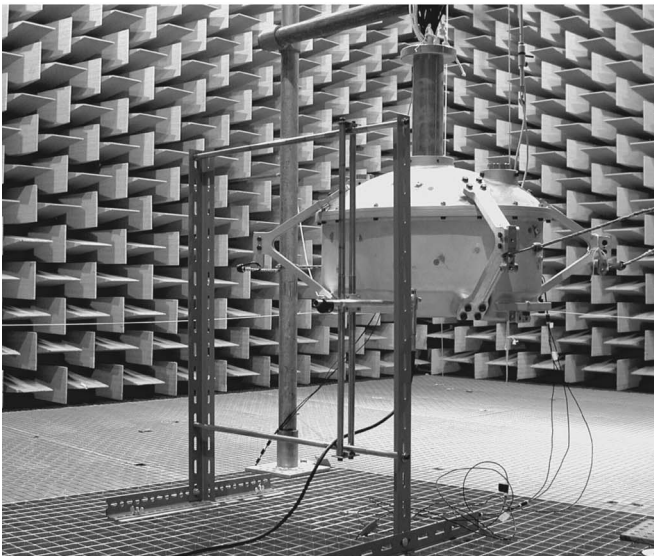


FIG. 5. Large panel driven at 1 kHz and disturbed by a monopole on the wrong side of the measurement plane. True (undisturbed) and predicted sound pressure (a) and true (undisturbed) and predicted particle velocity (b) in a diagonal across the prediction plane.



(a)



(b)

FIG. 6. Two sources in DTU's anechoic room, (a) the "monopole on a sphere" with the Microflow p - u intensity probe, and (b) the gearbox model driven by an internal exciter and the "manual robot" with the p - u probe.

locity, from the pressure determined from the pressure combined with the particle velocity determined from the particle velocity, and from pressure and particle velocity where both quantities have been determined from both quantities using the p - u method. Similar measurements (not shown) have been carried out with a lower level of the disturbing noise, and with a higher level. As expected the p - u method performs better than any other method when the signal-to-noise ratio is poor; otherwise the best solution is to predict the pressure from the pressure and the particle velocity from the particle velocity. However, if the disturbing noise level is much higher than the level generated by the source under test the p - u method breaks down.

V. DISCUSSION

A recent investigation of conventional NAH concluded that the accuracy of pressure-to-pressure predictions is comparable to the accuracy of velocity-to-velocity predictions,

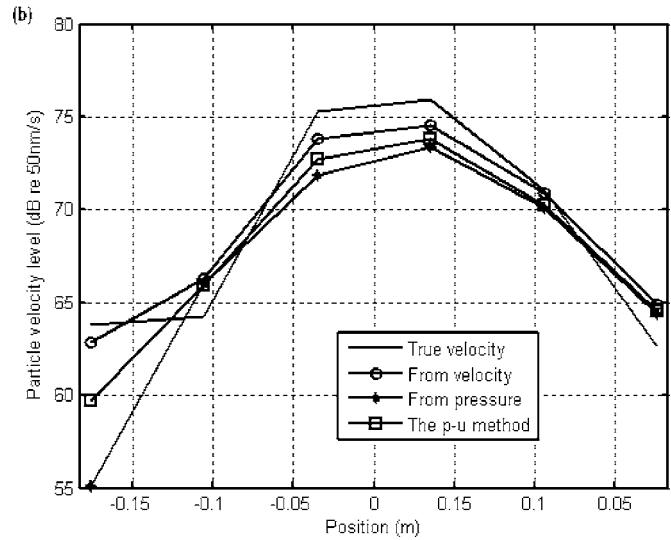
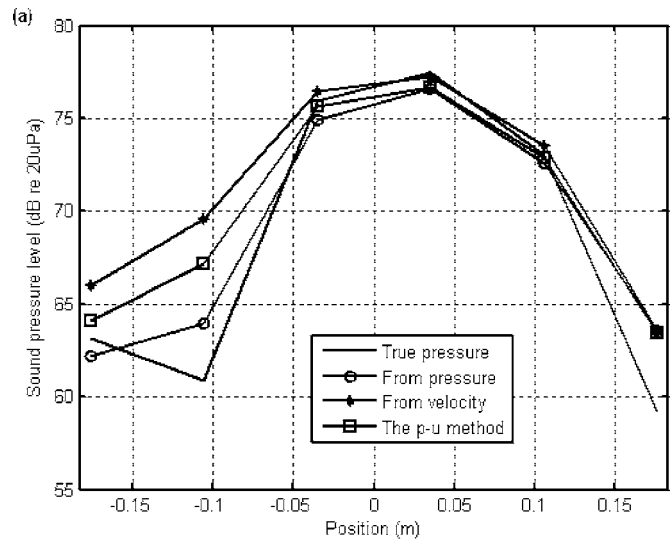


FIG. 7. Gearbox driven at 1 kHz. (a) True and predicted sound pressure, and (b) true and predicted particle velocity in a diagonal across the prediction plane.

whereas the accuracy of velocity-to-pressure predictions is far better than the accuracy of pressure-to-velocity predictions.⁷ It was also concluded that transducer mismatch has a significantly more serious influence on pressure-to-velocity predictions than on velocity-to-pressure predictions. In both cases the explanation was that the wave number ratio that occurs in the propagator when the pressure is predicted from the particle velocity in an inverse problem reduces high spatial frequencies associated with evanescent modes, whereas the reciprocal wave number ratio, which occurs when the particle velocity is predicted from the pressure, amplifies them.⁷ The tendencies observed with SONAH predictions are similar; on the whole velocity-to-pressure predictions perform better than pressure-to-velocity predictions, in particular if the transducers are not perfectly matched. The explanation is undoubtedly the same as with NAH; although transforms to the wave number domain are avoided, high spatial frequencies are nevertheless reduced in the former case and amplified in the latter case. An additional advantage

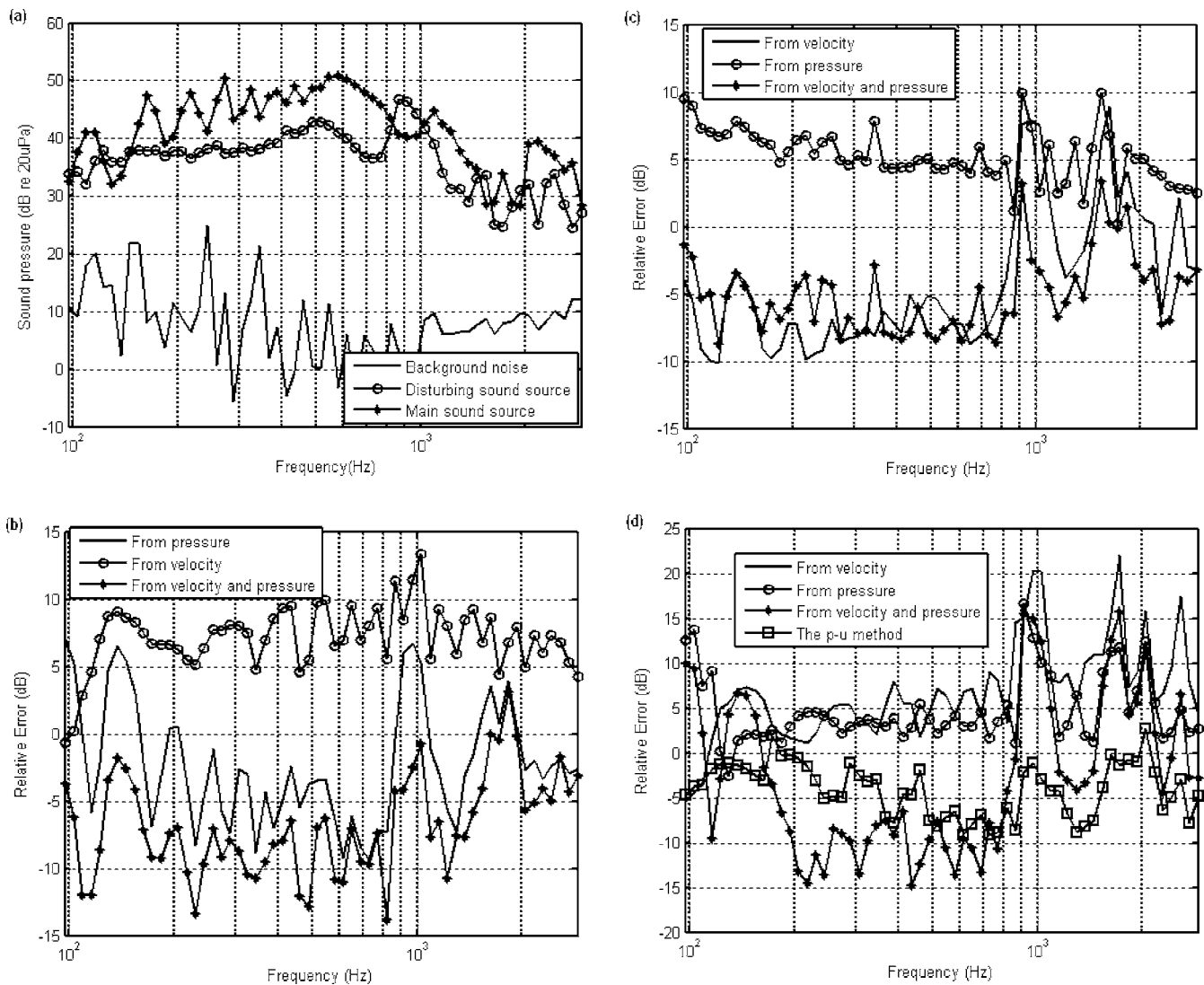


FIG. 8. Results obtained with the monopole disturbed by the monopole on a sphere. (a) Sound pressure level generated by the two sources at the center of the prediction plane; (b) relative error of predicted sound pressure, (c) particle velocity, and (d) sound intensity.

of particle velocity-based conventional NAH is that the normal component of the particle velocity because of the necessary, large measurement plane decreases faster toward the edges of the plane than the pressure does, which reduces spatial windowing effects.⁷ This effect is not relevant with the SONAH procedure.

VI. CONCLUSIONS

Statistically optimized near field acoustic holography has usually been based on measurement of the sound pressure in a plane, but may as well be based on measurement of the normal component of the particle velocity. A simulation study has shown and experimental results have confirmed that SONAH with advantage can be based on measurement of both quantities, in particular if the transducers are less than perfectly matched, since it in general is better to predict the sound pressure from the sound pressure and the particle velocity from the particle velocity than to predict either

quantity from the other quantity. Measuring both sound pressure and particle velocity has the additional advantage that it makes it possible—within limits—to reduce the influence of sound coming from the “wrong” side of the measurement array. In this case both the pressure and the particle velocity should be determined from both quantities.

ACKNOWLEDGMENT

The authors would like to thank Microflow for lending us a *p-u* sound intensity probe.

APPENDIX

The elements of the auto- and cross-correlation matrices $\mathbf{A}^H\mathbf{A}$, $\mathbf{A}^H\boldsymbol{\alpha}$, and $\mathbf{A}^H\boldsymbol{\beta}$ can be determined numerically as described in Ref. 4. This Appendix closely follows the derivation in Ref. 4 and describes how to calculate the elements of $\mathbf{A}^H\boldsymbol{\gamma}$.

In the limit of $M \rightarrow \infty$,

$$[\mathbf{A}^H \boldsymbol{\gamma}(\mathbf{r})]_n = \sum_m^M \Phi_m^*(\mathbf{r}_{h,n}) \rho c \frac{k}{k_z} \Phi_m(\mathbf{r})$$

$$\rightarrow \frac{\rho c}{\pi k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{j(k_x x_{h,n} + k_y y_{h,n} + k_z^* z_h)} e^{-j(k_x x + k_y y + k_z z)}}{k_z} dk_x dk_y,$$
(A1)

Changing to polar coordinates $[(k_x, k_y) = (K \cos \phi, K \sin \phi)]$ and introducing the function

$$w(K) = \frac{e^{j(k_z^* z_h - k_z z)}}{k_z} = \begin{cases} \frac{e^{-j\sqrt{k^2 - K^2}(z - z_h)}}{\sqrt{k^2 - K^2}} & \text{if } K^2 \leq k^2 \\ \frac{e^{-\sqrt{K^2 - k^2}(z + z_h)}}{-j\sqrt{K^2 - k^2}} & \text{if } K^2 > k^2 \end{cases}$$
(A2)

[cf. Eq. (3)], gives

$$[\mathbf{A}^H \boldsymbol{\gamma}(\mathbf{r})]_n \rightarrow \frac{\rho c}{\pi k} \int_0^{\infty} \int_0^{2\pi} w(K) \times e^{-jK((x-x_{h,n})\cos\phi + (y-y_{h,n})\sin\phi)} K d\phi dK$$

$$= \frac{\rho c}{\pi k} \int_0^{\infty} \int_0^{2\pi} w(K) \cos(KR_n \cos\phi) K d\phi dK$$

$$= \frac{2\rho c}{k} \int_0^k \frac{e^{-j\sqrt{k^2 - K^2}(z - z_h)}}{\sqrt{k^2 - K^2}} J_0(KR_n) K dK$$

$$+ \frac{j2\rho c}{k} \int_k^{\infty} \frac{e^{-\sqrt{K^2 - k^2}(z + z_h)}}{\sqrt{K^2 - k^2}} J_0(KR_n) K dK$$

$$= 2\rho c \int_0^{\pi/2} e^{-jk \cos\theta(z - z_h)} J_0(kR_n \sin\theta) \sin\theta d\theta$$

$$+ \frac{j2\rho c}{k(z + z_h)} \int_0^{\infty} J_0\left(kR_n \sqrt{1 + \frac{t^2}{k^2(z + z_h)^2}}\right) e^{-t} dt,$$
(A3)

where

$$K = k \sin \theta,$$
(A4)

$$R_n = \sqrt{(x_{h,n} - x)^2 + (y_{h,n} - y)^2},$$
(A5)

and

$$t = \sqrt{K^2 - k^2}(z + z_h).$$
(A6)

The two integrals of Eq. (A3) are finally evaluated using Gauss and Gauss-Laguerre quadrature, respectively.

¹J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography. I. Theory of generalized holography and the development of NAH," *J. Acoust. Soc. Am.* **78**, 1395–1413 (1985).

²E. G. Williams, *Fourier Acoustics—Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, 1999).

³R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," *Int. J. Acoust. Vib.* **6**, 83–89 (2001).

⁴J. Hald, "Patch near-field acoustical holography using a new statistically optimal method," *Proceedings of Inter-Noise 2003*, Jeju Island, Korea, pp. 2203–2210.

⁵Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized nearfield acoustical holography in cylindrical coordinates," *J. Acoust. Soc. Am.* **118**, 2355–2364 (2005).

⁶R. Raangs, W. F. Druyvesteyn, and H.-E. de Bree, "A low-cost intensity probe," *J. Audio Eng. Soc.* **51**, 344–357 (2003).

⁷F. Jacobsen and Y. Liu, "Near field acoustic holography with particle velocity transducers," *J. Acoust. Soc. Am.* **118**, 3139–3144 (2005).

⁸M. Tamura, "Spatial Fourier transform method of measuring transmission coefficients at oblique incidence. I: Theory and numerical examples," *J. Acoust. Soc. Am.* **88**, 2259–2264 (1990).

⁹F. Yu, J. Chen, C. Bi, W. Li, and X. Chen, "Experimental investigation of sound field separation technique with double holographic planes," *Chin. J. Acoust.* **24**, 111–118 (2005).

¹⁰J. Hald, "Patch holography in cabin environments using a two-layer handheld array with an extended SONAH algorithm," *Proceedings of Euronoise 2006*, Tampere, Finland.

¹¹J. Gomes, "Comparing parameter choice methods for the regularization in the SONAH algorithm," *Proceedings of Euronoise 2006*, Tampere, Finland.

¹²P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (SIAM, Philadelphia, 1997).

¹³F. Jacobsen and V. Jaud, "Statistically optimized near field acoustic holography based on particle velocity measurements," *Proceedings of the 13th International Congress on Sound and Vibration*, Vienna, Austria, 2006.

¹⁴S. Gade, N. Møller, J. Hald, and J. Alkestrup, "The use of volume velocity source in transfer measurements," *Proceedings of Inter-Noise 2004*, Prague, Czech Republic.

¹⁵F. Jacobsen and V. Jaud, "A note on the calibration of pressure-velocity sound intensity probes," *J. Acoust. Soc. Am.* **120**, 830–837 (2006).

Nearfield binaural synthesis and ambisonics

Dylan Menzies^{a)} and Marwan Al-Akaidi

Department of Computer Sciences and Engineering, De Montfort University, Leicester, Leicestershire LE1 6RS United Kingdom

(Received 20 June 2006; revised 22 December 2006; accepted 23 December 2006)

Ambisonic encodings can be rendered binaurally, as well as for speaker arrays. This process is developed for general high-order Ambisonic encodings of soundfields containing near as well as far sources. For sufficiently near sources an error is identified, resulting from the limited field of validity of the freefield harmonic expansion. A modified expansion is derived that can render such sources correctly. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434761]

PACS number(s): 43.60.Sx, 43.66.Pn, 43.66.Qp [EJS]

Pages: 1559–1563

I. INTRODUCTION

A. Ambisonic encoding

Ambisonics is a methodology developed for encoding soundfields, and decoding them onto speaker arrays, (Gerson, 1985, 1992). Initially it was used only to first order, with four signals that encode a full sphere of sound around a central listener. More recently Ambisonics has been employed at higher orders, whereby it is possible to not only increase the angular resolution of distant sources, but also extend the listening region and recreate accurately the soundfield from nearfield sources, (Daniel, 2003). We shall refer informally to an encoding of any order as *B format*, borrowing the original terminology for first order. Using high-order encodings, the listener receives distance cues about near sources exactly as they would for the real soundfield, because the soundfield around the listener can be reconstructed arbitrarily well.

The ambisonic encoded signals are defined by a spherical harmonic expansion of the freefield. Although our discussion does not depend on a particular representation, for definiteness we use signals, $B_{mn}^\sigma(k)$, defined with the real-valued $N3D$ spherical harmonics, (Daniel, 2000),

$$\Psi(\mathbf{r}, k) = \sum_m i^m j_m(kr) \sum_{n,\sigma} Y_{mn}^\sigma(\theta, \delta) B_{mn}^\sigma(k), \quad (1)$$

where

$$Y_{mn}^\sigma(\theta, \delta) = \sqrt{2m+1} \tilde{P}_{mn}(\sin \delta) \times \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \end{cases} \quad (2)$$

$$\tilde{P}_{mn}(\sin \delta) = \sqrt{(2 - \delta_{0,n}) \frac{(m-n)!}{(m+n)!}} P_{mn}(\sin \delta). \quad (3)$$

For $n=0$, σ only takes the value $+1$. θ here measures the angle around the coordinate symmetry axis, and $\pi/2 - \delta$ is the angle between the axis and the coordinate direction, so that δ would normally be called the elevation.

From here on we shall use a slightly simplified notation that removes the need for σ by extending n to negative values as used in the standard complex set

$$Y_{mn} = \begin{cases} Y_{mn}^{+1} & \text{if } n \geq 0 \\ Y_{m|n|}^{-1} & \text{if } n < 0 \end{cases}. \quad (4)$$

Similarly the encoded signals become $B_{mn}(k)$.

B. Conversion to binaural

In a binaural rendering system the listener is presented with one signal to each ear canal direct. Binaural signals can be derived from an ambisonic encoding, using head related transfer functions (HRTFs), as described below. The ambisonic encoding is easily rotated, which facilitates compensation for head movement.

Conversion to binaural can be achieved approximately by summing speaker array feeds that are each filtered by a HRTF matching the speaker position, (Jot and Wardle, 1998; McKeag and McGrath, 1996). Figure 1 illustrates the signal flow in this process.

A natural extension of this idea to an exact method for binaural signals is to transform the encoded soundfield into a plane wave expansion, and weight each component plane wave by the plane wave HRTF matching its direction and frequency, (Duraiswami *et al.*, 2005; Menzies, 2002). The process can be applied to high-order ambisonic encodings containing sources at various distances. A straightforward binaural approach would require HRTF sets for each source distance, however decoding the high-order signal requires only the plane wave HRTF set. This is not too surprising, as the HRTF sets are defined within the constraints of the wave equation, and so are all related. There is another less obvious advantage, which is that complex sources can be conveniently converted to binaural via high-order B format, as will be demonstrated in a future article. A single nearfield HRTF set cannot be applied in a simple way to a complex source description to yield the required binaural signals. Figure 2 depicts an overview of the encoding process from soundfield to binaural using a plane wave expansion.

The process is exact in the farfield, but as explained later, there is a subtle source of error which can affect near

^{a)}Author to whom correspondence should be addressed. Electronic mail: dylan@dmu.ac.uk

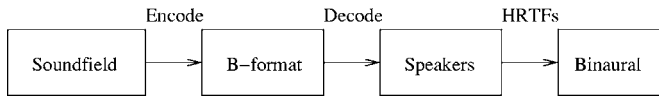


FIG. 1. Encoding a soundfield to binaural via virtual speakers.

sources. First we detail the steps to generate binaural signals from high-order B format using a plane wave expansion.

We aim to represent a source-free region by an expansion in plane waves, known as a Herglotz expansion, with coefficients $\mu(s, k)$ defined over unit vectors s , so that

$$\Psi(\mathbf{r}, k) = \frac{1}{4\pi} \int_{S_u} dS(s) e^{iks \cdot \mathbf{r}} \mu(s, k), \quad (5)$$

where integration is over the unit sphere. The spherical harmonic expansion of the source-free region, using standard complex spherical harmonics Y_m^n corresponding to N3D harmonics Y_{mn} , is

$$\Psi(\mathbf{r}, k) = \sum_m j_m(kr) \sum_n Y_m^n(\theta, \delta) A_m^n(k). \quad (6)$$

The terms $j_m(kr)$ are the spherical Bessel functions. From Eqs. (6) and (5) valid plane wave coefficients can be found in terms of the spherical harmonic coefficients A_m^n (Duraiswami *et al.*, 2005)

$$\mu(s, k) = \sum_{m,n} i^{-n} A_m^n(k) Y_m^n(s), \quad (7)$$

and in terms of the N3D convention

$$\mu(s, k) = \sum_{m,n} B_{mn}(k) Y_{mn}(s). \quad (8)$$

The lack of a complex factor in Eq. (8) reflects the fact that in ambisonics, plane waves with zero phase at the center have real-valued encodings, allowing the identification to be made between microphone polar patterns and the N3D harmonics. From the linear supposition of plane waves, the binaural signals, $\Psi^L(k)$, $\Psi^R(k)$ are found by integrating the plane wave weights with HRTF responses, $H^L(ks)$, $H^R(ks)$, over the sphere

$$\Psi^L(k) = \int_{S_u} dS(s) \mu(s, k) H^L(ks), \quad (9)$$

for the left side and similarly for the right. In practice the integral can be replaced by a quadrature sum, with very little loss of accuracy for a sufficient number of quadrature points, of order the number of spherical harmonics in Eq. (8) (Duraiswami *et al.*, 2005).

A high-order soundfield encoding can contain a variety of sources, both farfield and nearfield, so using the above method we are able to realize near sources binaurally with only knowledge of the the farfield HRTFs.

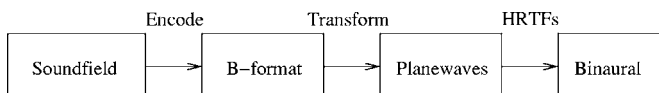


FIG. 2. Encoding a soundfield to binaural via a plane wave expansion.

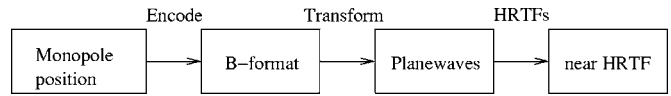


FIG. 3. Finding a nearfield HRTF from a monopole encoding.

Furthermore, by binaurally synthesizing a pure nearfield monopole of wave number k , the nearfield HRTF at that k , corresponding to the position of the monopole, is given immediately by $\Psi^L(k)$ in Eq. (9). Figure 3 summarizes this. The plane wave expansions of monopoles are investigated further in the next section.

Binaural rendering has the potential for greater realism than speaker array reproduction, because there is no constraint on the binaural signals delivered to the listener, whereas arrays must operate within a physical space and with the number of channels available, often in limited positions. On the other hand, binaural rendering has been hampered by the practical difficulties of ascertaining individual HRTFs and providing head tracking, both of which must be executed with precision. These problems are being addressed, and there are promising signs that they will be overcome. Therefore it is worthwhile to consider the process of binaural synthesis more carefully.

II. SCATTERING OF NEARFIELD SOURCES

A. Spherical expansion of a monopole

To study expansions of near sources we look in detail at the monopole. The important features are also true for general sources. A monopole source at noncentral position \mathbf{r}' has the following expansion in \mathbf{r} , valid only for $r < r'$, Morse and Ingard (1968)

$$\frac{e^{-ik|r-r'|}}{|r-r'|} = ik \sum_{m=0}^{\infty} j_m(kr) h_m(kr') \sum_{n=-m}^m Y_{mn}(\theta', \delta') Y_{mn}(\theta, \delta), \quad (10)$$

where $j_m(kr')$ and $h_m(kr)$ are the spherical Bessel function of the first kind and spherical Hankel function of the second kind. The positive frequency convention is chosen which gives an outward moving wave for a time piece $e^{i\omega t}$. For the following discussion it is important to emphasize the region of validity for the expansion is a sphere extending as far as the source, as illustrated in Fig. 4. A slightly expanded restriction applies to more general sources that may extend over a region. In this case the valid region extends as far as the maximum radius that does not enclose any source.

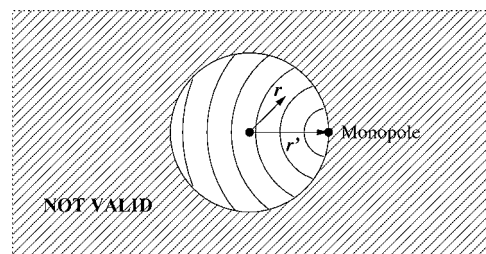


FIG. 4. Spherical expansion of a displaced monopole.

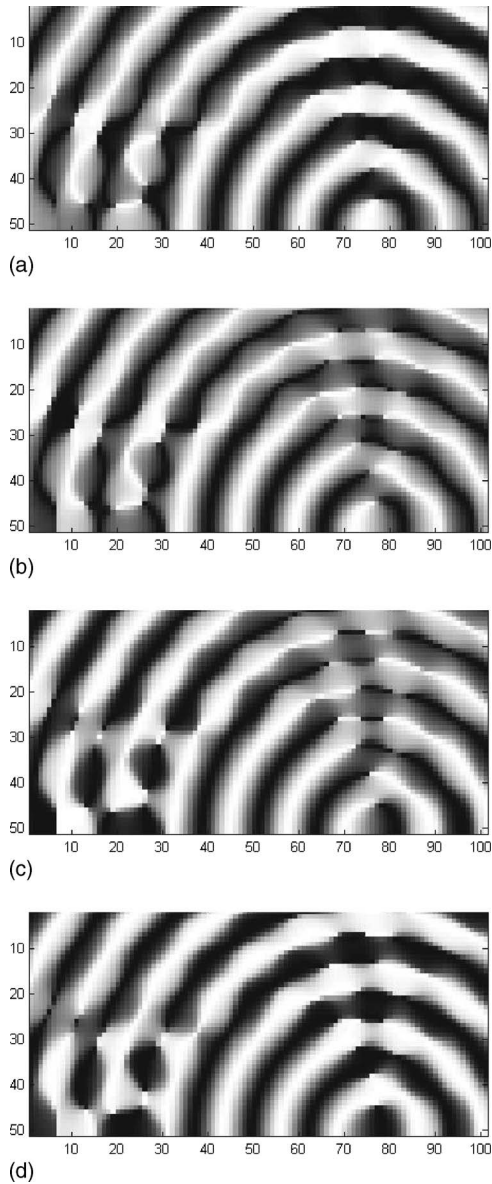


FIG. 5. Animation sequence, with $r'=2\lambda$, $m_{\max}=12$, $\omega t=0, \pi/4, 2\pi/4, 3\pi/4$.

A freefield expansion cannot be extended past the source completely because it must retain zero divergence everywhere. Figure 5 shows snapshots¹ taken at 1/4 cycle intervals of a half-plane cross section of an expansion with $r'=2\lambda$ and the maximum value of m , $m_{\max}=12$. The bessel functions $j_m(kr)$ are very close to zero for $kr < m$, so $m_{\max} \approx kr'$ is sufficient to synthesize accurately in the region $r < r'$. A detailed error analysis has been performed, (Duraiswami *et al.*, 2005). To aid visualization the plots are normalized $\text{Re}(\Psi)/|\Psi|$. The monopole radiates outwards from the source in the valid region, while just outside the valid region, the field radiates inwards, satisfying zero divergence.

B. Plane wave decomposition

As described by Eq. (8), the spherical expansion can be re-expressed as a plane wave expansion. Figure 6 shows the expansion in Fig. 5 re-expressed with 196 plane waves, whose directions are distributed around the sphere on Fliege

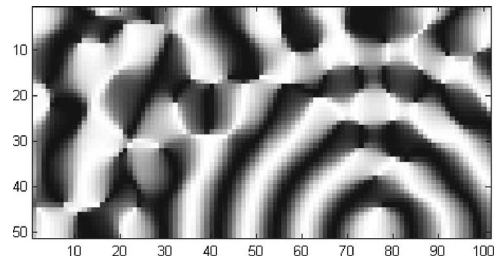


FIG. 6. Plane wave-from-spherical expansion of a monopole with $r'=2\lambda$, $m_{\max}=12$, 196 Fliege nodes.

nodes, (Fliege and Maier, 1999). Fliege nodes are positioned for minimal quadrature error over the sphere. The valid region matches closely, the outer region has transformed and remains invalid.

It appears at first that the angular bandwidth of the node set is sufficient if it matches that which is sufficient for a spherical harmonic expansion to radius r , $m_{\max} \approx kr$. This assumption was made in Duraiswami *et al.*, 2005. However, closer examination shows that a higher bandwidth for the node set, $m_{\max} \approx 2kr$, provides a nearer match.

C. Scattering validity of synthesized sources

A HRTF filter generates the signal at an ear resulting from the scattering of a plane wave by the listener. The phase of the plane wave is assumed to be zeroed to the center of the head. We can express the resultant field as a sum of the original unscattered field and the scattered component, $\Psi = \Psi_{\text{in}} + \Psi_{\text{scat}}$. The scattering can be formulated in terms of the Sommerfield radiation conditions, which state that Ψ_{scat} depends only on Ψ_{in} at the boundary of the scattering body. As we have seen, the ambisonic representation of a nearfield source is accurate only within a limited region, no matter how high the order of approximation. The derived plane wave decomposition can also only be accurate within the limited region. If part of the scattering body is outside this region, then Ψ_{in} is no longer correct on all of the scattering body. The binaural signals, found according to Eq. (9), are part of the resultant field Ψ , and so in general suffer loss of accuracy when the valid region does not enclose the scattering surface. Figure 7 illustrates this for a front view of a listener, where the source is close enough that everything from the shoulders downwards is excluded from the valid region. It has been shown previously that the torso plays a significant role in localization, (Algazi *et al.*, 2001).

The arrows in Fig. 7 show the flow of energy in the freefield expansion of the source, as illustrated previously in the animation sequence Fig. 5. In a field with a real source, both arrows would point away from the source. It is evident that the scattering in the shoulder region using the freefield will be quite different from the scattering with a source field, and cause differences in the resultant field at the ears.

D. Higher accuracy nearfield expansions

The above result is not too surprising in retrospect, because we should not expect to be able to construct the response from a source embedded in an arbitrarily complex

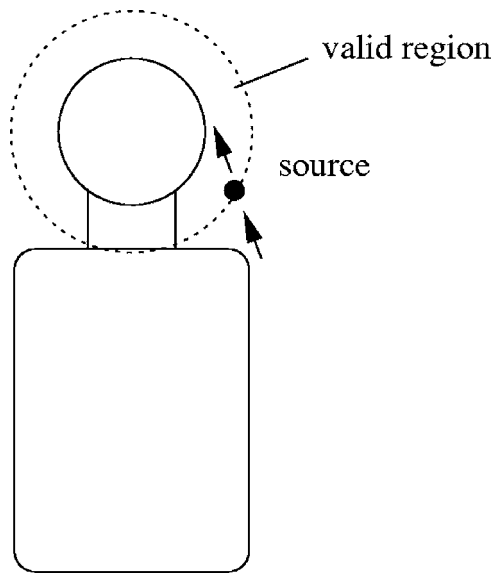


FIG. 7. Parts of scattering body outside the valid expansion region of a near source.

scattering geometry using only plane wave scattering responses. However, the question remains, how well can we do with plane wave HRTFs? Plane wave expansions have a useful property that spherical harmonic expansions lack, they can be translated and expressed about a different point simply by multiplying by phase factors: If a position relative to the new center is \mathbf{r}' , and the corresponding position relative to the old center is \mathbf{r} , then $\mathbf{r} = \mathbf{r}' + \mathbf{x}$, where \mathbf{x} is the translation from the old to the new center. So $e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot(\mathbf{x}+\mathbf{r}')} = e^{i\mathbf{k}\cdot\mathbf{x}} e^{i\mathbf{k}\cdot\mathbf{r}'}$. Therefore from Eq. (5) the expansion coefficients about the new center are $\mu'(\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{x}} \mu(\mathbf{k})$. Using this result, we can take the expansion for a source at a large radius, then shift the center so the source is at the required relative location. In the process the region of validity has been expanded, so a greater scattering body can be included, and the resulting binaural signals will be more accurate. Figure 8 illustrates this, showing a body entirely within the valid region for scattering, and with a source near to the head.

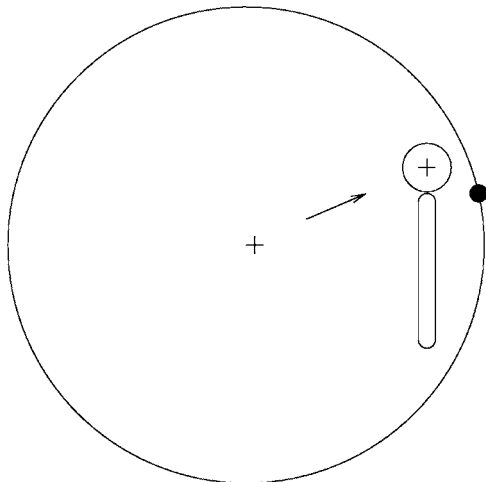


FIG. 8. Shifted plane wave expansion for full scattering with a near source.

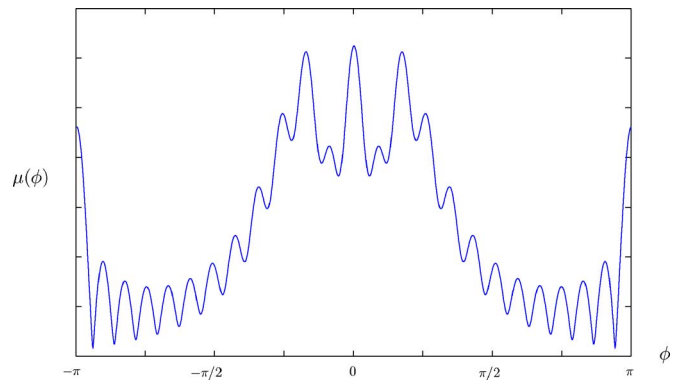


FIG. 9. Plane wave expansion coefficient for $r=2\lambda$, $m_{\max}=12$.

There is a cost to be paid for the improved rendering of the source, however. The new region of validity excludes other near source regions, which must be rendered separately with their own shifts, rather than from a single expansion. Also the new spherical expansion must be specified to higher order, because the region being scattered is at a greater radius in the freefield expansion. As a result of this the number of plane waves and hence HRTFs must be increased. The number of nodes required is $O(kr)$ where r is the radius of curvature of the region boundary.

The limit of the boundary of validity obtained by shifting is a plane through the source, so it would appear impossible, as conjectured earlier, to precisely generate binaural signals, using plane wave HRTFs, for sources in concave regions of the scattering body, such as under the chin. The fact that there exist a family of plane wave expansions about each point is a consequence of the presence of a source region, and the nonlocality of the plane wave. A source-free region on the whole plane, such as the harmonic expansion, has a unique plane wave expansion.

Figures 9 and 10 show two plots of the magnitude $|\mu(\phi)|$ of the plane wave decomposition of a freefield expansion of a monopole source at $r=2\lambda$, 4λ , respectively. The term ϕ is the spherical coordinate measuring the angle between the direction and the coordinate symmetry axis, with $\phi=0$ being the direction to the source. The decomposition of a source in a general position is just a rotation of $\mu(\phi)$. The greater detail seen in Fig. 10, compared with Fig. 9, reflects that it is sensitive to higher resolution HRTF data.

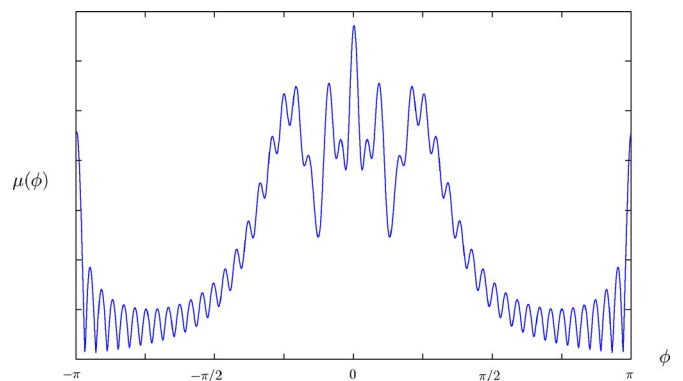


FIG. 10. Plane wave expansion coefficient for $r=4\lambda$, $m_{\max}=24$.

E. Complex sources

The argument developed above extends to complex localized sources, which can in general be written in terms of infinitesimal monopole clusters, or multipole expansions, and so subject to the same expansion region limitations. The valid region of a freefield expansion extends to the maximum radius that is free of any enclosed source.

The characterization of a general source by multipole expansion is a useful form of encoding. It can be transformed into a freefield expansion at the listener according to position and orientation, then into a plane wave expansion, and finally binaural signals. The details shall appear in a future publication.

F. Optimizing the expansion order

We go back now and consider the less special case of encoding sources that are not so close that the enlarge and shift process just described is necessary to include relevant scattering. If the radius of the scattering region, centered on the head center, is r , then the order of expansion required is $m_{\max}=kr$. There is nothing to prevent us from moving the center of the expansion downwards to the center of the total scattering body, in order to minimize r , and therefore m_{\max} . The plane wave expansion must be phase shifted as before, prior to HRTF processing. A further refinement would be to separate the encoded signal into several banded signals. The lower frequencies would be diffracted by the body and require a larger scattering region, although lower sampling rates, while higher frequencies would require a smaller region at full rate.

G. Speaker arrays

In principle the limitation of ambisonic encodings applied to HRTFs is also relevant to array reproduction, except in this case the listener's body will scatter in real nonvalid regions, rather than via the HRTF filtering process. In practice near sources could only be produced for one listener, negating the main advantage of array systems, and even this would be difficult to achieve due to other limitations mentioned earlier. It is conceivable, however, that a specialist auditory display might be designed with a speaker array around a single listener, and then the consideration of scattering validity is relevant.

III. CONCLUSION

Ambisonic encoding provides a convenient soundfield representation, that can be rendered into binaural signals via plane wave decompositions in a precise way. However, there exists a surprising limitation for binaural rendering of near sources, namely errors are introduced due to incorrect accounting for scattering. It is possible to synthesize improved binaural signals from a given source by creating a more detailed plane wave expansion that allows for scattering objects in a larger region. To cover a general soundscape, we combine a main expansion covering most sources, with additional expansions for close source regions.

We have not presented any tests applying our methods with real HRTF data. This is clearly an important and complex task, which we hope to address. Working with high quality personalized HRTFs will be a key factor. With the ongoing refinement of virtual reality systems, the considerations presented here are expected to become increasingly relevant.

¹A movie can be found at www.cse.dmu.ac.uk/~dylan/NearfieldBinauralSynthesis/freefieldMonopole.avi
Last viewed 2/7/06.

- Algazi, V., Avendano, C., and Duda, R. O. (2001). "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.* **109**, 1110–1122.
- Daniel, J. (2000). "Representation de champs acoustiques, application la transmission et la reproduction de scenes sonores complexes dans un contexte multimedia," Ph.D. thesis, University of Paris 6, Paris, France.
- Daniel, J. (2003). "Spatial sound encoding including near field effect," in *Proceedings of the AES 23rd International Conference*.
- Duraiswami, R., Zotkin, D., Li, Z., Grassi, E., Gumerov, N., and Davis, L. (2005). "High order spatial audio capture and its binaural head-tracked playback over headphones with hrtf cues," in *Proceedings of the AES 119th Convention*, New York.
- Fliege, J., and Maier, U. (1999). *IMA J. Numer. Anal.* **19.2**, 317–334.
- Gerzon, M. (1985). *J. Audio Eng. Soc.* **33**, 859–871.
- Gerzon, M. (1992). "General metatheory of auditory localization," in *Proceedings of the 92nd AES Convention*, Vienna.
- Jot, J., and Wardle, S. (1998). "Approaches to binaural synthesis," in *Proceedings of the AES 105th Convention*, San Francisco.
- McKeag, A., and McGrath, D. (1996). "Sound field format to binaural decoder with head tracking," in *Preprint 4302, AES Convention 6r*, Melbourne.
- Menzies, D. (2002). "W-panning and o-format, tools for object spatialization," in *Proceedings of the AES 22nd International Conference*, Helsinki.
- Morse, P., and Ingard, K. (1968). *Theoretical Acoustics* (Princeton University Press).

Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions

Christopher A. Shera^{a)}

Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Department of Otolaryngology, Harvard Medical School, Boston, Massachusetts 02115

Arnold Tubis

Department of Physics, Purdue University, West Lafayette, Indiana 47907 and Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093

Carrick L. Talmadge

National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

Egbert de Boer

Room D2-226, Academic Medical Center, University of Amsterdam, Meibergdreef 9 1105 AZ, Amsterdam, The Netherlands

Paul F. Fahey

Department of Physics, University of Scranton, Scranton, Pennsylvania 18510

John J. Guinan, Jr.

Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Department of Otolaryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 22 June 2006; revised 15 November 2006; accepted 16 November 2006)

Originally proposed as a method for measuring the power gain of the cochlear amplifier, Allen–Fahey experiments compare intracochlear distortion products and ear-canal otoacoustic emissions (OAEs) under tightly controlled conditions. In this paper Allen–Fahey experiments are shown to place significant constraints on the dominant mode of reverse energy propagation within the cochlea. Existing Allen–Fahey experiments are reviewed and shown to contradict the predictions of compression-wave OAE models recently proposed in the literature. In compression-wave models, distortion products propagate from their site of generation to the stapes via longitudinal compression waves in the cochlear fluids (fast waves); in transverse traveling-wave models, by contrast, distortion products propagate primarily via pressure-difference waves whose velocity and other characteristics depend on the mechanical properties of the cochlear partition (slow waves). Compression-wave models predict that the distortion-product OAEs (DPOAEs) measured in the Allen–Fahey paradigm increase at close primary-frequency ratios (or remain constant in the hypothetical absence of tuned suppression). The behavior observed experimentally is just the opposite—a pronounced decrease in DPOAE amplitude at close ratios. Since neither compression-wave nor simple conceptual “hybrid-wave” models can account for the experimental results—whereas slow-wave models can, via systematic changes in distortion-source directionality arising from wave-interference effects—Allen–Fahey and related experiments provide compelling evidence against the predominance of compression-wave OAEs in mammalian cochlear mechanics. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2405891]

PACS number(s): 43.64.Bt, 43.64.Kc, 43.64.Jb, 43.64.Ri, 43.64.Tk [BLM] Pages: 1564–1575

I. INTRODUCTION

Otoacoustic emissions (OAEs) appear inextricably linked to the biophysical and hydromechanical mechanisms that serve to amplify the motion of the cochlear partition. Most OAE models, therefore, assume that active forces exerted within the partition couple primarily into the transverse

motion of the basilar membrane (BM), and thereby into the slow traveling pressure wave that drives it (e.g., Zweig and Shera, 1995; Talmadge *et al.*, 1998; Nobili *et al.*, 2003). Such “slow-wave” models usually neglect the possibility that the motions of structures within the partition (e.g., the outer hair cells) may produce significant acoustic compression waves that propagate through the cochlear fluids at the speed of sound, uninfluenced by mechanically tuned but almost acoustically transparent structures such as the organ of Corti. In principle, however, back propagation via longitudinal

^{a)}Author to whom correspondence should be addressed. Electronic mail: shera@epl.meei.harvard.edu

compression waves (or “fast waves”) provides an additional—and perhaps predominant—mechanism for energy to escape from the inner ear. Indeed, recent measurements of distortion products in the ear canal (DPOAEs) and on the BM have been taken to confirm the suggestion (Wilson, 1980b) that the reverse propagation of OAEs occurs primarily if not exclusively via fast compression waves (Ren, 2004; Ren *et al.*, 2006; Ruggero, 2004 Siegel *et al.*, 2005).¹

Since the laws of physics permit both fast- and slow-wave OAEs, the existence of either is not in doubt—as in T. H. White’s fictional world of ants (White, 1958), so likely in the cochlea: Everything not forbidden is compulsory. The problem is to determine the relative strengths of the two emission modes and to understand their respective roles in cochlear mechanics. Unfortunately, the distortion-product measurements so far adduced to support the compression-wave model cannot convincingly address the issue (e.g., Ren, 2004; Ruggero, 2004); the probative power of these experiments is limited because they do not distinguish between distortion arising locally and distortion propagating to the measurement point from an origin elsewhere in the cochlea (cf., Dong and Olson, 2006; Shera *et al.*, 2006). Here, we take an alternative approach, asking whether compression-wave models of DPOAEs can account, even qualitatively, for a particularly relevant and well established result in cochlear mechanics: The so-called “Allen–Fahey” experiment. In a nutshell, we find that they cannot.² With additional assumptions, our arguments extend to simple purpose-built “hybrid-wave” models that combine both fast- and slow-wave coupling. Since slow-wave models by themselves *can* explain the results, Allen–Fahey and related experiments provide compelling evidence against a predominant role for compression-wave OAEs in mammalian cochlear mechanics.

II. ALLEN–FAHEY AND RELATED EXPERIMENTS

Allen–Fahey experiments involve measuring intracochlear distortion products (DPs) and ear-canal otoacoustic emissions under tightly controlled conditions (Allen and Fahey, 1992). In particular, the paradigm consists of measuring the ear-canal DPOAE (typically at $f_{dp}=2f_1-f_2$) as a function of the primary frequency ratio $r\equiv f_2/f_1$ while the intracochlear DP response is held constant at its characteristic place, x_{dp} (the DP frequency f_{dp} remains fixed throughout the measurement). We denote the DPOAE pressure measured under these conditions—the so-called “Allen–Fahey function”—by $P_{DPOAE}^\dagger(r)$, where the superscripted dagger signifies the microelectrode originally employed to monitor the intracochlear distortion product.

A. Methods for “calibrating” the intracochlear DP

Three different methods for monitoring the intracochlear response to the DP have been employed to date. In Allen and Fahey’s original experiment (Fahey and Allen, 1985), later repeated by Shera and Guinan (1997; 2007a), the DP amplitude at x_{dp} was determined by measuring the firing rate of an auditory-nerve fiber (ANF) tuned to f_{dp} . Shera and Guinan also performed an entirely otoacoustic version of the Allen–

Fahey experiment in which the intracochlear response near x_{dp} was monitored noninvasively using “secondary” distortion products.³ By introducing a third, fixed tone one can create a secondary DP whose ear-canal amplitude and phase provide a sensitive assay of the intracochlear value of the original DP near its characteristic place. Most recently, de Boer and colleagues have done an Allen–Fahey experiment using laser-Doppler velocimetry to measure the f_{dp} component of BM velocity at x_{dp} (de Boer *et al.*, 2005).

Each method for monitoring the intracochlear DP has strengths and weaknesses. The method based on ANF firing rate can be performed at any location in the cochlea by recording from a fiber with the corresponding characteristic frequency (CF). However, the method cannot be used to measure $P_{DPOAE}^\dagger(r)$ at values of r close to 1; at close ratios the fiber is driven both by the DP and by one (or both) of the primaries. The otoacoustic method overcomes this limitation and can be applied both at close ratios and at any frequency. In addition, because the otoacoustic method is noninvasive it can be employed in a wider variety of species. However, because the secondary DPOAE is generated over a *region* straddling x_{dp} , rather than at x_{dp} alone, the otoacoustic method is less tightly controlled. The BM-velocity method provides the most direct measure of mechanical motion in the cochlea. Like the otoacoustic method, it can be performed at any ratio. The method’s major disadvantage is that it can be employed only at a rather limited number of surgically accessible locations within the cochlea.

B. Synopsis of the experimental results

Figure 1 shows that despite significant differences in methodology—including the use of two different species (cat and guinea pig) and three different assays for measuring the intracochlear response to the DP (ANF firing rate, secondary DPOAE amplitudes, and BM velocity)—all four studies that have employed the Allen–Fahey paradigm to date report qualitatively similar results for $P_{DPOAE}^\dagger(r)$. In particular, as r decreases towards 1 from its maximal value (typically, $r_{max}\cong 1.5$), the magnitude of $P_{DPOAE}^\dagger(r)$ first stays roughly constant (or slowly increases) but then decreases rapidly, falling significantly below its value at r_{max} as r draws closer to 1 ($r<1.2$). [The abscissa in Fig. 1 shows f_1 and indicates the constant value of f_{dp} with a short vertical line; since $r=2-f_{dp}/f_1$, decreasing f_1 corresponds to decreasing r .] Although we count the original experiments performed by Fahey and Allen among the four extant studies consistent with this trend,⁴ the relatively high otoacoustic noise floor in their experiments precludes any compelling statement about the general behavior of $P_{DPOAE}^\dagger(r)$ based solely on their data.⁵

Results from experiments performed using the Allen–Fahey paradigm are broadly consistent with related but less tightly controlled comparisons of ear-canal DPOAEs and intracochlear DPs. For example, whereas DPOAEs recorded in the ear canal decrease dramatically at ratios less than the so-called “optimal” frequency ratio (i.e., $r<r_{opt}\cong 1.2$), psychophysical observations suggest that the intracochlear response to the DP *increases* at close ratios (e.g., Goldstein, 1967; Wilson, 1980a; Zwicker and Harris, 1990). Similar

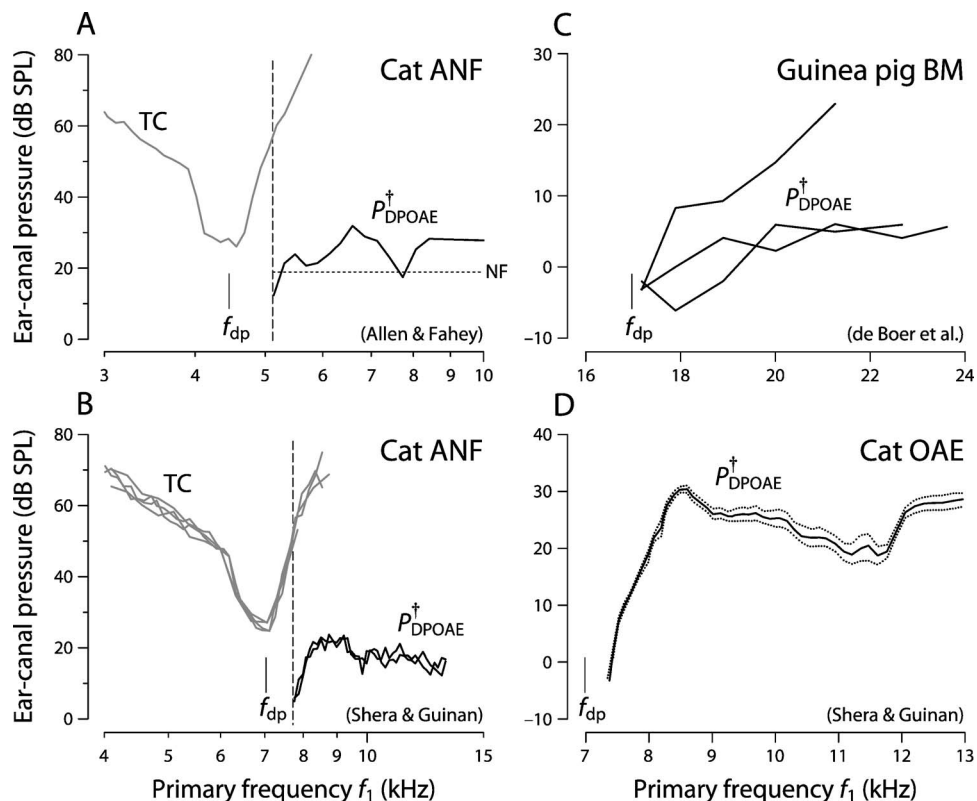


FIG. 1. Summary of results for Allen–Fahey experiments performed to date. In all panels solid black lines show measurements of the Allen–Fahey function $P_{\text{DPOAE}}^{\dagger}(r)$ vs the primary frequency f_1 . The DP frequency $f_{\text{dp}}=2f_1-f_2$ (short vertical line) is held constant during the measurement. Both f_1 and r increase along the abscissa. Labels indicate the species and method for monitoring the intracochlear response to the DP. (A) Sample result from the original Allen–Fahey experiment in cat (Allen and Fahey, 1992, Fig. 7). The DP response at x_{dp} was monitored using the auditory-nerve fiber (ANF) whose threshold tuning curve (TC) is shown in gray. The dashed vertical line indicates the frequency at which the f_1 primary started to excite the fiber so that the constant-DP condition could no longer be maintained. The dotted horizontal line (NF) represents the approximate noise floor for the DPOAE measurement. (B) Sample result from a repetition of the original ANF-based experiment in cat (Shera and Guinan, 2007a) featuring an increased frequency resolution and a lower acoustic noise floor (approximately -20 dB SPL). The multiple traces illustrate the reproducibility of the result. (C) Results from Allen–Fahey experiments in guinea pig in which laser-Doppler velocimetry was used to maintain constant BM velocity at x_{dp} (de Boer *et al.*, 2005). (D) Results from a noninvasive Allen–Fahey experiment in cat in which secondary distortion products were used to control the intracochlear response at x_{dp} (Shera and Guinan, 2007a). The data are from the same animal and CF range that provided the ANF data in (B). The dotted lines show the approximate uncertainty of the measurement based on ± 1 dB variations in the amplitude of the secondary DPOAE.

increases have been reported in measurements of BM motion, although the results are somewhat less conclusive (reviewed in Robles and Ruggero, 2001). Whereas some studies find a monotonic increase in the amplitude of the BM DP at close ratios (e.g., Robles *et al.*, 1997), others do not (e.g., Cooper and Rhode, 1997), perhaps because of suppressive effects.

C. Focus of the paper

The Allen–Fahey paradigm was developed to provide an indirect measure of the power gain of the cochlear amplifier (Allen and Fahey, 1992). Discussions of the paradigm—its experimental results, extensions, and theoretical basis—usually focus on the question: Does the Allen–Fahey paradigm yield a readily invertible measure of the amplifier gain? We do not specifically address that issue here (cf., Allen and Fahey, 1992; Kanis and de Boer, 1993; Shera, 2003; de Boer *et al.*, 2005; Shera and Guinan, 2007a); rather, we ask the more limited question: Irrespective of the amplifier gain or its relation to $P_{\text{DPOAE}}^{\dagger}(r)$, can compression-wave models of DPOAEs account for the qualitative results summarized in Fig. 1?

Sections III and IV show that the results of the Allen–Fahey experiment are incompatible with compression-wave DPOAE models. Readers who wish to understand the essential logic of the arguments without working through the details can skip ahead to the summary presented in the Sec. V A. Section V B reviews existing slow-wave explanations for Allen–Fahey experiments. Section V C rules out simple “hybrid-wave” models involving both fast- and slow-wave coupling. Regardless of whether Allen–Fahey experiments yield a readily interpretable measure of the power gain of the cochlear amplifier, they evidently do provide compelling evidence against the predominance of compression-wave DPOAEs in mammalian cochlear mechanics.

III. COMPRESSION-WAVE MODELS OF DPs AND DPOAEs

Figure 2 shows a schematic of the cochlea. For simplicity we consider a symmetric box model. $P_{\text{sv}}(x)$ and $P_{\text{st}}(x)$ represent the pressures in the two scalae just above and below the BM. The pressures also depend on the driving frequency, but we have left this dependence implicit.

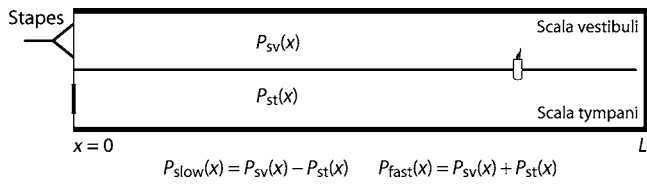


FIG. 2. Schematic of a symmetric box model of the cochlea. Without loss of generality, the pressures in the scala vestibuli and scala tympani— $P_{sv}(x)$ and $P_{st}(x)$, respectively—can be represented as linear combinations of sum and difference (fast- and slow-wave) components (Peterson and Bogert, 1950); only the slow-wave component drives the basilar membrane (BM). The “hair cell” straddling the BM represents a latent pressure source.

A. Fast waves and slow waves

As pointed out by Peterson and Bogert (1950), the pressures in the two scalae can be represented as linear combinations of sum and difference components with functional significance. The difference component of the pressure is the “slow wave,” defined by

$$P_{\text{slow}}(x) \equiv P_{sv}(x) - P_{st}(x). \quad (1)$$

The transverse slow wave drives the BM up and down and its propagation velocity and other characteristics depend on properties of the partition. The sum component is the compressional “fast wave,” defined by

$$P_{\text{fast}}(x) \equiv P_{sv}(x) + P_{st}(x). \quad (2)$$

True to its name, the longitudinal fast wave invades all soft tissues and propagates at the speed of sound in a manner independent of the BM. Although we use the names “fast” and “slow” for convenience, the two waves might more accurately be identified as longitudinal (compressional) and transverse (shearing) with respect to the spatial axis defined by position along the BM. Since arbitrary scalae pressures $P_{sv}(x)$ and $P_{st}(x)$ can be written as linear combinations of $P_{\text{slow}}(x)$ and $P_{\text{fast}}(x)$, we will henceforth work solely with the slow- and fast-wave components of the pressure.

B. Oscillating pressure sources

Forces generated by outer hair cells are thought to underlie the production and amplification of the reverse-traveling waves that give rise to DPOAEs. To represent these forces, the figure shows a quiescent pressure source (e.g., a hair cell) straddling the BM; for clarity, the figure shows only one of many such sources present in the cochlea. When activated, pressure sources generally couple into both the slow- and the fast-wave components of the pressure; the relative strength of the two coupling modes depends on the nature of the sources and their environment.

1. Slow-wave sources

Figure 3 illustrates the limiting case in which the vibrating source produces opposite pressures on the two sides of the partition. In this case $P_{sv}(x)$ and $P_{st}(x)$ sum to zero, and the source couples entirely into the slow, pressure-difference wave, $P_{\text{slow}}(x)$. The oscillating pressure difference acts on the BM and causes it to move up and down, generating waves (measurable both in the pressure and on the BM) that travel in both directions. At the base the slow wave vibrates

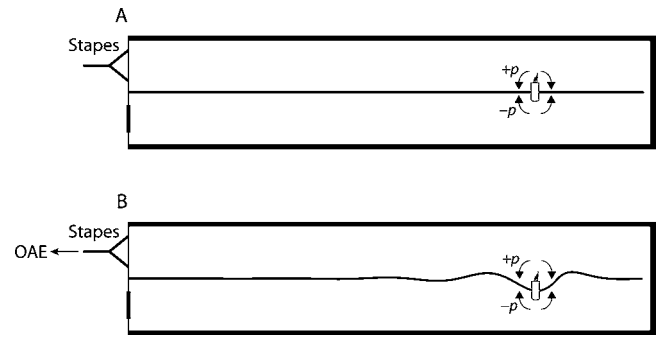


FIG. 3. An oscillating source that couples entirely into the slow wave (e.g., a transverse displacement source). In the case of DPOAEs, the latent source is activated by BM vibration at the frequencies of the primary tones (not shown). (A) The source produces a nonzero pressure gradient across the partition. (B) The resulting forces, transmitted by the fluids to varying degrees throughout the cochlea, move the BM and create slow pressure-difference waves (visible as “surface waves” on the BM) that propagate away in both directions. At the base, the pressure moves the stapes and round window, generating sound (an OAE) in the ear canal.

the stapes and round window, producing sound in the ear canal—an otoacoustic emission. Because the slow wave is the pressure component that drives the partition, models of cochlear mechanics (and of OAEs) have focused almost exclusively on this transverse mode.

2. Fast-wave sources

Our primary interest here, however, is with the other limit, illustrated in Fig. 4. In this case the oscillating “hair cell” produces the same pressure on both sides of the BM, and the source therefore couples entirely into the pressure-sum wave, $P_{\text{fast}}(x)$, which spreads throughout the cochlea nearly instantaneously. (The wavelength of the fast wave shown in the figure has been shortened considerably for the sake of presentation; at audio frequencies the wavelengths of sound in the fluid are typically long compared to the dimensions of the cochlea.⁶) We have presented versions of this toy model of fast waves elsewhere (Shera *et al.*, 2005; de Boer and Nuttall, 2006).



FIG. 4. An oscillating source that couples entirely into the fast wave (e.g., a volume source). (A) The source produces a compressional (sound) wave that quickly spreads throughout the cochlea. Although the fast wave moves the oval and round windows (generating an OAE) it does not itself directly drive the BM. (B) The asymmetry between the motions of the oval and round windows arising from their different acoustic impedances ($Z_{ow} \neq Z_{rw}$) generates a forward-traveling slow wave that moves the BM while propagating to its characteristic place.

The cochlea contains a distribution of sources whose strengths (amplitudes and phases) vary with position. Distortion-product sources oscillate at the DP frequency (in this case, $f_{dp}=2f_1-f_2$) and result from nonlinear interactions between the responses to the primaries. We let the complex function $q_{dp}^{fw}(x')$ represent the fast-wave DP source-strength density (or fast-wave pressure per unit length). The superscripted “fw” reminds us that the source is a fast-wave source rather than a more conventional slow-wave source. The function $q_{dp}^{fw}(x')$ depends on location x' and parametrically on the DP frequency and the stimulus parameters. Since the frequency f_{dp} is held constant in the Allen–Fahey paradigm we omit it from list of independent variables. In addition, we temporarily omit the stimulus parameters from the list for clarity; we include them later when needed.

To find the total fast-wave pressure at the base due to the distribution of sources we simply add their pressures up, weighting each by a propagator function $G_{fw}(x'|0)$ that accounts for any amplitude or phase shifts caused by fast-wave propagation from the site of generation (at $x=x'$) to the base (at $x=0$). In the simplest case (e.g., infinite propagation velocity) the weighting function $G_{fw}(x'|0)$ is unity. Adding up contributions from sources throughout the cochlea yields a general integral expression for $P_{dp}^{fast}(0)$:

$$P_{dp}^{fast}(0) = \int_0^L q_{dp}^{fw}(x') G_{fw}(x'|0) dx'. \quad (3)$$

To find $q_{dp}^{fw}(x')$ and $G_{fw}(x'|0)$ one could use appropriate pressure measurements (e.g., Dong and Olson, 2005) and/or a nonlinear cochlear model (e.g., Nobili and Mammano, 1996; Talmadge *et al.*, 2000; de Boer and Nuttall, 2000; Lim and Steele, 2002) extended to include fast waves and/or a nonzero compressibility of the organ of Corti (Shera *et al.*, 2005). It turns out, however, that these details are largely irrelevant for our purposes: We do not need to know either the fast-wave source density or the propagator to interpret the Allen–Fahey experiment in the fast-wave framework.

C. Producing OAEs from fast waves

When the compression wave reaches the base, it exerts forces on the oval and round windows. Just as in the slow-wave case, the middle ear transmits the motion of the oval window to the eardrum, creating an OAE in the ear canal [see Fig. 4(a)]. The linearity of middle-ear mechanics⁷ implies that P_{DPOAE} (the ear-canal pressure, or DPOAE) and $P_{dp}^{fast}(x=0)$ are directly proportional:

$$P_{DPOAE} = C_{me} P_{dp}^{fast}(0). \quad (4)$$

Because middle-ear transmission varies with frequency, the “constant” of proportionality, C_{me} , depends on f_{dp} . (In the Allen–Fahey paradigm, the DP frequency remains fixed so that C_{me} is, in fact, constant.) We emphasize that proportionality (4) does not depend on *how* the fast wave $P_{dp}^{fast}(0)$ is produced.

Although knowledge of the proportionality constant C_{me} is not needed to deduce qualitative fast-wave model predictions for the Allen–Fahey experiment, an explicit expression

for the DPOAE due to the fast-wave can be found using a general three-port description of the middle ear (Shera and Zweig, 1992b). The three-port description characterizes the acoustic and mechanical coupling between the ear canal and the oval and round windows. Although the general expression depends on middle-ear transfer functions and impedances in a complicated and relatively unenlightening way, the formula simplifies considerably if the impedances of the middle-ear cavity and round window are assumed negligible.⁸ In this limit,

$$P_{DPOAE} = \frac{1}{N_u + Z/Z_{mic}} P_{dp}^{fast}(0), \quad (5)$$

where N_u and Z are elements of the middle-ear transfer matrix (Shera and Zweig, 1992b).⁹ In particular, N_u is the volume-velocity transformer ratio (the ratio U_{ec}/U_{ow} when the cochlear load is “short-circuited”) and Z is the short-circuit transfer impedance (the corresponding ratio P_{ec}/U_{ow}). Z_{mic} is the Thévenin-equivalent impedance of the transducer in the ear canal.

D. Producing slow waves from fast waves

If the oval and round windows had identical acoustic impedances, the fast-wave pressure would move them in and out of the cochlea symmetrically. However, since the two windows have different impedances, they move with different volume velocities. This asymmetry between the motions of the windows launches a slow, pressure-difference wave that propagates down the cochlea, driving the motion of the BM en route. Thus, although the fast wave does not itself directly move the BM, the boundary conditions at the base—in particular, the asymmetry between the windows—creates a slow-wave component in the pressure that subsequently vibrates the BM [see Fig. 4(b)].

The three-port description of middle-ear mechanics implies that the amplitude of the induced *forward-traveling* slow wave is directly proportional to the *reverse-traveling* fast wave that produces it:

$$P_{dp}^{slow}(0) = C_{sof} P_{dp}^{fast}(0), \quad (6)$$

where $P_{dp}^{slow}(0)$ is the initial ($x=0$) amplitude of the forward-traveling slow wave launched from the stapes. The subscript “sof” on the proportionality constant stands for “slow/fast.” Just as with C_{me} in Eq. (4), the constant C_{sof} depends on f_{dp} . Solving the three-port equations yields an explicit expression for C_{sof} :

$$P_{dp}^{slow}(0) = \frac{Z_c(Z_{ow} - Z_{rw})}{Z_c(Z_{ow} + Z_{rw}) + 4Z_{ow}Z_{rw}} P_{dp}^{fast}(0), \quad (7)$$

where Z_c is the cochlear input impedance,¹⁰ and Z_{ow} and Z_{rw} are the impedances of the oval and round windows, respectively. The impedance Z_{ow} in Eq. (7) is the total impedance of the oval window seen from the cochlea; in other words, it includes contributions from the annular ligament, stapes, and other ossicles, as well as from the eardrum and any transducers present in the ear canal.¹¹ In this context, where the middle ear is driven from within the cochlea, the subscript “ow” might best be understood

as an abbreviation for “outside world” rather than merely for “oval window.”

Equation (7) shows explicitly how the slow wave launched from the base depends on the asymmetry between the windows. As expected, $P_{\text{dp}}^{\text{slow}}(0)$ goes to zero in the limit that Z_{ow} approaches Z_{rw} (as it might, for example, if the stapes were detached from the annular ligament). In the normal case, however, the impedances of the two windows are very different, and $P_{\text{dp}}^{\text{slow}}(0)$ is therefore nonzero. In the limit $Z_{\text{rw}} \rightarrow 0$ considered above, Eq. (7) simplifies to¹²

$$P_{\text{dp}}^{\text{slow}}(0) = P_{\text{dp}}^{\text{fast}}(0), \quad (8)$$

so that $C_{\text{sof}} = 1$.

E. Slow-wave propagation to the DP place

Once launched from the base, the forward-traveling DP slow wave propagates along the cochlea, moving the BM as it travels. The resulting BM velocity¹³ at the wave’s characteristic place is represented by the function

$$V_{\text{BM}}(P_0; r, P_r) \equiv \text{BM velocity at } (x_{\text{dp}}, f_{\text{dp}}), \quad (9)$$

where $P_0 \equiv P_{\text{dp}}^{\text{slow}}(0)$ is the slow wave at the stapes. For clarity, we now include in the list of independent parameters the amplitudes of the primaries (represented by the variable $P_r = \{P_1, P_2\}$) and the ratio of their frequencies, $r = f_2/f_1$. At fixed values of r and P_r , the BM velocity $V_{\text{BM}}(P_0; r, P_r)$ is a monotonically increasing, compressive function of P_0 .

Our primary interest here is not $V_{\text{BM}}(P_0; r, P_r)$ but its functional inverse; that is, the function that gives the value of the slow wave P_0 necessary to produce a given BM velocity at x_{dp} :

$$P_0 = P_{\Lambda}(V_{\text{BM}}; r, P_r). \quad (10)$$

As a mnemonic, we use the subscript “ Λ ” because it resembles an inverted “ V .” The function $P_{\Lambda}(V_{\text{BM}}; r, P_r)$ depends nonlinearly on V_{BM} because of cochlear compression (or “self-suppression”); the function depends on r and P_r because of possible “two-tone suppression” of the DP response by the primaries.¹⁴ Two-tone suppression is strongest when the responses to the primaries overlap with the region of DP amplification just basal to x_{dp} . Maximal overlap occurs at values of r close to 1. At larger values of r the spatial separation between the responses increases and suppression weakens. At the largest values of r the primaries do not appreciably suppress the response to the DP, and $P_{\Lambda}(V_{\text{BM}}; r, P_r)$ becomes independent of r and P_r . In other words,

$$\lim_{r \rightarrow \infty} P_{\Lambda}(V_{\text{BM}}; r, P_r) = P_{\Lambda}^0(V_{\text{BM}}), \quad (11)$$

where the superscript “0” on P_{Λ}^0 signifies the presumed absence of suppression by the primaries. In practice, the suppression-free limit is well approximated at rather modest values of r ; even at $r = 1.5$, the closest primary is an octave above f_{dp} .

To quantify the suppression at other values of r , we define the function

$$S(V_{\text{BM}}; r, P_r) \equiv \frac{P_{\Lambda}(V_{\text{BM}}; r, P_r)}{P_{\Lambda}^0(V_{\text{BM}})}. \quad (12)$$

In the absence of suppression by the primaries $S(V_{\text{BM}}; r, P_r) = 1$. Because suppression decreases the response to the DP, so that a larger value of P_0 is needed to achieve the same BM velocity,

$$|S(V_{\text{BM}}; r, P_r)| \geq 1 \quad (\text{at fixed } V_{\text{BM}}). \quad (13)$$

At low levels, the largest values of S (greatest suppression) generally occur at values of r close to or slightly above 1 (e.g., Pang and Guinan, 1997).

IV. COMPRESSION-WAVE PREDICTIONS FOR ALLEN–FAHEY EXPERIMENTS

We are now in a position to obtain compression-wave model predictions for the Allen–Fahey paradigm (i.e., to find the ear-canal DPOAE when the response to the DP is held constant at x_{dp}). In the following discussion we assume that the DP response measure is BM velocity, and denote its constant criterion value by V_{BM}^{\dagger} :

$$V_{\text{BM}}^{\dagger} \equiv \text{criterion value of } V_{\text{BM}}(x_{\text{dp}}, f_{\text{dp}}). \quad (14)$$

Henceforth, a superscripted dagger on any variable means that the quantity is measured or evaluated when the BM velocity at x_{dp} is V_{BM}^{\dagger} .

A. Predictions without tuned suppression

For simplicity, we first neglect any suppression by the primaries (e.g., Allen and Fahey, 1992). In other words, we take $S(V_{\text{BM}}; r, P_r) = 1$ so that $P_{\Lambda} = P_{\Lambda}^0(V_{\text{BM}})$. According to Eq. (10), the stapes slow-wave pressure $P_{\text{dp}}^{\text{slow}}(0)^{\dagger}$ necessary to produce the criterion BM velocity V_{BM}^{\dagger} is then given by

$$P_{\text{dp}}^{\text{slow}}(0)^{\dagger} = P_0^{\dagger} = P_{\Lambda}^0(V_{\text{BM}}^{\dagger}). \quad (15)$$

Working backwards to find the corresponding DPOAE, we find that the fast-wave pressure $P_{\text{dp}}^{\text{fast}}(0)^{\dagger}$ needed to produce the slow-wave pressure $P_{\text{dp}}^{\text{slow}}(0)^{\dagger}$ follows from proportionality (6):

$$P_{\text{dp}}^{\text{fast}}(0)^{\dagger} = P_{\text{dp}}^{\text{slow}}(0)^{\dagger} / C_{\text{sof}}. \quad (16)$$

In the fast-wave Allen–Fahey paradigm, the value of $P_{\text{dp}}^{\text{fast}}(0)^{\dagger}$ at any given value of r is obtained by adjusting the primaries P_r until the distribution of fast-wave sources $\mathcal{Q}_{\text{dp}}^{\text{fw}}(x; r, P_r)$ combine [as in Eq. (3)] to yield $P_{\text{dp}}^{\text{fast}}(0; r, P_r) = P_{\text{dp}}^{\text{fast}}(0)^{\dagger}$; by definition this occurs at $P_r = P_r^{\dagger}(r)$. Once the necessary value of $P_{\text{dp}}^{\text{fast}}(0)^{\dagger}$ has been obtained, proportionality (4) relating P_{DPOAE} and $P_{\text{dp}}^{\text{fast}}(0)$ implies that

$$P_{\text{DPOAE}}^{\dagger} = C_{\text{me}} P_{\text{dp}}^{\text{fast}}(0)^{\dagger}. \quad (17)$$

Combining Eqs. (15)–(17) yields the formula

$$P_{\text{DPOAE}}^{\dagger} \Big|_{\text{no sup}} = \frac{C_{\text{me}}}{C_{\text{sof}}} P_{\Lambda}^0(V_{\text{BM}}^{\dagger}) = \text{const.} \quad (18)$$

Thus, in the absence of tuned suppression, fast-wave models imply that when the BM velocity $V_{\text{BM}}(x_{\text{dp}}, f_{\text{dp}})$ is held constant (at V_{BM}^{\dagger}), the DPOAE measured in the ear canal remains constant (at $P_{\text{DPOAE}}^{\dagger}$), independent of r . Refer-

ence to the synopsis of Allen–Fahey experiments shown in Fig. 1 demonstrates that a constant value of P_{DPOAE}^\dagger is not the result obtained experimentally, especially at values of r close to 1. Does this discrepancy between fast-wave predictions and experiment arise because of suppression?

B. Predictions with tuned suppression

In the more realistic case, the primaries partially suppress the response to the DP slow wave, modifying the function $P_\Lambda^0(V_{\text{BM}}^\dagger)$ by the factor $S(V_{\text{BM}}^\dagger; r, P_r^\dagger)$ as described by Eq. (12). As a result, the analog of Eq. (18) becomes

$$\begin{aligned} P_{\text{DPOAE}}^\dagger(r) &= \frac{C_{\text{me}}}{C_{\text{sof}}} P_\Lambda(V_{\text{BM}}^\dagger; r, P_r^\dagger) \\ &= S(V_{\text{BM}}^\dagger; r, P_r^\dagger) P_{\text{DPOAE}}^\dagger|_{\text{no sup}}, \end{aligned} \quad (19)$$

where $P_r^\dagger = P_r^\dagger(r)$ and $P_{\text{DPOAE}}^\dagger|_{\text{no sup}}$ is the constant DPOAE amplitude measured in the absence of suppression [Eq. (18)].

In Sec. III E we argued that at fixed V_{BM} , the suppression function $|S(V_{\text{BM}}; r, P_r)|$ generally increases as $r \rightarrow 1$ [Eq. (13)]. Thus, when tuned suppression is taken into account fast-wave models imply that the ear-canal DPOAE measured using the Allen–Fahey paradigm *increases* at close primary ratios (i.e., as f_1 decreases towards f_{dp}). However, this is exactly the opposite of the trend observed experimentally. As reviewed in Sec. II, the common feature of Allen–Fahey experiments performed to date is a general *decrease* in the value of $P_{\text{DPOAE}}^\dagger(r)$ as $r \rightarrow 1$ [Fig. 1].¹⁵ *The results of Allen–Fahey experiments cannot be explained using fast-wave models of OAEs.*

V. DISCUSSION

A. Recapitulation

Our analysis demonstrates that compression-wave model proposed in the literature (e.g., Wilson, 1980b; Ren, 2004; Ren *et al.*, 2006; Ren and Nuttall, 2006) cannot account for results of Allen–Fahey and related experiments. Indeed, the contradiction with experiment extends to any compression-wave model, independent of the details by which the distortion sources (“hair cells”) are presumed to couple sound into the cochlear fluids. In the absence of suppression, compression-wave models predict a constant value of $P_{\text{DPOAE}}^\dagger(r)$. Rigorously derived in Secs. III and IV, this prediction can be deduced immediately from the bottom panel of Fig. 5: In fast-wave models, fixing the DP response at x_{dp} is equivalent to fixing both fast and slow waves at the stapes [Eqs. (15) and (16)]; this, in turn, is equivalent to fixing the DPOAE in the ear canal [Eq. (17)]. Note that the prediction of constant $P_{\text{DPOAE}}^\dagger(r)$ applies *independent of the power gain of the cochlear amplifier*. No matter where the fast-wave DP originates, no matter how the source-region F in Fig. 5 moves around as r varies, the forward-traveling slow-wave that drives—and may be amplified by—the cochlear partition always originates at the stapes and, therefore, always traverses the full extent of the amplifier gain region on its way to x_{dp} .

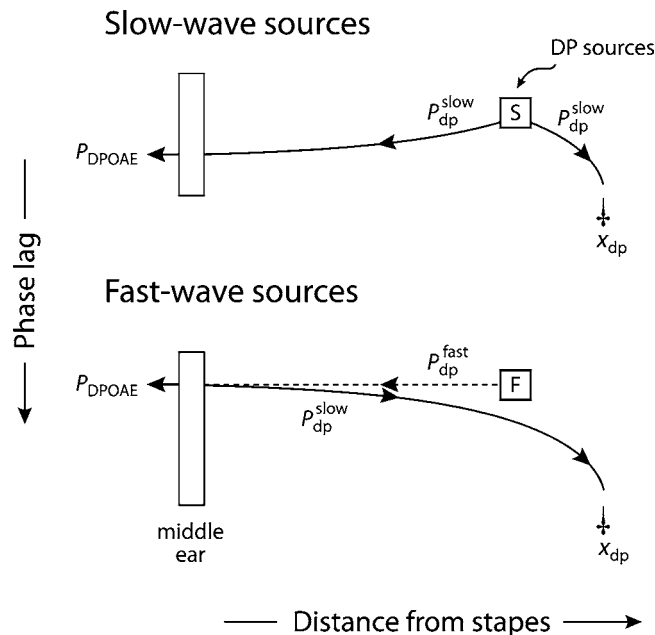


FIG. 5. Schematic illustration of the generation of slow- and fast-wave model DPs at $2f_1 - f_2$. Both panels show wave phase lag (increasing downward) vs cochlear location. In each case nonlinear interactions between stimulus slow waves at frequencies f_1 and f_2 (not shown) create a region of DP sources (square boxes). In the top panel, slow-wave DP sources S produce slow waves $P_{\text{dp}}^{\text{slow}}$ (curved solid lines) that travel along the BM in both directions. The curved phase trajectories indicate that the wavelength depends on the spatially varying properties of the cochlear partition. In the bottom panel, fast-wave DP sources F produce a compressional wave $P_{\text{dp}}^{\text{fast}}$ (straight dashed line) that quickly spreads throughout the cochlea, reaching the stapes with minimal phase lag. At the stapes, the fast wave creates a forward-traveling slow wave (curved solid line). Note that both slow- and fast-wave scenarios produce a forward-traveling slow wave that drives the BM at x_{dp} . In the slow-wave model, the forward-traveling wave radiates directly from the source region, whose location and effective “directionality” depend on the primary frequency ratio, r . In the fast-wave model, the slow wave is generated at the stapes, whose location is fixed, independent of r . The Allen–Fahey paradigm prescribes that the BM response to the forward-traveling slow wave remain constant at x_{dp} (indicated by the inverted dagger used as an arrow). For simplicity, the diagram ignores complications (largely irrelevant for our purposes) arising from multiple internal reflections (e.g., the reflection of slow waves from the stapes and/or the generation of fast waves near x_{dp}). Adapted from Shera *et al.* (2006).

How does the existence of suppression modify these predictions? As r decreases towards 1, the primaries draw closer to x_{dp} and their suppressive action reduces the response to the DP at x_{dp} . To maintain the constant response mandated by the Allen–Fahey paradigm, the DP source output must be increased (e.g., by boosting the levels of the primary tones). When the source output is increased, the fast-wave pressure at the stapes and the DPOAE in the ear canal both increase correspondingly [Eq. (19)]. In the presence of tuned suppression, fast-wave models therefore predict that the ear-canal DPOAE will increase at close ratios. Neither of the two fast-wave model predictions—constant $P_{\text{DPOAE}}^\dagger(r)$ without suppression, increasing $P_{\text{DPOAE}}^\dagger(r)$ with suppression—appears consistent with the experimental result, which indicates that the ratio of ear-canal to intracochlear DPs falls as $r \rightarrow 1$ (cf. Fig. 1).

Our conclusion about fast-wave models contradicts that of a recent paper by Ren and Nuttall (2006), who argue that the results of the Allen–Fahey experiment “imply that the

cochlea-generated emission reaches the stapes...as a compression [fast] wave rather than as a backward traveling [slow] wave.” We question the logic of their argument on one crucial point: It overlooks the fact that in the fast-wave models they posit, the ear-canal DPOAE and the forward-traveling slow wave that drives the BM are both proportional to $P_{dp}^{fast}(0)$ [Eqs. (4) and (6); see also Fig. 5]; as a result, the two cannot vary independently, as their argument requires.¹⁶

B. Slow-wave interpretations of Allen–Fahey experiments

Although compression-wave models fail to account for the data, the results of Allen–Fahey and related experiments *can* be understood in the slow-wave framework. At least two different mechanisms have been proposed, both of which implicitly exploit the additional flexibility (in this case, r -dependence) inherent in the slow-wave mechanism. The slow-wave cartoon shown in the top panel of Fig. 5 illustrates this flexibility graphically. Note, in particular, that the forward- and reverse DP waves that emanate from S need not maintain the same amplitude ratio at all values of r . Contrast this with the fast-wave model illustrated in the bottom panel. In this case, the ratio of reverse fast wave to forward slow wave is determined, independent of r , by impedance relationships at the stapes that remain constant for a given DP frequency [Eq. (6)].

Allen and Fahey’s (1992) slow-wave interpretation of their results was based on the observation that a decrease in $P_{DPOAE}^{\dagger}(r \rightarrow 1)$ would occur if the power gain of the cochlear amplifier were less than 1; that is, if the acoustic power in the slow wave driving the BM were absorbed by the organ of Corti in the region basal to the peak [rather than being amplified, as solutions to the cochlear inverse problem indicate (e.g., Zweig, 1991; de Boer, 1995a, b)]. As Allen and Fahey demonstrate, the reason for the dependence on power gain (or loss) is that as $r \rightarrow 1$, the slow wave radiating apically from the source region S toward its characteristic place traverses an ever decreasing fraction of the amplifier gain/loss region on its way to the measurement point at x_{dp} .

Unfortunately for this simple and intuitive explanation, a direct relationship between the behavior of $P_{DPOAE}^{\dagger}(r)$ and the gain of the cochlear amplifier requires that the wave pattern radiating from the DP source region remain fixed, independent of r (Allen and Fahey, 1992; Shera and Zweig, 1992a). Although individual DP sources (e.g., hair cells) may always radiate symmetrically, a distribution of DP sources whose relative phases vary systematically with position can become highly directional in a manner that depends strongly on r (Shera, 2003; Shera and Guinan 2007b). Needless to say, these are just the types of distributions expected in the cochlea. Indeed, model calculations show that at close ratios realistic distributions of (slow-wave) distortion sources radiate much more strongly toward x_{dp} than they do back toward the stapes (Shera and Guinan, 2007a).¹⁷ As a result, and despite possible countervailing effects of amplifier gain and tuned suppression, fixing the response at x_{dp} causes the corresponding ear-canal DPOAE to fall as $r \rightarrow 1$. Thus, in slow-wave models, changes with r in the effective *directionality* of

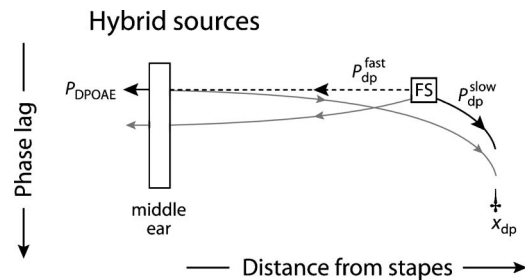


FIG. 6. Schematic illustration of a conceptual hybrid-wave model that combines elements of the slow- and fast-wave source models shown in Fig. 5. Nonlinear interactions between stimulus slow waves at frequencies f_1 and f_2 (not shown) create a region FS of hybrid DP sources that couple strongly into both slow and fast waves. By hypothesis, the slow- and fast-wave coupling strengths have somehow been adjusted so that: (1) The DPOAE in the ear canal arises predominantly by coupling into fast waves and (2) the BM vibration at x_{dp} arises predominantly by direct coupling into slow waves within FS rather than indirectly through impedance relationships at the stapes. Other contributing but non-dominant pathways are shown in gray.

the waves radiated from the DP source region can qualitatively account for results of Allen–Fahey experiments, even when the gain of the cochlear amplifier is greater than one (Shera, 2003; de Boer *et al.*, 2005; Shera and Guinan, 2007a).

This latter slow-wave explanation for the results of Allen–Fahey experiments is consistent with a number of theoretical studies that trace the bandpass form of the so-called “DP filter” to interference effects within the distributed source region (van Hengel, 1996; Talmadge *et al.*, 1998; Kemp and Knight, 1999; Shera, 2003). Fahey *et al.* (2006) have recently provided experimental evidence supporting the suggestion that the decrease in DPOAE amplitude as $r \rightarrow 1$ arises primarily from vector cancellation among multiple sources. In addition, measurements of the relative amplitudes of the “distortion” and “reflection” components of DPOAEs (e.g., Knight and Kemp, 2000) provide additional experimental evidence for changes in distortion-source “directionality” as a function of f_2/f_1 (reviewed in Shera and Guinan, 2007b).

C. Hypothetical hybrid-wave models

The contradiction between compression-wave models and Allen–Fahey experiments hinges on the compression-wave model’s eponymous assumption that distortion sources couple predominantly into compression waves—as a consequence, both the ear-canal DPOAE and the cochlear response to the DP at its characteristic place ultimately arise from coupling into fast waves.¹⁸ Despite the failure of fast-wave models, the success of the slow-wave framework raises the intriguing possibility that a purpose-built “hybrid-wave” model, artfully adjusted to combine fast- and slow-wave coupling as needed, might yield fast-wave dominated DPOAEs while maintaining consistency with the Allen–Fahey experiment. In particular, consider the hybrid-wave possibility illustrated in Fig. 6. The DP sources in this conceptual model couple into both slow and fast waves. By hypothesis, their relative coupling strengths are arranged so that whereas the ear-canal DPOAE arises predominantly by fast-wave coupling (as in compression-wave models), the cochlear re-

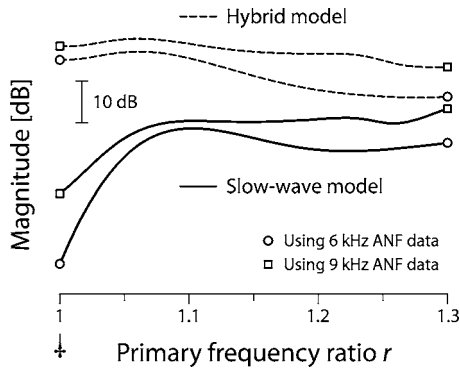


FIG. 7. Hybrid- and slow-wave Allen–Fahey functions simulated in the cat. The figure shows the magnitude of the Allen–Fahey function $P_{\text{DPOAE}}^{\dagger}(r)$ computed using equations derived from a one-dimensional transmission-line model of the cat cochlea extended to include both slow- and fast-wave DP sources (Shera *et al.*, 2005). The dashed lines show hybrid-wave results; the solid lines show slow-wave results reproduced from Fig. 13 of Shera and Guinan (2007a). The DP frequency was fixed at 7 kHz and responses were computed as a function of the primary frequency ratio, whose value increases along the abscissa. As described elsewhere (Shera and Guinan, 2007a), the model was customized for cat by solving the inverse problem using ANF-derived measurements of the amplitude and phase of cochlear tuning (van der Heijden and Joris, 2003). Each pair of curves shows model results based on tuning measurements at CFs of 6 and 9 kHz (\circ and \square , respectively). All model responses have been offset vertically and grouped for clarity.

sponse to the DP at x_{dp} arises predominantly by slow-wave coupling (as in slow-wave models). Does this hypothetical hybrid-wave model predict the results of the Allen–Fahey experiment?

Although we cannot address the question in full generality, the computational results shown in Fig. 7 indicate that the hybrid-wave model qualitatively disagrees with the data, at least when it employs the same simple and theoretically natural spatial distributions of DP sources that work well in the slow-wave framework (Shera, 2003; Shera and Guinan, 2007a). The figure shows Allen–Fahey functions computed for a hybrid model in which the fast-wave source density at $2f_1 - f_2$ was taken to have the form

$$\varrho_{\text{dp}}^{\text{fw}}(x) \propto V_{\text{BM}}^2(x; f_1) V_{\text{BM}}^*(x; f_2), \quad (20)$$

where $V_{\text{BM}}(x; f)$ is the BM traveling wave at the indicated frequency. Source density functions of this form result from approximating the nonlinearity by the cubic term of a power-series expansion in the local BM velocity (e.g., Talmadge *et al.*, 1998). Estimates of BM traveling waves [e.g., $V_{\text{BM}}(x; f_1)$ and $V_{\text{BM}}(x; f_2)$] were obtained for the cat cochlea by converting ANF-derived measurements of the amplitude and phase of cochlear tuning (van der Heijden and Joris, 2003) to spatial responses using local scaling (e.g., Zweig, 1976).

The fast-wave DPOAE was then computed using Eqs. (3) and (4), with the fast-wave propagator, $G_{\text{fw}}(x|0)$, set equal to 1. A constant real propagator means that the waves are assumed to propagate with negligible phase shift in a manner independent of the properties of the cochlear partition. The integral becomes (Shera *et al.*, 2005)

$$P_{\text{DPOAE}} \propto P_{\text{dp}}^{\text{fast}}(0) = \int_0^L \varrho_{\text{dp}}^{\text{fw}}(x) dx. \quad (21)$$

The BM velocity at x_{dp} was computed using an analogous integral (e.g., Shera, 2003; Shera and Guinan, 2007a):

$$V_{\text{BM}}(x_{\text{dp}}; f_{\text{dp}}) \propto P_{\text{dp}}^{\text{slow}}(x_{\text{dp}}) = \int_0^{x_{\text{dp}}} \varrho_{\text{dp}}^{\text{sw}}(x) G_{\text{sw}}(x|x_{\text{dp}}) dx, \quad (22)$$

where $G_{\text{sw}}(x|x_{\text{dp}})$ is the slow-wave propagator at the DP frequency. Unlike the fast-wave propagator, $G_{\text{sw}}(x|x')$ depends on the mechanical characteristics (i.e., tuning) of the cochlear partition. For simplicity, the slow-wave source density, $\varrho_{\text{dp}}^{\text{sw}}$, was taken proportional to $\varrho_{\text{dp}}^{\text{fw}}$ (Shera *et al.*, 2005). Aside from an overall constant of proportionality, the Allen–Fahey function $P_{\text{DPOAE}}^{\dagger}(r)$ is then simply the ratio $P_{\text{DPOAE}}/V_{\text{BM}}$ computed from Eqs. (21) and (22). Note that these hybrid-model calculations ignore contributions from the concomitant—but, by hypothesis, nondominant—pathways shown by the gray lines in Fig. 6.¹⁹

Figure 7 demonstrates that the conceptual hybrid-wave model outlined above does not reproduce the substantial downturn characteristic of Allen–Fahey functions at close ratios. Because our quasilinear calculation ignores the influence of suppression, the disagreement is worse than it appears in the figure; including suppressive effects would further increase the model value of the Allen–Fahey function at close ratios (see also Shera and Guinan, 2007a). The hybrid-wave results contrast strongly with those computed for the slow-wave model, implemented here using all identical assumptions but one: Rather than reaching the stapes via fast waves, the dominant contribution to the DPOAE is assumed to propagate via slow waves, making it subject to additional filtering due to the mechanical tuning of the partition. For the slow-wave model, Eq. (21) is modified to read

$$P_{\text{DPOAE}} \propto P_{\text{dp}}^{\text{slow}}(0) = \int_0^L \varrho_{\text{dp}}^{\text{sw}}(x) G_{\text{sw}}(x|0) dx. \quad (23)$$

This relatively simple change substantially increases the amount of destructive interference among the distortion-source wavelets, causing the ear-canal DPOAE to fall at close ratios.

Thus, although we cannot definitively rule out all hybrid-wave models, computations performed using a seemingly natural implementation do not reproduce the data. Our conclusion is necessarily qualified because the additional freedom inherent in hybrid-wave models, combined with a lack of knowledge of relevant biophysical details, means that obtaining hybrid-wave predictions for the Allen–Fahey experiment requires more assumptions than were needed to rule out compression-wave models. Because of the additional uncertainties that accompany these assumptions, our analysis and computations cannot invalidate all hybrid-wave models. Nevertheless, they do indicate that at least one simple class of hybrid-wave models predicts Allen–Fahey functions inconsistent with experiment. The model’s key shortcoming is that it does not provide enough destructive

interference to produce a substantial decrease in the fast-wave DPOAE as $r \rightarrow 1$. Furthermore, the hybrid-wave model suffers on purely theoretical grounds: Although simplicity and parsimony can be unreliable guides, the model's requirement that slow- and fast-wave sources radiate similar amounts in opposite directions seems suspiciously *ad hoc*, relying as it does on an apparently fortuitous collusion of coupling strengths for which there is as yet no biophysical explanation.

D. A source of potential confusion

Before concluding, we address a confusion apparent in the literature and latent in many descriptions of the so-called “hydrodynamic model” of OAEs (Nobili *et al.*, 2003). Faced with a continuing disavowal of cochlear traveling waves²⁰ in preference for language proclaiming that “emissions are propagated almost instantaneously through the fluid” (Vet\v{e}šník *et al.*, 2006), readers may misunderstand the hydrodynamic model as involving fast-wave coupling of the sort proposed by others (Wilson, 1980b; Ren, 2004; Ren *et al.*, 2006; Ruggero, 2004; Siegel *et al.*, 2005). Any correspondence between the model and experiment might then naturally be regarded as theoretical support for fast-wave mechanisms. So let us be clear: *The hydrodynamic model is a slow-wave model* in the sense used here. As a review of the relevant equations readily verifies (Mammano and Nobili, 1993; Nobili *et al.*, 2003), the active sources in the hydrodynamic model do *not* couple into fast, compression waves of the type illustrated schematically in Fig. 4; rather, they couple entirely into slow, pressure-difference waves, as illustrated in Fig. 3.

In both slow- and fast-wave models, mechanical forces are communicated throughout the cochlea almost instantaneously by the fluids. No one disputes this basic physical fact, which follows from the near incompressibility of the cochlear contents. Why then do the two models yield different otoacoustic predictions (e.g., for Allen–Fahey experiments)? For one simple reason: Because fast- and slow-wave forces act differently on the structures within and bounding the cochlea. In particular, slow-wave forces—but not fast-wave forces—drive the motion of the BM as they propagate. Thus, pressure differences associated with slow-wave coupling at one location drive BM motion and activate BM mechanisms at other locations. Each of these mechanical reactions produces in turn new slow-wave forces that act to drive BM motion at other locations, and so on. All these myriad slow-wave interactions combine with appropriate amplitudes and phases. By this collective process the emission measured in the ear canal emerges with its own characteristic wave-travel delay.

All of this interconnected slow-wave action is correctly captured by the mathematics of the hydrodynamic model. But some accompanying descriptions (e.g., “emissions are propagated almost instantaneously through the fluid”) have the potential to confuse, not only because they suggest (incorrectly) that the hydrodynamic model involves fast-wave mechanisms, but also because the instantaneous propagation of individual hair-cell forces is not the same thing as instan-

aneous propagation of *emissions*. Otoacoustic emissions do not arise from single, isolated sources; they are a collective response whose measured properties (e.g., time course, effective propagation velocity) can be very different from those of the instantaneous “conceptual responses” into which they can be mathematically decomposed.

E. Additional evidence for slow-wave OAEs

Although our aim in this paper has been to understand the constraints that Allen–Fahey experiments place on modes of DPOAE propagation, our conclusions limiting the role of fast-wave DPOAEs are consistent with other independent arguments for the dominance of slow-wave otoacoustic emissions. These arguments include: (1) Evidence for mechanical filtering of DPOAEs presented by the strikingly different phase-gradient delays manifest by lower- and upper-sideband distortion products (Shera *et al.*, 2006); (2) the near-zero phase-gradient delays (i.e., “wave-fixed” behavior) predicted, in contradiction to experiment, by compression-wave models of SFOAEs (Shera *et al.*, 2005; de Boer and Nuttall, 2006); (3) hints of reverse DP propagation on the basilar membrane revealed by population studies of the auditory-nerve response to combination tones (see Fig. 6, Kim *et al.*, 1979; or Fig. 5, Kim *et al.*, 1980); and (4) the argument from reciprocity: If hair cells generate compression waves strong enough to dominate the motion of the stapes in the reverse direction, compression waves generated by the stapes might be expected to drive the hair cells directly in the forward direction, in contradiction both to the long first-spike latencies in fibers innervating the apex and measurements of cochlear responses to identical pressures applied at the oval and round windows (Voss *et al.*, 1996).

ACKNOWLEDGMENTS

We thank Nigel Cooper, Christopher Bergevin, Elizabeth Olson, Robert Withnell, and the anonymous reviewers for helpful comments on the manuscript, Tianying Ren for valuable discussions, and Marcel van der Heijden for graciously sharing his data. This work was supported by Grant Nos. RO1 DC03687 (C.A.S.) and RO1 DC00235 (J.J.G.) from the NIDCD, National Institutes of Health.

¹The slow-wave/fast-wave controversy was rehearsed during the discussion session at the Eighth Mechanics of Hearing Workshop (see Allen, 2003, p. 583ff) and then again at the Ninth (see Allen, 2006, p. 534ff).

²This paper amplifies an abridged (nine-sentence) argument published in the proceedings of the Ninth Mechanics of Hearing Workshop held 23–28 July 2005 (Shera *et al.*, 2006).

³Secondary distortion products have been used by others to establish that distortion products propagate within the cochlea much like traveling waves excited by external tones (e.g., Goldstein *et al.*, 1978; Whitehead *et al.*, 1993).

⁴Allen and Fahey (1992) summarize their results as follows: “From the extended data base, we noticed that as f_2 decreases [$r \rightarrow 1$], p_d [i.e., $P_{\text{DPOAE}}^+(r)$ in our notation] tends to slightly (e.g., 6 dB) decrease.”

⁵As indicated in Fig. 1, the DPOAE noise floor in the original experiments was at or just below 20 dB SPL (Fahey and Allen, 1985; Allen and Fahey, 1992). The neural units measured had CF thresholds in the range 20–40 dB SPL and most often around 30 dB SPL (most were medium spontaneous-

rate units). As a result, the measurements of $P_{\text{DPOAE}}^{\dagger}(r)$ were confined to the relatively small dynamic range between the otoacoustic noise floor and the level at CF.

⁶In humans, for example, the ratio c/fL —where c is the speed of sound in water, f is the frequency, and L is the cochlear length (≈ 35 mm)—is about 40 at 1 kHz.

⁷For simplicity, we are ignoring level-dependence arising from activation of the middle-ear reflex that might occur if the experiment were performed in an unanesthetized subject.

⁸These approximations are not unrealistic, especially in animal experiments in which the middle-ear cavities are opened widely to the atmosphere. In addition, measurements in the cat indicate that the impedance of the round window is negligible except at very low frequencies (Lynch *et al.*, 1982).

⁹The 2×2 transfer matrix ${}^{\text{ec}}T_{\text{ow}} = \begin{pmatrix} N_p^{\dagger} & Z \\ Y & N_u \end{pmatrix}$ is defined by $\begin{pmatrix} P_{\text{ec}} \\ U_{\text{ec}} \end{pmatrix} \equiv {}^{\text{ec}}T_{\text{ow}} \begin{pmatrix} P_{\text{ow}} \\ U_{\text{ow}} \end{pmatrix}$, where variables P and U represent pressures and volume velocities. Note that $P_{\text{ow}} = P_{\text{sv}}(0)$ and $U_{\text{ow}} = U_{\text{sv}}(0)$. The principle of reciprocity implies that $\det {}^{\text{ec}}T_{\text{ow}} = N_u/N_p - ZY = 1$ (Shera and Zweig, 1991). The matrix ${}^{\text{ec}}T_{\text{ow}}$ is elsewhere written with a diacritical hat to indicate that the pressures P_{ec} and P_{ow} are measured with respect to a “ground” located in the tympanic cavity (Shera and Zweig, 1992b). For simplicity, we suppress that notation here since we are assuming (see Note 8) that the cavity impedance can be neglected (or, alternatively, that the cavities have been opened widely to the atmosphere).

¹⁰The cochlear input impedance is defined as $Z_c \equiv (P_{\text{slow}}/U_{\text{slow}})|_{x=0}$, where the slow-wave volume velocity $U_{\text{slow}}(x)$ is defined in terms of the volume velocities in the two scalae by the equation $U_{\text{slow}} = \frac{1}{2}(U_{\text{sv}} - U_{\text{st}})$.

¹¹More specifically, the three-port description indicates that

$$Z_{\text{ow}} = \frac{N_p(Z + N_u Z_{\text{mic}})}{1 + N_p Y Z_{\text{mic}}},$$

where the elements of the transfer matrix ${}^{\text{ec}}T_{\text{ow}}$ are defined in Note 9 and Z_{mic} is the Thévenin-equivalent impedance of the microphone.

¹²The simplification requires $|Z_{\text{rw}}| \ll |Z_{\text{ow}}|$ and $|Z_{\text{rw}}| \ll |Z_c|/4$. See also Note 8.

¹³Although we focus here on BM velocity, other response measures used in Allen–Fahey experiments (e.g., ANF firing rate) could be used as well.

¹⁴In addition to suppressing the response to the DP, the primaries also suppress one another. The effects of such “mutual suppression” between the primaries are implicitly included in the discussion through their effect on the form of $P_A(V_{\text{BM}}; r, P_r)$. Mutual suppression also plays an important role in determining the DP source strength, and thus the value of $P_r^{\dagger}(r)$ introduced in Sec. IV A.

¹⁵This statement requires some qualification in order not to oversimplify the complexities of the experimental situation. Fahey and Allen (1985) measured the level of a suppressor tone that would be required to suppress a tone that was approximately 6 dB above the neural threshold down to neural threshold. For suppressors both below and above CF a 65 dB SPL suppressor would suppress a tone at CF by 6 dB. Whereas below CF this effect was fairly independent of the frequency of the suppressor, to obtain 6 dB of suppression above CF requires that the suppressor frequency be just barely greater than a frequency that would excite the neuron (i.e., just above the upper edge of the tuning curve). Therefore, when suppression plays a significant role in an ANF-based Allen–Fahey experiment, it will be for the higher neural threshold units, because in these units the tuning of high-frequency suppression has levels greater than the distortion tuning curve. This is also where the signal to noise ratio is adequate in the original Allen–Fahey experiment. Their original data show no evidence for suppression of a fast wave component [i.e., for an increase in $P_{\text{DPOAE}}^{\dagger}(r)$ under conditions where it would have been predicted [e.g., Fig. 6 of Allen and Fahey (1992) and in about a dozen unpublished units]. However, the data also show no universal trend of a decrease in $P_{\text{DPOAE}}^{\dagger}(r)$ as $r \rightarrow 1$ of the magnitude shown in Figs. 1(c) and 1(d) (see Note 4). Although the dynamic range of the experiment is rather limited, the most common trend in the original Allen and Fahey data set was neither an overall increase nor an overall decrease as $r \rightarrow 1$. In summary, the data of the original Allen–Fahey experiment are not generally consistent with the predictions of the fast-wave mechanism; the data from the other three experiments shown in Fig. 1 are strongly inconsistent with the fast-wave mechanism.

¹⁶In addition, Ren and Nuttall’s argument presumes that the amplitude of the forward-traveling pressure-difference wave decreases as it approaches its characteristic place. However, pressure measurements in gerbils indicate that the slow-wave pressure *increases* near the characteristic place (Olson, 2001), as predicted by active and/or three-dimensional cochlear models.

¹⁷Fast-wave models also generally exhibit vector cancellation at close ratios, although the destructive interference is not generally as strong as it is in slow-wave models (see Fig. 7). Regardless of its strength, however, the cancellation in fast-wave models does not cause the Allen–Fahey function to fall. In fast-wave models, the DPOAE level is tightly coupled to the response at x_{dp} because the slow wave is triggered by the fast wave at the stapes: Aside from the possible influence of suppression, fixing the response at one location (e.g., x_{dp}) fixes it at the other (i.e., the ear canal).

¹⁸The coupling can, of course, be indirect: Recall that in fast-wave models the forward slow wave vibrating the cochlear partition at x_{dp} is launched from the stapes by a fast wave interacting with the mechanical asymmetry between the oval and round windows.

¹⁹Because of the impedance asymmetry between the windows, compression waves invariably trigger slow waves at the stapes [see Eq. (7)]. Furthermore, in the hybrid-wave models we have considered to date (e.g., Shera *et al.*, 2005), it is impossible to produce forward slow waves without also generating reverse slow waves. We expect the pathways shown in gray—and other higher-order pathways, such as those arising from multiple internal reflection of slow waves at the DP place and/or the stapes—to produce interference ripples in the Allen–Fahey function.

²⁰The criticism directed at wave-equation or transmission-line approaches to modeling OAEs (e.g., Nobili *et al.*, 2003; Vetešník *et al.*, 2006) has no basis in the physics of slow-wave propagation through the fluid. Whether analyzed using Green’s functions or equivalent wave equations, the slow-wave physics is manifest mathematically as an identity between the long-range force propagator for the three-dimensional hydrodynamic model and the corresponding propagator for a one-dimensional transmission line (Shera *et al.*, 2004). With respect to the long-range force propagation important for modeling OAEs, the hydrodynamic model *is* a transmission line. Thus, when advocates of the hydrodynamic model draw a sharp rhetorical divide between the “transmission-line concept” for which “quantitative predictions...are still lacking” and their own “much more predictive” approach (Vetešník *et al.*, 2006), they force a verbal distinction where no physical or mathematical difference exists. The two approaches to modeling OAEs are equivalent; they employ the same hydrodynamical equations and must, therefore, make the same predictions. An earlier publication elaborates on this physical equivalence in some detail (Shera *et al.*, 2004).

Allen, J. B. (2003). “Moderated discussion session,” in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 563–592.

Allen, J. B. (2006). “Moderated discussion session,” in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer, (World Scientific, Singapore), pp. 521–543.

Allen, J. B., and Fahey, P. F. (1992). “Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane,” *J. Acoust. Soc. Am.* **92**, 178–188.

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.

de Boer, E. (1995a). “The ‘inverse problem’ solved for a three-dimensional model of the cochlea. I. Analysis,” *J. Acoust. Soc. Am.* **98**, 896–903.

de Boer, E. (1995b). “The ‘inverse problem’ solved for a three-dimensional model of the cochlea. II. Application to experimental data sets,” *J. Acoust. Soc. Am.* **98**, 904–910.

de Boer, E., and Nuttall, A. L. (2000). “The mechanical waveform of the basilar membrane. III. Intensity effects,” *J. Acoust. Soc. Am.* **107**, 1497–1507.

de Boer, E., and Nuttall, A. L. (2006). “Amplification by ‘compression waves’ in the cochlea—A parable,” *Assoc. Res. Otolaryngol. Abs.* **29**, 989.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (2005). “The Allen–Fahey experiment extended,” *J. Acoust. Soc. Am.* **107**, 1260–1266.

Dong, W., and Olson, E. S. (2005). “Two-tone distortion in intracochlear pressure,” *J. Acoust. Soc. Am.* **117**, 2999–3015.

Dong, W., and Olson, E. S. (2006). “Tuning and travel of two tone distortion in intracochlear pressure,” in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer, (World Scientific, Singapore), pp. 56–62.

Fahey, P. F., and Allen, J. B. (1985). “Nonlinear phenomena as observed in the ear canal and at the auditory nerve,” *J. Acoust. Soc. Am.* **77**, 599–612.

Fahey, P. F., Stagner, B. B., and Martin, G. K. (2006). “Mechanism for bandpass frequency characteristic in distortion product otoacoustic emission generation,” *J. Acoust. Soc. Am.* **119**, 991–996.

- Goldstein, J. L. (1967). "Auditory nonlinearity," *J. Acoust. Soc. Am.* **41**, 676–689.
- Goldstein, J. L., Buschsbaum, G., and Furst, M. (1978). "Compatibility between psychophysical and physiological measurements of aural combination tones," *J. Acoust. Soc. Am.* **63**, 474–485.
- Kanis, L. J., and de Boer, E. (1993). "The emperor's new clothes: DP emissions in a locally active nonlinear model of the cochlea," in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 304–314.
- Kemp, D. T., and Knight, R. (1999). "Virtual DP reflector explains DPOAE 'wave' and 'place' fixed dichotomy," *Assoc. Res. Otolaryngol. Abs.* **22**, 396.
- Kim, D. O., Molnar, C. E., and Matthews, J. W. (1980). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure," *J. Acoust. Soc. Am.* **67**, 1704–1721.
- Kim, D. O., Siegel, J. H., and Molnar, C. E. (1979). "Cochlear nonlinear phenomena in two-tone responses," *Scand. Audiol. Suppl.* **9**, 63–81.
- Knight, R. D., and Kemp, D. T. (2000). "Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 area study," *J. Acoust. Soc. Am.* **107**, 457–473.
- Lim, K. M., and Steele, C. R. (2002). "A three-dimensional nonlinear active cochlear model analyzed by the WKB-numeric method," *Hear. Res.* **170**, 190–205.
- Lynch, T. J., Nedzelnitsky, V., and Peake, W. T. (1982). "Input impedance of the cochlea in cat," *J. Acoust. Soc. Am.* **72**, 108–130.
- Mammano, F., and Nobili, R. (1993). "Biophysics of the cochlea: Linear approximation," *J. Acoust. Soc. Am.* **93**, 3320–3332.
- Nobili, R., and Mammano, F. (1996). "Biophysics of the cochlea. II: Stationary nonlinear phenomenology," *J. Acoust. Soc. Am.* **99**, 2244–2255.
- Nobili, R., Vetešník, A., Turicchia, L., and Mammano, F. (2003). "Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model," *J. Assoc. Res. Otolaryngol.* **4**, 478–494.
- Olson, E. S. (2001). "Intracochlear pressure measurements related to cochlear tuning," *J. Acoust. Soc. Am.* **110**, 349–367.
- Pang, X. D., and Guinan, J. J. (1997). "Growth rate of simultaneous masking in cat auditory-nerve fibers: Relationship to the growth of basilar-membrane motion and the origin of two-tone suppression," *J. Acoust. Soc. Am.* **102**, 3564–3575.
- Peterson, L. C., and Bogert, B. P. (1950). "A dynamical theory of the cochlea," *J. Acoust. Soc. Am.* **22**, 369–381.
- Ren, T. (2004). "Reverse propagation of sound in the gerbil cochlea," *Nat. Neurosci.* **7**, 333–334.
- Ren, T., He, W., and Nuttall, A. L. (2006). "Backward propagation of otoacoustic emissions in the cochlea," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 79–85.
- Ren, T., and Nuttall, A. L. (2006). "Cochlear compression wave: An implication of the Allen–Fahey experiment," *J. Acoust. Soc. Am.* **119**, 1940–1942.
- Robles, L., and Ruggero, M. A. (2001). "Mechanics of the mammalian cochlea," *Pestic. Sci.* **81**, 1305–1352.
- Robles, L., Ruggero, M. A., and Rich, N. C. (1997). "Two-tone distortion on the basilar membrane of the chinchilla cochlea," *J. Neurophysiol.* **77**, 2385–2399.
- Ruggero, M. A. (2004). "Comparison of group delays of $2f_1-f_2$ distortion product otoacoustic emissions and cochlear travel times," *Acoust. Res. Lett. Online* **5**, 143–147.
- Shera, C. A. (2003). "Wave interference in the generation of reflection- and distortion-source emissions," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer, (World Scientific, Singapore), 439–453.
- Shera, C. A., and Guinan, J. J. (1997). "Measuring cochlear amplification and nonlinearity using distortion-product otoacoustic emissions as a calibrated intracochlear sound source," *Assoc. Res. Otolaryngol. Abs.* **20**, 51.
- Shera, C. A., and Guinan, J. J. (2007a). "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive calibration of intracochlear distortion sources," *J. Acoust. Soc. Am.* **121**, 1003–1016.
- Shera, C. A., and Guinan, J. J. (2007b). "Mechanisms of mammalian otoacoustic emission," in *Active Processes and Otoacoustic Emissions*, edited by G. A. Manley, B. L. Lonsbury-Martin A. N. Popper, and R. R. Fay, in press (Springer-Verlag, New York).
- Shera, C. A., Tubis, A., and Talmadge, C. L. (2004). "Are there forward and reverse traveling waves in the cochlea? Countering the critique of Nobili et al.," *J. Assoc. Res. Otolaryngol.* **5**, 349–359, [A version correcting JARO's numerous typesetting errors is available from the authors upon request (see also <http://mit.edu/apg>)].
- Shera, C. A., Tubis, A., and Talmadge, C. L. (2005). "Coherent-reflection models of reflection-source OAEs with and without slow transverse retrograde waves," *Assoc. Res. Otolaryngol. Abs.* **28**, 657.
- Shera, C. A., Tubis, A., and Talmadge, C. L. (2006). "Four counterarguments for slow-wave OAEs," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 449–457.
- Shera, C. A., and Zweig, G. (1991). "Phenomenological characterization of ear drum transduction," *J. Acoust. Soc. Am.* **90**, 253–262.
- Shera, C. A., and Zweig, G. (1992a). "Analyzing reverse middle-ear transmission: Noninvasive Gedankenexperiments," *J. Acoust. Soc. Am.* **92**, 1371–1381.
- Shera, C. A., and Zweig, G. (1992b). "Middle-ear phenomenology: The view from the three windows," *J. Acoust. Soc. Am.* **92**, 1356–1370.
- Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (2005). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," *J. Acoust. Soc. Am.* **118**, 2434–2443.
- Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (1998). "Modeling otoacoustic emission and hearing threshold fine structures," *J. Acoust. Soc. Am.* **104**, 1517–1543.
- Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (2000). "Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure," *J. Acoust. Soc. Am.* **108**, 2911–2932.
- van der Heijden, M., and Joris, P. X. (2003). "Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies," *J. Neurosci.* **23**, 9194–9198.
- van Hengel, P.W.J. (1996). "Emissions from cochlear modelling," Ph.D. thesis, Rijksuniversiteit Groningen.
- Vetešník, A., Nobili, R., and Gummer, A. (2006). "How does the inner ear generate distortion product otoacoustic emissions? Results from a realistic model of the human cochlea," *J. Otorhinolaryngol. Relat. Spec.* **68**, 347–352.
- Voss, S. E., Rosowski, J. J., and Peake, W. T. (1996). "Is the pressure difference between the oval and round window the effective stimulus for the cochlea?," *J. Acoust. Soc. Am.* **100**, 1602–1616.
- White, T. H. (1958). *The Once and Future King* (Putnam, New York).
- Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (1993). "Measurement of $2f_1-f_2$ excitation at the distortion-frequency place in the cochlea using ear-canal distortion products," *Assoc. Res. Otolaryngol. Abs.* **16**, 395.
- Wilson, J. P. (1980a). "The combination tone, $2f_1-f_2$, in psychophysics and ear-canal recording," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft), pp. 43–50.
- Wilson, J. P. (1980b). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," *Hear. Res.* **2**, 527–532.
- Zweig, G. (1976). "Basilar membrane motion," in *Cold Spring Harbor Symposium on Quantitative Biology, Vol. XL, 1975*, (Cold Spring Harbor Laboratory, Cold Spring Harbor), pp. 619–633.
- Zweig, G., (1991). "Finding the impedance of the organ of Corti," *J. Acoust. Soc. Am.* **89**, 1229–1254.
- Zweig, G., and Shera, C. A. (1995). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," *J. Acoust. Soc. Am.* **98**, 2018–2047.
- Zwicker, E., and Harris, F. P. (1990). "Psychoacoustical and ear canal cancellation of ($2f_1-f_2$)-distortion products," *J. Acoust. Soc. Am.* **87**, 2583–2591.

Examination of bone-conducted transmission from sound field excitation measured by thresholds, ear-canal sound pressure, and skull vibrations

Sabine Reinfeldt, Stefan Stenfelt,^{a)} Tobias Good, and Bo Håkansson

Department of Signals and Systems, Division of Biomedical Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

(Received 14 August 2006; revised 22 December 2006; accepted 25 December 2006)

Bone conduction (BC) relative to air conduction (AC) sound field sensitivity is here defined as the perceived difference between a sound field transmitted to the ear by BC and by AC. Previous investigations of BC-AC sound field sensitivity have used different estimation methods and report estimates that vary by up to 20 dB at some frequencies. In this study, the BC-AC sound field sensitivity was investigated by hearing threshold shifts, ear canal sound pressure measurements, and skull bone vibrations measured with an accelerometer. The vibration measurement produced valid estimates at 400 Hz and below, the threshold shifts produced valid estimates at 500 Hz and above, while the ear canal sound pressure measurements were found erroneous for estimating the BC-AC sound field sensitivity. The BC-AC sound field sensitivity is proposed, by combining the present result with others, as frequency independent at 50 to 60 dB at frequencies up to 900 Hz. At higher frequencies, it is frequency dependent with minima of 40 to 50 dB at 2 and 8 kHz, and a maximum of 50 to 60 dB at 4 kHz. The BC-AC sound field sensitivity is the theoretical limit of maximum attenuation achievable with ordinary hearing protection devices. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434762]

PACS number(s): 43.64.Bt, 43.66.Vt, 43.50.Hg [BLM]

Pages: 1576–1587

I. INTRODUCTION

A hearing sensation is produced by sound vibrations transmitted to the cochlea in three main ways. The first, air conduction (AC), is the most dominant one for ordinary hearing in a sound field; this is the transmission of sound to the cochlea via the ear canal, the eardrum, and the middle ear ossicles. The vibrations in the air also induce vibrations in the skull bone, which are transmitted in several ways to the cochlea. This second form of transmission is termed bone conduction (BC). A third kind of transmission, called body conduction, occurs when vibrations are induced in the body and transmitted to the cochlea via the soft tissues, skeleton, and fluids. In practice, it is difficult to differentiate the latter two kinds without isolating the body or the head. Here, both are referred to as bone conduction (BC).

The question now is, for a person in a sound field, how much of the sound that the person hears is transmitted by BC and how much by AC? Rephrasing this: How much of the sound energy in the cochlea is transmitted by BC versus AC, when the stimulation is a sound field? This ratio, i.e., the sound energy transmitted by BC relative to that of AC from a sound field to the cochlea, is hereafter referred to as the bone conduction relative to air conduction sound field sensitivity (BC-AC sound field sensitivity). It should be noted that we define this ratio for an open ear canal; the BC sensitivity alteration of occluding the ear canal (occlusion effect) is not included in the BC-AC sound field sensitivity.

The BC-AC sound field sensitivity sets the theoretical attenuation limit for ordinary hearing protection devices (HPDs), such as earplugs and earmuffs. Although the AC sound is attenuated by wearing ordinary HPDs, the perceived attenuation is also influenced by the BC contribution to the sound. Above a certain level of attenuation of the AC sound, the perceived sound is dominated by BC sound; an increase in the AC attenuation above this level is not perceived by the listener. This phenomenon implies that there is a maximum attenuation limit for HPDs. Thus, the attenuation limit equals the ratio of sound energy transmitted by BC relative to that by AC from a sound field to the cochlea, the BC-AC sound field sensitivity.

When the hearing threshold is obtained for a person in a sound field, this is defined as the AC sound field hearing threshold. After obstructing the ear canals by very efficient HPDs, to the extent that the AC transmission is totally obstructed, the sound field hearing threshold is obtained again; this time it is designated the BC sound field hearing threshold. Since the occlusion effect appears when the ear canals are blocked (the BC transmitted sound is stronger at low frequencies), the BC sound field hearing threshold must be adjusted accordingly. The BC sound field hearing threshold relative to the AC threshold is, then, the amount of sound that the person hears from BC relative to AC in a sound field. This is the BC-AC sound field sensitivity.

Several researchers (Zwislocki, 1957; Nixon and von Gierke, 1959; Schroeter and Poesselt, 1986; Berger, 1983; Berger *et al.*, 2003; Ravicz and Melcher, 2001) have already investigated the ratio between the BC and AC parts of the sound perceived from a sound field. Their aim was to exam-

^{a)}Current address: Department of Neuroscience and Locomotion, Division of Technical Audiology, Linköping University, SE-581 85 Linköping, Sweden.

ine the sound attenuation from different HPDs; in some cases it was also to determine the BC-AC sound field sensitivity itself. However, there are discrepancies in their results. At some frequencies, their estimates of the BC-AC sound field sensitivity vary by up to 20 dB. Berger (1983) termed the BC-AC sound field sensitivity “the BC limit.” However, there is a slight, but important, difference between our BC-AC sound field sensitivity and Berger’s BC limit; we exclude the BC sensitivity alteration caused by the HPD (occlusion effect) whereas Berger determined the BC-AC sound field sensitivity with the HPD in place and, by that, included the occlusion effect. None should be considered more correct; Berger’s BC limit is the actual attenuation achievable with that specific HPD fitted in that specific way (it is the attenuation experienced by a person wearing that specific HPD, or combination of HPDs, in that specific way). One should be able to calculate the BC limit from the BC-AC sound field sensitivity knowing the occlusion effect produced by the HPD or HPDs; if they do not produce any occlusion effect, the BC limit should equal the BC-AC sound field sensitivity.

To achieve an estimate of the BC-AC sound field sensitivity, three methods are used in the present investigation: real ear attenuation at threshold (REAT), microphone in real ear (MIRE), and vibration measurements of the skull bone with an accelerometer. The REAT method is a comparison of the sound field hearing thresholds with an open ear and with different HPDs. The MIRE method is used to find the difference in the sound pressure level, i.e., insertion loss, close to the eardrum with an open ear canal and with HPDs. Except that MIRE is usually obtained by sound pressure measurements at the ear canal entrance, the two measurement methods mentioned above are standard techniques to evaluate HPDs (Berger, 1986).

In the skull bone vibration method, the vibration level of the skull bone at the AC sound field hearing threshold is compared with the skull bone vibration level at the BC hearing threshold (stimulation by a BC transducer); the difference is an estimate of the BC-AC sound field sensitivity. Although the REAT and MIRE methods have been used in several investigations, to our knowledge, measurements of the skull bone vibrations to estimate the BC-AC sound field sensitivity have been used only by von Békésy (1948). Since these methods may be applicable to various frequency ranges and give different estimates, a combination and comparison of them should produce a good estimate of the BC-AC sound field sensitivity.

An area where the BC-AC sound field sensitivity provides valuable information is in design of BC microphones and more general BC communication systems (Ono, 1977; Zheng *et al.*, 2003). It is generally accepted that a BC microphone produces a better signal-to-noise ratio (SNR) than an AC microphone in front of the mouth in a surrounding of high level noise (Oyer, 1955; Black, 1957). It has been estimated that almost equal amounts of the perceived sound of one’s own voice are transmitted by AC and BC (von Békésy, 1949; Pörschmann, 2000). If the AC and BC parts of one’s own voice are of equal proportions, the AC and BC microphones record the same amount of the voice. The SNR is

then determined by the amount of the surrounding noise that is recorded by the microphones. Provided airborne noise and no other vibration coupling to the skull, the noise vibration in the skull bones can be estimated by the BC-AC sound field sensitivity. As a consequence, the SNR improvement using a BC microphone compared with an AC microphone can be approximated by the BC-AC sound field sensitivity. It should be noted that using the skull bone vibrations to record one’s own voice with a BC microphone is only expected to have merits in environments with high noise levels.

II. MATERIALS AND METHODS

A. Subjects

Fifteen normal hearing subjects (nine male and six female) with an average age of 29 years (24 to 36 years) participated in all measurements in the study. They all had otologically normal ears and, with the exception of two subjects who had hearing thresholds of up to 40 dB HL in one ear and at one frequency, their hearing thresholds were no worse than 25 dB HL in any ear in the frequency range 125 Hz to 8 kHz.

B. Measurement setup

All measurements were made in a sound insulated room of 16 m³. The hearing thresholds, either with TDH39 headphones, in a sound field, or with BC stimulation, were obtained with a digital audiometer (Interacoustics AC40). For the baseline audiometry, the output from the audiometer was fed to the TDH39 headphones, and the subjects’ thresholds were obtained using the Hughson-Westlake procedure in accordance with ISO 8253-1 (1989). The sound field and BC thresholds were obtained with a pulsed Békésy procedure (ISO 8253-1, 1989). For the sound field hearing thresholds, the output from the audiometer was fed to a loudspeaker (Bose Interaudio XL) placed 1 m in front of the test subject. Finally, when testing BC thresholds, the output from the audiometer was fed to a BC transducer (Radioear B71) placed at the center of the forehead.

The setup for measuring the ear canal sound pressure close to the eardrum is depicted in Fig. 1. The ear canal sound pressure was measured with a four-channel input, two-channel output Brüel & Kjær Pulse ver. 9.0 system (measurement and analysis system primarily for sound and vibration signals). The output from the measurement system was fed to a BC transducer (Radioear B71) or a loudspeaker (QLN Acoustics Qubic) by a power amplifier (Sony TA-N220 or Rotel RB-976MkII, respectively). It should be noted that due to practical reasons, the audiometer testing utilized a different loudspeaker than when the testing was conducted using the Pulse measurement system; both loudspeakers were placed symmetrically at the frontal axis and the deviation from this axis was bounded by 5°. The output from the amplifiers was used as input to the measurement system together with the signal from the probe tube microphone (Ety-motic Research ER-7C), which represents the ear canal sound pressure. The measurement system produced a frequency response function of input 2 (probe tube microphone signal) divided by input 1 (input to BC transducer or loud-

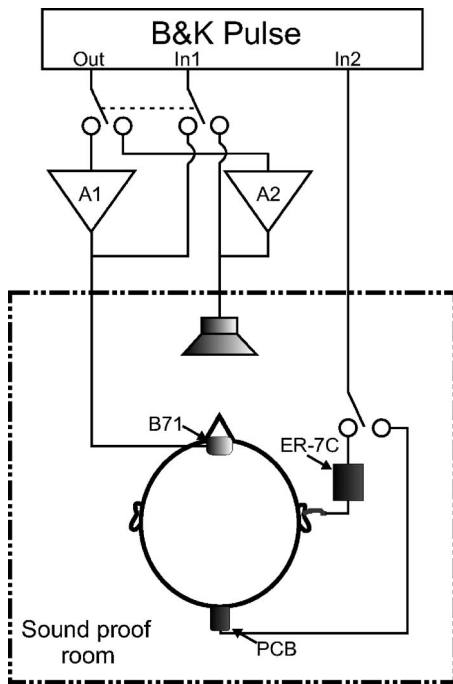


FIG. 1. Setup for the MIRE and vibration measurements. A stepped sine signal was fed from the Brüel & Kjær Pulse system (B&K Pulse) output (Out) to either a BC transducer (B71) or a loudspeaker, via power amplifier A1 or A2, respectively. The amplifier output was used as input 1 (In1) for the Pulse measurement system, while input 2 (In2) was the signal from either the probe microphone (ER-7C) for the MIRE measurements or the accelerometer (PCB) for the vibration measurements. The dash-dotted square indicates the sound insulated room.

speaker). The person equipped with the BC transducer and the probe tube microphone, together with the loudspeaker, was placed in the sound insulated room; the rest of the equipment was outside the room.

For the vibration measurements, the setup was similar to that for the ear canal sound pressure measurements, except that instead of the microphone an accelerometer (PCB Piezotronics 353B66) was used. This setup is also described in Fig. 1. The frequency response function was, then, the accelerometer signal divided by the signal to either the loudspeaker or the BC transducer.

C. The real ear attenuation at threshold (REAT) procedure

When the real ear attenuation at threshold (REAT) was measured, the person was placed in a chair in the sound insulated room. To keep the position of the head stable, the chair was fitted with a foam covered neck-rest. The test subject was given a button and the instructions to keep it pushed down when a pulsed tone was heard, otherwise to release it. Using this von Békésy procedure, the sound field hearing thresholds in the frequency range of 125 Hz to 8 kHz were obtained for the following four conditions:

- (1) The ear canal was open.
- (2) A pair of earplugs (EAR Classics) was bilaterally inserted 18 mm into the ear canals (relative to the tip of the tragus).¹
- (3) With the plugs still inserted, a pair of circumaural ear-

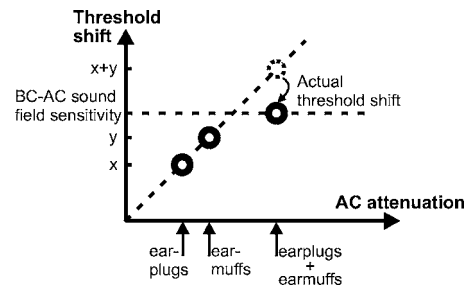


FIG. 2. Relation between AC attenuation and sound field hearing threshold shift. The hypothesis is that an increase of AC attenuation (by using combinations of HPDs) is reflected by the same increase in threshold shift (diagonal line) until the BC-AC sound field sensitivity limit is reached (horizontal line). x is the threshold shift from the earplug, y is the threshold shift from the earmuff, and the calculated sum of these is shown as a dashed circle ($x+y$). The BC-AC sound field sensitivity is estimated as the actual threshold shift with the combination of earplugs and earmuffs if, and only if, the sum of their separate threshold shifts ($x+y$) exceeds the threshold shift obtained with the combination.

muffs (Peltor OptimeI H510B) with a volume of approximately 30 cm³ was fitted, i.e., both plugs and earmuffs were used in this test.

- (4) The plugs were removed, leaving only the earmuffs in place.

In other words, the sound field hearing thresholds were obtained with an open ear canal, using only earplugs, using only earmuffs, and using both earplugs and earmuffs. The REAT was then calculated as the difference between the sound field hearing thresholds obtained with HPDs and with an open ear canal. Consequently, the REAT for the earplugs was the hearing threshold with earplugs in dB HL minus the open ear hearing threshold in dB HL. Then the REATs for earmuffs and for the combination of earmuffs and earplugs were obtained similarly.

The hypothesis is that the attenuation of the sound field by the HPDs corresponds to a threshold shift equal to the attenuation. Thus, if the earplugs cause a threshold shift of x dB and the earmuffs cause a threshold shift of y dB, using both earmuffs and earplugs can be expected to cause a threshold shift of $x+y$ dB. However, if the attenuation of a single HPD or a combination of HPDs is large enough, the AC path can be attenuated to the point at which the BC path becomes dominant. This inference is illustrated in Fig. 2. The perceived hearing from the sound field is transmitted by BC; the threshold shift is an estimate of the BC-AC sound field sensitivity. According to the measurement procedure, the BC-AC sound field sensitivity can be estimated as the threshold shift for the combination of earplugs and earmuffs if, and only if, the threshold shift with the combination of earplugs and earmuffs is less than the sum of the separate threshold shifts with earplugs and with earmuffs.

The controversy of this hypothesis is that the attenuation of the earmuff and earplug is considered independent and that they add linearly. There are primarily two mechanisms that would introduce dependent and nonlinear attenuation: (1) alteration of the load impedances seen from the HPDs and (2) mechanical coupling between the HPDs. Ravicz and Melcher (2001) found the sum of earplug and earmuff at-

tenuation to be similar to the attenuation of earplug and earmuff combination for frequencies below 500 Hz, i.e., the attenuation of the earplug and earmuff could be added independently. They further argue that at higher frequencies the interaction between earmuffs and earplugs would be negligible since the radiation impedance is close to the impedance of the volume under the earmuff at frequencies above 1 kHz (Schroeter and Poesselt, 1986) and the volume of the earplug is only a few percent of the total volume under the ear muff (5% in this study). Further, due to the spring coupling between the earmuff and the skull, the earmuff is vibrationally decoupled from the skull at higher frequencies. As a consequence, we assume that the attenuation of the earmuffs and earplugs can be added linearly.

D. The microphone in real ear (MIRE) procedure

The procedure for the MIRE measurements was similar to the REAT procedure just described (in Sec. II C). The difference was that, instead of the person's subjective sound field hearing threshold, the sound pressure in the ear canal was measured. The four test conditions described for the REAT measurements were used for the MIRE measurements as well: open ear canal, earplugs, earmuffs, and earplugs and earmuffs combined. In reality, both the REAT and MIRE measurements were obtained for each condition before the HPD(s) were changed.

The ear canal sound pressure was measured using a stepped sine procedure in the frequency range 0.1 to 10 kHz with a frequency resolution of 12 points/octave. A third-octave band analysis was made by integrating the sound energy in each third-octave band; the resulting MIRE data were presented in the frequency range 125 Hz to 8 kHz. All ear canal sound pressure measurements were made with the probe tube microphone inserted 25 mm into the ear canal, measured from the tip of the tragus, for all test subjects. With the ear canal open or when only earmuffs were used, a mark on the microphone probe tube served to indicate the insertion depth. The foam earplug had a probe tube positioned through a perforation made by a needle catheter. The foam material of the earplug made the fit of the microphone probe tube tight, which prevented sound leakage. Moreover, since the insertion depth of the earplugs was controlled, the insertion depth of the microphone probe tube was the same when earplugs were used as when the ear canal was open.

When the earplugs were used, the sound pressure at the microphone probe tube opening could be up to 40 dB lower than at the outside of the earplug. This means there is a risk that if sound is leaking through the microphone silicone probe tube, the result can be affected. To improve the sound insulation of the microphone probe tube, the part of the tube from the microphone casing to the foam plug was covered with an extra silicone tube with an internal diameter of 2 mm. To avoid sound leakage between the microphone electrical wire and the earmuffs, the wire was fed through a small hole drilled in the plastic of the earmuff; this hole was sealed with epoxy glue. Even with these precautions, the results were affected by sound leakage at frequencies above 5 kHz.²

As in the REAT procedure, the BC-AC sound field sen-

sitivity was estimated from the measurement with the combination of earplugs and earmuffs. If the attenuation obtained with the combination was less than the sum of the separate attenuations of earplugs and earmuffs, the sound pressure shift with the combination of HPDs was taken as an estimate of the BC-AC sound field sensitivity.

E. The skull vibration measurements procedure

An accelerometer (PCB 353B66) was attached to a plastic plate with a circular area of 113 mm². The plastic plate was pressed against the head at the center of the occiput (back of the head) with an elastic head band. The hypothesis is that the vibration of the skull bone mimics the vibration of the cochlea. To reduce the influence of the skin, the plastic plate was pressed against the head with a static force of approximately 4 N.³ The stimulation for the vibration measurement was either a sound field, produced by the loudspeaker, or a BC sound, produced by the BC transducer at the center of the forehead. The stimulation signal was the same as with the MIRE measurement, and the resulting data, the vibration of the skull bone relative to the stimulation, was analyzed in third-octave bands in the frequency range of 125 Hz to 8 kHz.

The measurements were repeated for two other positions of the accelerometer: the right and the left temporal bones 30 to 40 mm superior to the ear canal opening, close to the pinna. The resulting vibration levels were related to each subject's individual AC and BC hearing thresholds. For example, if the subject's sound field hearing threshold was 20 dB SPL at a given frequency, and the sound field stimulation level at the same frequency was 90 dB SPL during the vibration measurement (accelerometer), the stimulation level was 70 dB above threshold. Consequently, the vibration level was 70 dB higher than at the sound field hearing threshold and the result was shifted 70 dB to give the skull vibration level at threshold, assuming linearity. Similar computations were done for the data obtained with BC stimulation. These two results were compared, i.e., the skull vibrations at the BC threshold relative to the skull vibrations at the sound field threshold. The ratio was used as an estimate of the BC-AC sound field sensitivity.

This method is not valid at higher frequencies, since the compliance of the skin together with the mass of the circular plate and accelerometer form a mass-spring system shunting the vibrations at high enough frequencies. Flottorp and Solberg (1976) estimated the skin compliance to be around 4.7×10^{-6} m/N for an area and static force close to that used here. Due to the mass of the accelerometer and plastic plate (4.83 g), the upper frequency limit of the method is around 1 kHz. Consequently, the results from the vibration measurements are valid only for frequencies below 1 kHz.

von Békésy (1948), Håkansson *et al.* (1994), and Stenfelt and Goode (2005a), have shown that the head vibrates as a rigid body at low frequencies, which means that when a BC transducer stimulates the forehead, the accelerometer at the occiput position measures in the stimulation direction (the accelerometer only measures in one direction); however, on the sides of the head, the accelerometer measures in a

direction perpendicular to the stimulation direction. It was shown by Stenfelt and Goode (2005a), that the low-frequency skull vibration level could be up to 20 dB lower in the perpendicular plane than in the stimulation direction. This means that the vibration measurements at the sides of the head are not suitable at the low frequencies where the skull exhibits primarily rigid body motion.

In addition, Stenfelt and Goode (2005a), showed that, the skull moves with true rigid body motion up to frequencies of 300 to 400 Hz. Above this frequency the stiffness of the skull affects the motion, and the vibration at the occiput does not mimic the vibration at the cochleae. At even higher frequencies, above the first skull resonance at around 1 kHz (Håkansson *et al.*, 1994), the skull vibrates in all three space dimensions without any dominant direction. Although the vibrations from the sides of the head (temporal bones) would be a good estimate of the vibration transmitted to the cochleae at these higher frequencies, the skin attenuates the vibration measured by the accelerometers, which is why this method is not valid at these high frequencies. Hence, an estimate of the BC-AC sound field sensitivity is equal to the vibration result at the occiput position for frequencies up to 400 Hz.

F. The occlusion effect

BC sound is transmitted to the inner ear and stimulates the basilar via several routes (for an overview, see Stenfelt and Goode, 2005b). One of the transmission routes is BC vibration that causes the ear canal walls to vibrate and produces a sound pressure in the ear canal that is transmitted to the inner ear via the eardrum and middle ear ossicles. Sound pressure in the external ear canal due to BC stimuli can be elevated at frequencies below 2 kHz when the opening of the external ear canal is occluded; this is called the occlusion effect. Huizing (1960) explained the occlusion effect by a change in the resonance properties of the ear canal: a closed tube has resonance properties that differ from those of an open tube. Tonndorf (1966) presented another explanation: The mass effect of the air column in the ear canal, together with the compliance of the air in the ear canal and the tympanic membrane, function as a high-pass filter on the ear canal sound pressure. When the ear canal is occluded, the high-pass filter effect is more or less eliminated, which results in an increase of low frequency sound. Tonndorf's (1966) explanation is correct at low frequencies, where the mass and compliance of the ear canal air determine the acoustic properties, whereas Huizing's (1960) explanation is correct above 2 kHz, where resonances and anti-resonances determine the acoustic properties of the ear canal. Since HPDs were used in the present investigation to attenuate the AC sound, the ear canal was occluded, hence producing an occlusion effect. This occlusion effect had to be taken into account. The REAT and MIRE data were affected not only by the damping of the HPD used, but also by the enhancement of BC sound due to the occlusion effect. This means that if no compensation were made, the BC-AC sound field sensitivity based on the REAT and MIRE results would be underestimated.

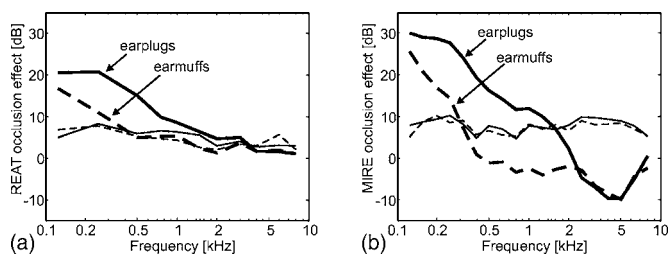


FIG. 3. Average of the measured occlusion effect with earplugs (solid line) and earmuffs (dashed line) from (a) sound field hearing threshold shifts (REAT) and (b) from ear canal sound pressure shifts (MIRE). The standard deviations are shown in thinner lines.

The occlusion effect was estimated for each subject with each HPD (earplug and earmuff). As described in Sec. II B, the hearing threshold with BC stimulation applied at the forehead was obtained with the ear canals open. These hearing thresholds were also measured with the earplugs or earmuffs applied. The alteration of the threshold data was used as a measure of the occlusion effect and also for the compensation of the REAT data. The MIRE data were compensated similarly for the occlusion effect as follows. The ear canal sound pressure was measured with a probe tube microphone (see Sec. II D) when the stimulation was a BC vibration at the forehead; the ear canals were open, occluded with earplugs, or occluded with earmuffs. These measurements gave the occlusion effect for the MIRE data and the results were compensated accordingly.

The average results of the occlusion effect measurements and their standard deviations are shown in Fig. 3. It should be noted that the occlusion effects measured with the REAT and MIRE procedures differ. For frequencies below 2 kHz, the occlusion effect from the ear canal sound pressure (MIRE) was 5 to 10 dB greater than the occlusion effect from the threshold data (REAT). This is due to the different BC pathways to the cochlea. With the ear canal open, the sound radiated in the ear canal is not the dominating contributor for BC hearing. According to Huizing (1960) and Stenfelt *et al.* (2003), the part of the BC sound that radiates into the ear canal is about 10 dB below the dominant BC pathway at frequencies below 1 kHz. When the ear canal is occluded, the perception of BC sound, for frequencies up to at least 1 kHz, is dominated by the outer ear contribution. As a consequence, the occlusion effect for the MIRE result, which is obtained by the ear canal sound pressure, is approximately 10 dB greater than for the REAT results that is obtained by perceptual hearing thresholds. The standard deviations of the MIRE occlusion effect are slightly greater than for the REAT occlusion effect.

Since the measurements of skull vibrations did not involve an occluded ear canal, the occlusion effect does not affect those data and no corrections of the accelerometer measurements were required.

G. Calibrations

The sound field in the sound attenuated room, either from the audiometer or the Brüel & Kjær Pulse system, was calibrated using a Brüel & Kjær type 4134 $\frac{1}{2}$ -in. microphone.

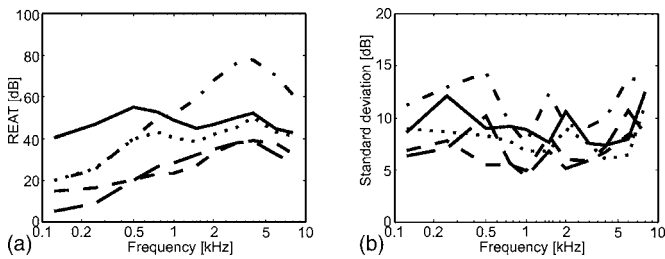


FIG. 4. Results from the measurements using hearing threshold shifts (REAT method); (a) shows the averages and (b) shows the standard deviations. Dashed line: earplugs; long dashed line: earmuffs; dotted line: combination of earplugs and earmuffs; dash-dotted line: calculated sum of earplugs and earmuffs. The solid line is the combination (of earplugs and earmuffs), which has been compensated for the occlusion effect presented in Fig. 3(b).

According to the manufacturer, this microphone has a maximum level deviation of 2 dB in the frequency range 0.1 to 10 kHz. First, the sensitivity of the microphone was determined with a Brüel & Kjør type 4230 sound level calibrator. Then the $\frac{1}{2}$ -in. microphone was positioned approximately where the center of the head of a test subject would be during a test session, and the sound field was determined. The probe tube microphone (ER-7C) was also calibrated against the $\frac{1}{2}$ -in. microphone. The probe tube opening was placed 1 mm from the $\frac{1}{2}$ -in. microphone, a sound field was introduced, and the calibration curve of the ER-7C was obtained for the frequency range 0.1 to 10 kHz. The reference for the sound field hearing thresholds was the standard ISO 226 (2003).

The output from the B71 transducer was calibrated according to ISO 389-3 (1994). The transducer was placed on the Brüel & Kjør type 4930 artificial mastoid and the output from the mastoid was related to a dynamic force of the B71 transducer. The airborne sound radiated from the B71 transducer was measured to exclude any effect on the BC measurements with open ear canal. The ear canal sound pressure and the sound pressure 2 cm in front of the BC transducer position at the forehead were measured. Then, the same measures were taken with the BC transducer at approximately the same position but without touching the forehead. The measurements were equalized for the same amount of sound radiated from the B71 (measured by the microphone in front of the transducer) and the ear canal sound pressures measured by the probe tube microphone were compared. The sound radiated from the B71 transducer was more than 10 dB below the sound pressure generated by BC sound in the ear canal at all frequencies in the range 0.1 to 10 kHz.

III. RESULTS

A. The real ear attenuation at threshold (REAT)

The average results for the sound field hearing threshold shifts are shown in Fig. 4(a); the standard deviations of the data are depicted in Fig. 4(b). The threshold shifts, REAT, are given for the different HPDs: the threshold shift with earplugs is shown by a dashed line, the threshold shift with earmuffs is shown by a long-dashed line, the threshold shift with the combination of earplugs and earmuffs is shown by a dotted line, while the calculated sum of threshold shifts for

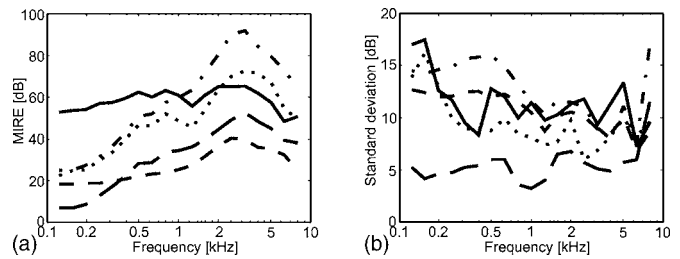


FIG. 5. Results from the measurements of the ear canal sound pressure (MIRE method); (a) shows the averages and (b) shows the standard deviations. Dashed line: earplugs; long dashed line: earmuffs; dotted line: combination of earplugs and earmuffs; dash-dotted line: calculated sum of earplugs and earmuffs. The solid line is the combination (of earplugs and earmuffs), which has been compensated for the occlusion effect presented in Fig. 3(b).

earplugs and earmuffs is shown by a dash-dotted line. According to the methodology, the BC-AC sound field sensitivity is reached only when the sum of the separate threshold shifts for earplugs and earmuffs is greater than the threshold shift for the combination of earplugs and earmuffs (without compensation for the occlusion effect). It can be seen in Fig. 4(a) that these two lines are close at frequencies at and below 500 Hz, while above this frequency the calculated sum (dash-dotted line) is greater than the combination of HPDs (dotted line). This indicates that the BC-AC sound field sensitivity limit was reached in the frequency range from above 0.5 to 8 kHz.

As illustrated previously (in Sec. II F), the occlusion effect influences the low-frequency BC sound; to estimate the true BC-AC sound field sensitivity, correction must be made for the occlusion effect. The solid line in Fig. 4(a) is the combination of earplugs and earmuffs compensated for the occlusion effect (Fig. 3); this represents the estimate of the BC-AC sound field sensitivity at frequencies above 500 Hz. The occlusion effect compensation is accomplished by an algebraic addition of the measured individual occlusion effect values (Fig. 3). According to the REAT approach, the BC-AC sound field sensitivity lies between 42 and 55 dB in the frequency range 0.5 to 8 kHz.

B. The microphone in real ear (MIRE)

The results of the ear canal sound pressure measurements (MIRE) given in Fig. 5(a) are similar to the REAT results in Fig. 4(a). Also, the standard deviations of the MIRE data [Fig. 5(b)] are similar to those obtained according to the REAT procedure. The sound pressure level difference, MIRE, is illustrated for the different HPDs: earplugs (dashed line), earmuffs (long-dashed line), combination of earplugs and earmuffs (dotted line), and the calculated sum for earplugs and earmuffs (dash-dotted line). The analysis of the MIRE results is also similar to that of the REAT results. The limit for reaching stimulation by BC is determined by comparing the calculated sum of separate earplug and earmuff attenuation and the attenuation from the combination of earmuffs and earplugs. As with the REAT data, the two lines differ at frequencies of 500 Hz and above.

The solid line in Fig. 5(a) is the combined earplug + earmuff data, which has been compensated for the occlu-

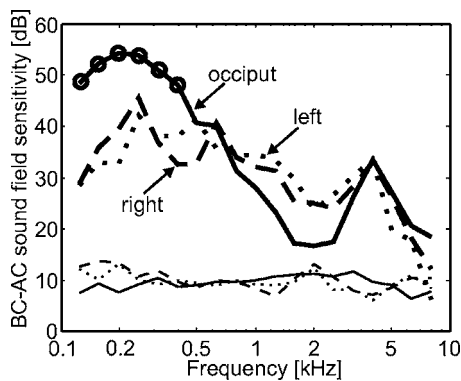


FIG. 6. The average BC-AC sound field sensitivity calculated from the skull vibration measurements using an accelerometer at three positions on the head. The thinner lines show the standard deviations. Solid line: accelerometer positioned at the center of occiput; dashed line: accelerometer positioned on the temporal bone, above the ear canal opening close to the pinna at the right side of the head; dotted line: same position as before but at the left side of the head. The circles indicate the position and the frequency range where the method is valid for the BC-AC sound field sensitivity estimation.

sion effect (Fig. 3). It should be noted here that the occlusion effect is not the same for the REAT and MIRE data. This means that in terms of ear canal sound pressure, the MIRE attenuation compensated for the occlusion effect gives an estimate of the BC-AC sound field sensitivity for frequencies of 500 Hz and above. However, as stated in Sec. II D, sound leakage through the walls of the probe microphone contaminates the result at 5 kHz and above. Consequently, the frequency range for the BC-AC sound field sensitivity estimate obtained by measuring the ear canal sound pressure (MIRE) is between 0.5 and 5 kHz, and it has values between 55 and 66 dB.

C. Skull bone vibration measurement

The skull bone vibration during AC and BC stimulation was measured with an accelerometer, at three positions on the head, and related to the individual hearing thresholds (without HPDs). The analysis of these measurements, displayed as the BC-AC sound field sensitivity, is plotted in Fig. 6 for three positions: the center of the occiput (solid line), the right temporal bone (dashed line), and the left temporal bone (dotted line). The standard deviations of the vibration measurements are included as thin lines. The BC-AC sound field sensitivities estimated from the temporal bone locations are similar, while the occiput result differs from the other two by 10 to 15 dB at the low frequencies (below 500 Hz). The low-frequency difference between the occiput and temporal bone results is caused by the vibration direction of the skull bone when stimulating at the center of the forehead (Sec. II E). Since the skull moves as a rigid body up to 300 to 400 Hz (Stenfelt and Goode, 2005a), and the accelerometer at the occiput measures in the vibration direction during BC transducer stimulation, the vibration obtained at the occiput is approximately the same as the vibration of the cochleae at these low frequencies. Hence, the result from the occiput up to 400 Hz can be used as an estimate of the BC-AC sound field sensitivity. This is shown in Fig. 6 as circles on the occiput line in the valid frequency range. Below 400 Hz, the

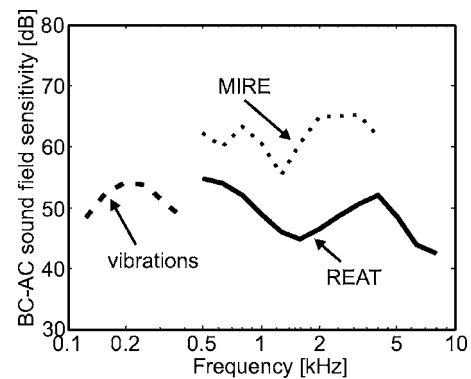


FIG. 7. Estimate of the BC-AC sound field sensitivity from the three methods in the frequency ranges where they produced valid results. Solid line: REAT; dotted line: MIRE; dashed line: vibrations. The final estimate of the BC-AC sound field sensitivity consists of the estimate from the vibration measurements and the REAT measurements, since the estimate from the MIRE measurements produced values that were too high (see text).

BC-AC sound field sensitivity is estimated to be 49 to 55 dB; above 400 Hz the method is not valid for the BC-AC sound field sensitivity estimation.

D. The BC relative to AC sound field sensitivity estimate

The estimates of the BC-AC sound field sensitivity based on the three methods are summarized in Fig. 7. The REAT estimate of the BC-AC sound field sensitivity is the solid line, and the MIRE estimate is the dotted line. According to the hypothesis that the BC-AC sound field sensitivity is reached when the sum of the separate earplug and earmuff data exceeds the combined data, the estimate from the MIRE measurements is valid in the frequency range 0.5 to 5 kHz, while the estimate from the REAT measurements is applicable at 0.5 to 8 kHz. It can be seen that the BC-AC sound field sensitivity estimate using the MIRE method produces 10 to 20 dB higher levels than the other methods. This divergence is attributed to the relative importance of the various contributors to BC sound (Stenfelt and Goode, 2005b); the ear canal component is not as effective as other contributing components.

It has been shown that the radiation in the ear canal, as a part of the BC sound, is 10 to 20 dB lower than other contributing parts (Huizing, 1960; Stenfelt *et al.*, 2003; Stenfelt, 2006). This means that when the attenuation from the HPDs have reached the limit for BC-AC sound field sensitivity and the BC path becomes dominant for the perceived sound, the sound pressure in the ear canal can be attenuated 10 to 20 dB more before the sound pressure in the ear canal becomes dominated by the BC sound radiated in the ear canal. As a consequence, the sound pressure in the ear canal (producing the MIRE estimate) can be reduced some 10 to 20 dB more than the measured threshold shift (producing the REAT data). Due to this, the MIRE produces a higher and biased estimate of the BC-AC sound field sensitivity. This problem was previously addressed by Berger (1986).

One may argue that the MIRE estimate presented here is meaningless since it cannot be used to derive the BC-AC sound field sensitivity. However, the MIRE measure serves

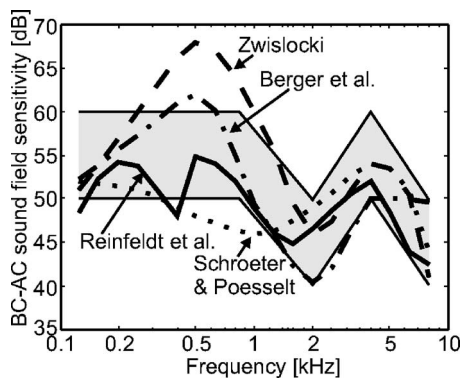


FIG. 8. Estimates of the BC-AC sound field sensitivity from this and some earlier investigations. Dashed line: Zwislowski (1957); dotted line: Schroeter and Poesselt (1986); dash-dotted line: Berger *et al.* (2003); solid line: current study. The gray area is the proposed BC-AC sound field sensitivity using the results from the current and above mentioned studies.

two purposes. First, it provides an upper bound on the BC-AC sound field sensitivity; since it is biased towards higher values, the BC-AC sound field sensitivity should stay below the MIRE estimate. Second, it gives a reality check of the BC-AC sound field estimate obtained by the other methods since we know that it should be some 10 to 20 dB below the MIRE data. It should be remembered that the bias in the MIRE data only occurs for high attenuation where the BC-AC sound field sensitivity is reached. For lower attenuations, that are normal for single application of HPDs, the MIRE results correspond to REAT data.

The estimate from the vibration method, the dashed line in Fig. 7, is reliable below 400 Hz; the vibration modes of the skull and the vibration attenuation caused by the skin produces unreliable data at higher frequencies (see Sec. II E).

The final proposed estimate of the BC-AC sound field sensitivity from this study is based on the estimate from the vibration measurements in the frequency range 125 to 400 Hz, and the estimate from the REAT measurements at frequencies between 0.5 and 8 kHz. There is a small gap between these frequency ranges. However, neither the vibration measurement nor the REAT data provide any values between 400 and 500 Hz. If the BC-AC sound field sensitivity at 400 Hz were estimated by linear interpolation from the REAT results at 250 and 500 Hz, the obtained value is slightly above the estimate given by the vibration data at 400 Hz. The BC-AC sound field sensitivity estimate is assumed to be continuous using the vibration measurement result at 400 Hz and the REAT estimate at 500 Hz.

The final proposed estimate of the BC-AC sound field sensitivity is presented as the solid line in Fig. 8; it varies between 48 and 55 dB at frequencies between 125 and 900 Hz. At 900 Hz the estimate is 50 dB, decreases to 45 dB at 1.6 kHz, increases to 52 dB at 4 kHz, and finally decreases to 42 dB at 8 kHz.

IV. DISCUSSION

The BC-AC sound field sensitivity has been addressed in previous studies. The most important results from these studies, together with those from the current study, are given in Fig. 8. In addition to the result from this study (solid line),

the data from Zwislowski (1957), Schroeter and Poesselt (1986), and Berger *et al.* (2003) are included.

A. Comparison with previous data

Often, the data from Zwislowski (1957) are used as the supreme standard for the BC-AC sound field sensitivity. However, there are some aspects of his report that should be investigated further. Zwislowski argued that the attenuation of multiple HPDs (e.g., using both earplugs and earmuffs) was influenced by the mechanical coupling of the HPDs. In his study, no increase of the attenuation was obtained when adding an earmuff to an earplug, but the low-frequency attenuation improved considerably when the earplug was positioned in the bony part of the ear canal. We suggest the following explanation of Zwislowski's findings: With the earplugs in the cartilage part of the ear canal, the BC transmission dominates the attenuation of the sound field; adding the earmuff produces no additional attenuation. When the earplug is inserted further into the bony part of the ear canal, the occlusion effect is reduced, which improves the low-frequency attenuation. The authors believe that the mechanical coupling of the earmuff and earplug, suggested as the dominant path of sound transmission by Zwislowski (1957), is of minor importance (as stated in Sec. II C).

Compared to results from other groups, Zwislowski (1957) obtained 5 to 10 dB greater values for the BC-AC sound field sensitivity at frequencies between 0.2 and 1.5 kHz. Zwislowski's data between 0.4 and 1.5 kHz were obtained from only three test persons using special resonant earplugs tuned to different resonance frequencies; only the highest attenuation obtained among the three subjects was used for the BC-AC sound field sensitivity estimate. In the same study he also applied ordinary earplugs together with heavy earmuffs; with these fitted, the low frequency attenuation was 5 to 15 dB lower than the maximum obtained with the resonant earplugs. The data of ordinary earplugs combined with heavy earmuffs are similar to other studies and also to the data presented here.

The method of using only the maximum level obtained in a small frequency range, which Zwislowski used, leads to slight overestimation of the normal attenuation; his result can be seen as the addition of the true attenuation and the uncertainties associated with the estimation method. Berger *et al.* (2003) noted that if they only used the results from a few subjects showing the greatest attenuations, they could achieve a BC-AC sound field sensitivity close to Zwislowski's results. A more reliable method is to use averages from several subjects, several tests, or both; this was done in the other investigations given in Fig. 8, as well as in the present one. Although not probable, the impedance change in the ear canal, caused by the resonant earplugs, might be reflected through the middle ear and reduce the BC sensitivity, thereby overestimating the BC-AC sound field sensitivity in the study by Zwislowski. Furthermore, an occlusion inserted in the bony part of the ear canal could have caused a static pressure difference over the tympanic membrane. Such a pressure difference was shown by Huizing (1960) to decrease the BC sensitivity by approximately 10 dB at frequen-

cies below 1 to 2 kHz. According to the above, the low frequency data presented by Zwislocki (1957) as the BC-AC sound field sensitivity are believed to be slightly too high.

Berger *et al.* (2003) used deep fitted foam earplugs and earmuffs with integrated active noise reduction (ANR) to estimate the BC-AC sound field sensitivity. They used a REAT procedure to obtain the estimate. Although the ANR was in its on-mode, which could give amplifier noise and therefore overestimate the BC-AC sound field sensitivity, the earplugs reduced it enough. Measurements were made with the ANR both on and off, and the result was only a small difference at 125 Hz, found to be insignificant. Therefore, it did not really matter which ANR setting was used, and they chose the estimate with ANR in its on-mode. Compared with the result of the current study, their data are slightly higher at 500 Hz. This is the frequency zone where our data are slightly uncertain; the REAT result is on the borderline of validity and the vibration data are not valid. It might be that around 500 Hz, the Berger *et al.* estimate of the BC-AC sound field sensitivity is more accurate than our estimate. Since they used deep fitted earplugs (down to the bony part of the ear canal), they did not compensate for the occlusion effect. Although deep fitting reduces the occlusion effect, some elevation of BC sound may still be present at the lowest frequencies (below 200 Hz), which may explain their slightly lower estimate at the lowest frequencies.

Schroeter and Poesselt (1986) reported data obtained with large custom designed earmuffs, with speakers incorporated for low frequency sound cancellation. This setup was used at frequencies below 2 kHz, whereas at and above 2 kHz they combined the attenuation box with deeply inserted earplugs. Below 1.2 kHz, their estimate is lower than the BC-AC sound field sensitivity presented here. The slightly lower estimate of low-frequency BC-AC sound field sensitivity made by Schroeter and Poesselt (1986) arises probably because they did not use any earplugs to attenuate the air borne sound. Although they achieved up to 70 dB attenuation by using an attenuation box with ANR, their reference position was 2 cm outside the ear canal opening and the attenuation achieved at the eardrum level could have been less than that at the reference microphone. This is indicated by the fact that they obtained the greatest attenuation at the lowest frequencies, where the sound pressure is the most uniform; at higher frequencies the wave length is shorter which means the ANR attenuation is worse.

The method of measuring the vibrations of the skull bone and calculating the amplitude of vibrations at the hearing thresholds to estimate the BC-AC sound field sensitivity has, as far as the authors know, been used only by von Békésy (1948). He estimated the BC-AC sound field sensitivity to be 48 to 53 dB at 200 to 800 Hz, which is similar or slightly lower than the one estimated by the current study.

By comparing results of the previous studies with those of the present one, the conclusion was drawn that the BC-AC sound field sensitivity probably lies in the grayed area of Fig. 8. It is relatively constant between 50 and 60 dB at frequencies up to 900 Hz. Then it decreases to 40 to 50 dB at 2 kHz, directly increases to 50 to 60 dB as a maximum at 4 kHz, and finally decreases to 40 to 50 dB at 8 kHz.

At the highest frequencies, between 4 and 8 kHz, our data as well as the data from Zwislocki and the data from Schroeter and Poesselt fall off from between 52 and 55 dB to between 41 and 45 dB. The exception is the Berger *et al.* data (dashed-dotted line in Fig. 8), which are almost constant at 50 dB in the same frequency range. The difference between the studies is that Berger *et al.* used diffuse sound field while the others used frontally incident sound. Poulsen (1987) measured the insertion loss on acoustic test fixtures with four different earmuffs using both frontally incident sound and diffuse sound field stimulation. That study reported around 5 dB greater attenuation for diffuse sound field compared with frontally incident stimulation in the 4 to 8 kHz frequency range with three of the four earmuffs used. Consequently, for a diffuse sound field the data of Berger *et al.* might be the best estimate of the BC-AC sound field sensitivity at the highest frequencies.

B. Estimation methods used

To achieve an estimate of the BC-AC sound field sensitivity for the whole frequency range of 125 Hz to 8 kHz, three estimation methods were combined. It was found that the vibration measurements method was applicable at frequencies between 100 and 400 Hz, while the threshold shifts method produced data applicable at frequencies between 0.5 and 8 kHz; the measurements of the ear canal sound pressures gave values for the BC-AC sound field sensitivity that were too high.

The method of threshold shifts could have been extended to lower frequencies by using high attenuation hearing protectors and, perhaps, also by better fitting of the earplugs and earmuffs. However, the method of threshold shifts (REAT) is biased by physiological noise at the very low frequencies. According to Berger and Kerivan (1983), this shift may differ by as much as 5 dB at 125 and 250 Hz, depending on what HPD is used. Due to the occlusion effect, noise produced by the body itself (such as heartbeat etc.) is enhanced, which masks at the lowest frequencies. Such physiological noise would produce worse hearing thresholds when a hearing protector is applied, especially when earplugs are used. That means that, even if the hearing protectors did have greater low-frequency attenuation enabling REAT measurements at lower frequencies, the REAT method would have produced an unreliable estimate of the BC-AC sound field sensitivity at the lowest frequencies. However, a deeper fit of the earplug would have reduced the occlusion effect as well as the affect of physiological noise. Consequently, a deeper fit of the earplug could have provided REAT data at lower frequencies. Due to the microphone probe tube penetrating the earplug, it was not possible to fit the earplug deeper. To obtain deeper earplug fitting, another method for ear canal sound pressure measurements is required.

One limitation of the vibration measurement method for the BC-AC sound field sensitivity stems from the soft tissue coupling between the accelerometer and the skull bones. We measured the vibration of the skull bone through the skin which, in itself, gives an upper frequency limit for reliable

data (in our case approximately 1 kHz). However, the vibration modes of the skull restricted the data to frequencies below 400 Hz for this type of method (the accelerometer measured only in one direction), while the choice of stimulation at the forehead was limited to vibration measured at the occiput. At these low frequencies the skull exhibits near-rigid body motion in the stimulation direction; thus the vibration at the occiput is close to that of the cochleae, and the AC and BC vibration thresholds can be compared.

To extend this method to higher frequencies, the vibration measurement should be made close to the cochleae, which is unsuitable for live human test subjects. One can argue that the vibration of the occiput and cochleae are not comparable when a sound field is used, since the head would be stimulated in all space dimensions. With sound field stimulation, the vibration was found to be similar at all three measurement positions, when comparing the absolute vibration levels at frequencies below 500 Hz (the difference seen in Fig. 6 is due to differences with BC stimulation at the forehead). If one adds all three vibration dimensions, the total vibration energy is increased by almost 5 dB compared with a one-dimensional measurement. This suggests that the estimation of the cochlear stimulation level by the accelerometer at the occiput is within 5 dB of the actual vibration level.

The use of vibration data up to frequencies of 400 Hz may be slightly high. When investigating the mechanical point impedance in cadaver heads, the resonance frequency was found between 150 and 400 Hz; most were between 250 and 300 Hz (Stenfelt and Goode, 2005a). It is only below the resonance frequency that the skull moves as a true rigid body and the occiput moves similarly as the cochleae. This may be the explanation for the reduction of the BC-AC sound field sensitivity according to the vibration measurement at frequencies above 250 Hz. Consequently, the BC-AC sound field sensitivity between 300 and 500 Hz given here is slightly uncertain.

Despite some uncertainty, the comparison of the skull vibration at AC and BC thresholds at the occiput produced the best estimate for the BC-AC sound field sensitivity at the lowest frequencies. It is assumed that the accelerometer only measured the vibrations of the skull bone at the measurement position; however, for sound field stimulation, the accelerometer might have picked up a small amount of vibrations from the air. If so, the vibration level of the AC stimulation would be overestimated; this would give the BC-AC sound field sensitivity too low values. However, since the errors of only using one-dimensional vibration measurements and the influence of AC stimulation directly to the accelerometer are opposite, the resulting error associated with the vibration estimate is assumed limited at the lowest frequencies.

The third method used, ear canal sound pressure measurements (MIRE), did not produce any reliable data. The reason is that the ear canal sound pressure does not necessarily reflect the BC sensitivity; it does so even less after the ear canal is manipulated by fitting earmuffs or earplugs. This can be understood by imagining an earplug that almost fills the whole ear canal, leaving a tiny volume in front of the eardrum. Since almost the whole ear canal wall area is ob-

structed, the sound radiated into the remaining ear canal volume is minimal. However, the perceived BC sound is unaltered since it is almost independent of the ear canal sound pressure when no occlusion effect is present. While great attenuation is obtained by comparing the open ear canal sound pressure with the earplug fitted ear canal sound pressure, the perceived attenuation is limited by the BC transmission to the inner ear. This implies that if HPDs, or combination of HPDs, are used whose attenuation reaches the BC-AC sound field sensitivity, ear canal sound pressure should not be used to assess the perceived attenuation.

In a recent study, Ravicz and Melcher (2001) used the MIRE method in a way similar to ours to measure the impact of HPDs and even a helmet covering the whole skull. Their average MIRE results for combined earmuffs and earplugs are similar to ours when we use our MIRE data before compensating for the occlusion effect (dotted line in Fig. 5); Ravicz and Melcher (2001) did not compensate for the occlusion effect. They further found their MIRE data to produce up to 15 dB greater attenuation than that measured by REAT (they only measured REAT data at five frequencies between 0.5 and 3 kHz), also in line with our findings.

C. The occlusion effect

It was also found that the occlusion effect altered both the threshold and the ear canal sound pressure data, and these results had to be compensated accordingly. Without this compensation, the BC-AC sound field sensitivity based on REAT data would be underestimated, since low-frequency occluded thresholds are improved. Many of the previous reports on this subject did not compensate for the occlusion effect. Zwislocki (1957) calculated the occlusion effect from the results of measurements made by von Békésy (1948), but concluded that no compensation was necessary because there was no relative displacement between the skull and the lower jaw in a sound field. Zwislocki (1957) believed that the occlusion effect is caused by the compression of the auditory canal due to relative movements between the lower jaw and the skull; it is now known that the occlusion effect is independent of the lower jaw relative motion (Allen and Fernandez, 1960; Howell and Williams, 1989; Stenfelt *et al.*, 2003).

Neither Schroeter and Poesselt (1986) nor Berger *et al.* (2003) took the occlusion effect into account when analyzing the data. It should be noted that a deep enough occlusion down into the bony part of the ear canal does not alter the perceived BC sound. However, such occlusion needs to be within a couple of millimeters from the eardrum. The above mentioned studies may have accomplished this and compensation for a perceived occlusion effect was not required. It should also be noted that the aim of their studies was slightly different from ours—we define the BC-AC sound field sensitivity as the BC transmission with an open ear canal (i.e., without occlusion effect) whereas, for example, Berger *et al.* (2003) use the maximum attenuation achieved with combination of earmuffs and earplugs as the BC limit (i.e., with the occlusion effect, if any, present).

D. Spread in data

A greater variation than expected was found between the test subjects; the standard deviations of the measurements are given in Figs. 3–6. However, when comparing our standard deviations with previous investigations, our data spread is in the same range of or slightly greater than Berger (1983), Berger and Kerivan (1983), Ravicz and Melcher (2001), and Berger *et al.* (2003). Although most parameters were controlled, some uncertainties still remain. Some of the variation is believed to originate in the room acoustics. The sound insulated room was well damped, but there may still have been reflections causing a nonhomogenous sound field. Another source of the individual variations may be the HPD fittings. Although the fitting of the earplugs was well monitored, the distance between the earplug and the eardrum might have differed, since we did not check the individual ear canal length. This distance affects the occlusion effect, as it depends on the air volume enclosed between the earplug and the eardrum. Furthermore, some differences in ear canal sound pressure may be due to small variations of the probe microphone tube position. Even if the insertion depth was well controlled, its position in regard to the canal walls could not be verified.

E. Applications of the results

The most obvious area where the estimate of the BC-AC sound field sensitivity is important is in attenuations by HPDs. The BC-AC sound field sensitivity can be seen as a measure of how worse a sound field is transmitted by BC than AC. Since the HPD attenuates the AC path, once the attenuation is equal to the BC-AC sound field sensitivity, the BC path becomes dominant and no further attenuation by the HPD is effective for the wearer. However, a HPD will normally cause an occlusion effect and the maximum possible attenuation by specific HPD fitted in a specific way is lowered according to the occlusion effect by that specific device fitted in that specific way. This means that the maximum attenuation for a HPD can be reduced considerably due to the occlusion effect. For example, using the data in this study, the BC-AC sound field sensitivity at 250 Hz is 50 to 60 dB and the occlusion effect of an earplug is typically 20 dB. This means that the maximum attenuation of any earmuff with the earplug fitted in this way is the BC-AC sound field sensitivity reduced by 20 dB, i.e., 30 to 40 dB. Consequently, the maximum achievable attenuation for any combination of HPDs can be calculated from the BC-AC sound field sensitivity as long as the occlusion effect is known.

Another area where the BC-AC sound field sensitivity provides valuable information is, as stated in the Introduction, in the understanding and design of BC microphones. Although the use of BC microphones has been proposed in areas where an ordinary microphone cannot be used due to practical limitations (e.g., unable to have a microphone in front of the mouth for various reasons), the usefulness of a BC microphone is primarily in an environment with high levels of surrounding noise; we will restrict our discussion to the latter situation. The benefit of using a BC microphone to record one's own voice in a noisy environment manifest it-

self in a greater speech intelligibility than if a noise-cancelling AC microphone is used (Ono, 1977). This improved speech intelligibility relies on an improved SNR of the speech signal. However, it is not known exactly how much the SNR is improved with a BC microphone compared with an AC microphone. The BC-AC sound field sensitivity can be used to derive a rough estimate of this SNR improvement. We will not conduct a thorough analysis of this SNR improvement but merely provide an approximation based on several assumptions. The first assumption is that an approximately equal amount of one's own voice is transmitted by AC and BC (von Békésy, 1949; Pörschmann, 2000); this approximation holds for frequencies up to 2 to 3 kHz. The sound pressure caused by one's own voice at a microphone in front of the mouth and at the ear differs by 10 to 15 dB (von Békésy, 1949), meaning that an AC microphone in front of the mouth records 10 to 15 dB greater signal compared with the BC microphone when related to perception thresholds. Also, a directional AC microphone improves the speech signal by increasing the SNR by around 15 dB compared with an omni-directional microphone. Hence, the assumption is that recording the speech signal in front of the mouth using a directional microphone improves the signal by around 30 dB compared with recording it at the ear (the AC path of the own voice). This means that the signal recorded by a directional AC microphone is, when related to the perceived own voice, around 30 dB above that recorded by the BC microphone.

However, the air borne noise affects the AC microphone more than the BC microphone; in relation to the perception via either AC or BC, this difference is the same as the BC-AC sound field sensitivity. For brevity, we assume that the BC-AC sound field sensitivity is frequency independent at 50 dB. We can then relate the difference between the SNRs by using an AC and BC microphone for recording one's own voice. According to the assumptions, the noise exposure of the AC microphone relative to the BC microphone equals the BC-AC sound field sensitivity and the AC microphone's SNR increases by 15 dB due to the placement close to the mouth and by 15 dB due to the usage of a directional microphone. Hence, a rough estimate is a SNR improvement of 20 dB (50 dB–15 dB–15 dB) using a BC microphone instead of an AC microphone. It should be remembered that these numbers are rough estimates and, for a specific situation, the usage of a BC microphone can improve the SNR both more or less than the computed 20 dB. Also, a BC microphone has low sensitivity at frequencies above 2 kHz, which may severely distort the speech signal.

V. CONCLUSIONS

Three methods were investigated for estimating the BC-AC sound field sensitivity.

- (1) REAT: The sound field hearing thresholds with and without HPDs were compared, which gave a BC-AC sound field sensitivity between 42 and 55 dB in the frequency range 0.5 to 8 kHz. The lower frequency limit of 500 Hz is caused by limited AC attenuation of the HPDs.
- (2) MIRE: The sound pressure levels close to the eardrum

with and without HPDs were compared; according to this method the BC-AC sound field sensitivity is 55 to 65 dB between frequencies of 0.5 and 5 kHz. The low frequency limitation is due to the insufficient AC attenuation at low frequencies. The upper limit at 5 kHz is caused by sound leakage in the probe tube microphone at higher frequencies. As a consequence of the different BC pathways to the cochlea, the sound pressure in the ear canal does not exactly reflect the BC transmission to the cochlea and the BC-AC sound field sensitivity obtained by the MIRE method is too high.

(3) Vibration measurements: The accelerometer was attached at three positions on the head, to measure the skull vibrations with AC and BC stimulation. The vibration levels of the stimulation modalities at the hearing thresholds were compared, to estimate a BC-AC sound field sensitivity between 48 and 55 dB at frequencies below 400 Hz. The frequency range is limited by the vibration pattern in the skull bone and the vibration attenuation properties of the skin at higher frequencies.

The total estimate of the BC-AC sound field sensitivity from the present study, based on the REAT and the vibration methods, together with estimates from some earlier studies are shown in Fig. 8. They seem to agree fairly well, and the BC-AC sound field sensitivity for a normal hearing person is proposed to be between 50 and 60 dB at low frequencies, decreasing to a minimum of 40 to 50 dB at 2 kHz, increasing to a maximum of 50 to 60 dB at 4 kHz, and decreasing again to 40 to 50 dB at 8 kHz.

The BC-AC sound field sensitivity is useful in designing BC microphones; it gives an estimate of the possible improvement in SNR of an individual's own recorded voice when using a BC microphone instead of an AC microphone in a noisy environment. It is also the theoretical limit of the greatest attenuation achievable by ordinary HPDs.

ACKNOWLEDGMENTS

This study was supported by grant from the Swedish Research Council (621-2002-5624). We wish to thank Elliot Berger for discussions and suggestions improving the paper.

¹According to Berger and Kerivan (1983) plug insertion depths, our insertions would be between a partial and standard insertion.

²The leakage was verified by sealing the probe tube opening of the microphone. This sealing was verified to provide more than 60 dB attenuation by comparing the output signal from the probe tube microphone opened and sealed when the probe tube was introduced in a small cavity with a sound stimuli present. This sound stimulus was required to increase more than 60 dB at any frequency in the range 0.1 to 10 kHz for the output from the sealed probe tube to match the open probe tube. Next, the output from an open and sealed probe tube was compared when the whole probe tube microphone assembly was positioned in a sound field. Here it was found that it was only below 5 kHz that sound insulation was at least 60 dB.

³Dirks (1964) found that the reliability of BC measurements increased when the static force exceeded 4 N. We found that by using this static force the influence from the skin was not significant at frequencies below 1 kHz.

Allen, G., and Fernandez, C. (1960). "The mechanism of bone conduction," *Ann. Otol. Rhinol. Laryngol.* **69**, 5–28.

Berger, E. H. (1983). "Laboratory attenuation of earmuffs and earplugs both singly and in combination," *Am. Ind. Hyg. Assoc. J.* **44**, 321–329.

Berger, E. H. (1986). "Methods of measuring the attenuation of hearing protection devices," *J. Acoust. Soc. Am.* **79**, 1655–1687.

Berger, E. H., and Kerivan, J. E. (1983). "Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold," *J. Acoust. Soc. Am.* **74**, 81–94.

Berger, E. H., Kieper, R. W., and Gauger, D. (2003). "Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways," *J. Acoust. Soc. Am.* **114**, 1955–1967.

Black, R. D. (1957). "Ear-insert microphone," *J. Acoust. Soc. Am.* **29**, 260–264.

Dirks, D. (1964). "Factors related to bone conduction reliability," *Arch. Otolaryngol.* **79**, 551–558.

Flottorp, G., and Solberg, S. (1976). "Mechanical impedance of human headbones (forehead and mastoid portion of the temporal bone) measured under ISO/IEC conditions," *J. Acoust. Soc. Am.* **59**, 899–906.

Håkansson, B., Brandt, A., Carlsson, P., and Tjellström, A. (1994). "Resonance frequencies of the human skull *in vivo*," *J. Acoust. Soc. Am.* **95**, 1474–1481.

Howell, P., and Williams, M. (1989). "Jaw movement and bone-conduction in normal listeners and a unilateral hemi-mandibulectomy," *Scand. Audiol.* **18**, 231–236.

Huizing, E. H. (1960). "Bone conduction—The influence of the middle ear," *Acta Oto-Laryngol., Suppl.* **155**, 1–99.

ISO 226. (2003). "Acoustics—Normal equal-loudness-level contours."

ISO 389-3. (1994). "Reference zero for the calibration of audiometric equipment. Part 3: Reference equivalent threshold force levels for pure tones and bone vibrators."

ISO 8253-1. (1989). "Acoustics—Audiometric test methods—Part 1: Basic pure tone air and bone conduction threshold audiometry."

Nixon, C. W., and von Gierke, H. E. (1959). "Experiments on the bone-conduction threshold in a free sound field," *J. Acoust. Soc. Am.* **31**, 1121–1125.

Ono, H. (1977). "Improvement and evaluation of the vibration pick-up-type ear microphone and two-way communication device," *J. Acoust. Soc. Am.* **62**, 760–768.

Oyer, H. J. (1955). "Relative intelligibility of speech recorded simultaneously at the ear and mouth," *J. Acoust. Soc. Am.* **27**, 1207–1212.

Pörschmann, C. (2000). "Influences of bone conduction and air conduction on the sound of one's own voice," *Acta Acust.* **86**, 1038–1045.

Poulsen, T. (1987). "Objective determination of insertion loss of hearing protectors," Internal Rept. No. 29, The Acoustics Laboratory, Technical University of Denmark, Denmark.

Ravicz, M. E., and Melcher, J. R. (2001). "Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: Examination of noise conduction through the ear canal, head, and body," *J. Acoust. Soc. Am.* **109**, 216–231.

Schroeter, J., and Poesselt, C. (1986). "The use of acoustical test fixtures for the measurement of hearing protector attenuation. Part II: Modeling the external ear, simulating bone conduction, and comparing test fixture and real-ear data," *J. Acoust. Soc. Am.* **80**, 505–527.

Stenfelt, S. (2006). "Middle ear ossicles motion at hearing thresholds with air conduction and bone conduction stimulation," *J. Acoust. Soc. Am.* **119**, 2848–2858.

Stenfelt, S., and Goode, R. L. (2005a). "Transmission properties of bone conducted sound: Measurements in cadaver heads," *J. Acoust. Soc. Am.* **118**, 2373–2391.

Stenfelt, S., and Goode, R. L. (2005b). "Bone conducted sound: Physiological and clinical aspects," *Otol. Neurotol.* **26**, 1245–1261.

Stenfelt, S., Wild, T., Hato, N., and Goode, R. L. (2003). "Factors contributing to bone conduction: The outer ear," *J. Acoust. Soc. Am.* **113**, 902–913.

Tonndorf, J. (1966). "Bone conduction. Studies in experimental animals," *Acta Oto-Laryngol., Suppl.* **213**, 1–132.

von Békésy, G. (1948). "Vibration of the head in a sound field and its role in hearing by bone conduction," *J. Acoust. Soc. Am.* **20**, 749–760.

von Békésy, G. (1949). "The structure of the middle ear and the hearing of one's own voice by bone conduction," *J. Acoust. Soc. Am.* **21**, 217–232.

Zheng, Y., Liu, Z., Zhang, Z., Sinclair, M., Droppo, J., Deng, L., Acero, A., and Huang, X. (2003). "Air and bone conductive integrated microphones for robust speech detection and enhancement," *IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 249–254.

Zwislocki, J. (1957). "In search of the bone-conduction threshold in a free sound field," *J. Acoust. Soc. Am.* **29**, 795–804.

Bias due to noise in otoacoustic emission measurements

Bradford C. Backus^{a)}

Center for Auditory Research, The Ear Institute, University College London, 332 Gray's Inn Rd, London WC1X 8EE, United Kingdom

(Received 17 August 2006; revised 22 December 2006; accepted 29 December 2006)

Measurements of otoacoustic emission (OAE) magnitude are often made at low signal/noise ratios (SNRs) where measurement noise generates bias and variability errors that have led to the misinterpretation of OAE data. To gain an understanding for these errors and their effects, a two part investigation was carried out. First, the nature of OAE measurement noise was investigated using human data from 50 stimulus-frequency OAE experiments involving medial olivocochlear reflex (MOCR) activation. The noise was found to be reasonably approximated by circular Gaussian noise. Furthermore, when bias errors were taken into account, measurement variability was not found to be affected by MOCR activation as had been previously reported. Second, to quantify the errors circular Gaussian noise produces for different methods of OAE magnitude estimation for distortion-product, stimulus-frequency, and spontaneous OAEs, simulated OAE measurements were analyzed via four different magnitude estimation methods and compared. At low SNRs (below ~ 6 dB), estimators involving Rice probability density functions produced less biased estimates of OAE magnitudes than conventional estimation methods, and less total rms error—particularly for spontaneous OAEs. They also enabled the calculation of probability density functions for OAE magnitudes from experimental data. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434831]

PACS number(s): 43.64.Jb, 43.64.Yp, 43.66.Yw, 43.60.Cg [BLM]

Pages: 1588–1603

I. INTRODUCTION

The discovery of otoacoustic emissions (OAEs) in 1978 by D. T. Kemp (Kemp, 1978) launched a subfield within hearing research dedicated to measuring and interpreting OAE signals. That work led to the adoption of universal OAE-based hearing screening for neonates in the United States and abroad. Inspired by two goals, (1) to learn more about mechanisms of hearing and (2) to develop other clinical uses, OAE research continues.

The sound pressures of otoacoustic emissions are typically reported as magnitudes, often without adequately taking into account biases introduced by measurement-noise. This paper will investigate OAE measurement-noise and deal with the biases and other errors produced by this noise during frequency-domain analysis—when the signal becomes represented by a magnitude and a phase, e.g., via a fast Fourier transform (FFT)—and where the signal magnitude is constant. Such signals can result from measurements of distortion-product (DP) OAEs, stimulus-frequency (SF) OAEs, and spontaneous (S) OAEs. The implications of these results for transiently evoked (TE) OAEs are also discussed.

All magnitude estimation methods contain bias errors. Bias errors arise because magnitudes are always ≥ 0 . They can be appreciated by thinking about the problem of estimating the magnitude of a signal with true (but unknown) magnitude $A=0$ embedded in measurement-noise. Most estimators of A will generate positive nonzero estimates—biased estimates. Bias also occurs when the signal does not exactly equal 0, but when the signal is still small compared with the

measurement-noise. The amount of bias depends upon the signal/noise ratio (SNR) and the estimator used. For example, an estimator that compares observed magnitudes to a model of how those magnitudes are likely to be distributed for a given SNR can be used to reduce bias.

A second kind of error, which will be termed *variability error*, arises from the randomness of measurement-noise. Like bias error, variability error depends upon both the SNR and the estimator. If an estimator involves averaging, for example, variability error can be reduced.

Both bias and variability error contribute to the total rms error for an estimator—a measure of its reliability. For OAE studies, it is important minimize these errors and to be able to quantify them, for instance by calculating accurate 95% confidence intervals. This is particularly important for studies that seek subtle correlations between physical OAE measures and psychophysical ones. This manuscript quantifies bias, variability, and total rms error due to measurement-noise for four magnitude estimation methods via computer simulations. The results indicate that for low SNR measurements ($\text{SNR} < \sim 6$ dB) Bayesian estimators that use Rice probability density functions (PDFs) can help reduce estimation errors and can produce realistic confidence interval estimates where other estimators fail to do so.

II. METHODS

A. Overview

“What is the best way to estimate the magnitude of an OAE?” The answer to this question depends upon the nature of OAE measurement-noise. To gain insight into the problem, OAE measurement-noise was investigated and then

^{a)}Electronic mail: bradford.c.backus@alum.dartmouth.org

computer simulations were done in which a known hypothetical OAE signal was embedded in artificial noise. The signal's magnitude was then estimated using four different estimation methods. The bias, variability, and total rms error each estimator produced were compared, as were the accuracies of their estimated 95% confidence intervals (CIs).

An outline for this paper is as follows: after the experimental methods are explained and the four estimators are introduced, the nature of OAE measurement-noise is investigated. This is done in part because two of the four estimators: Bayesian estimators and maximum *a posteriori* (or maximum likelihood) estimators involve a statistical model that relies on assumptions about the measurement-noise. Finally, the performance of the four estimators is compared via computer simulations.

B. Experimental methods

This manuscript includes data from experiments to investigate the nature of OAE measurement-noise and to test the assumptions that could enable improved magnitude estimation for OAEs. The experiments monitored an SFOAE and measured the steady-state change in that emission induced by activation of the medial olivocochlear reflex (MOCR). These OAE/MOCR experiments were used because they provided separate measurements of (1) an OAE signal and (2) the measurement-noise. The theoretical portion of the paper, however, is not limited to this kind of experiment.

Twenty-five adult subjects with normal audiograms were measured. For each measurement, a subject was seated in a sound-absorbing chamber and fitted with two Etymotic Research ER10c earphones (each equipped with two sound sources and a microphone) for several ~30-min measuring sessions.

To generate a SFOAE in one ear, a continuous 40 dB SPL probe-tone near 1 kHz (within 12%) was played bilaterally¹ through one sound source in each ER10c earphone. To elicit a change in that emission, the medial olivocochlear reflex (MOCR) was activated by a 4-s 60-dB SPL contralateral wideband (100 Hz to 10 kHz) noise-burst (5 ms rise/fall) presented through the second source of the opposite earphone at 8-s intervals. A stimulus "presentation" consisted of 4-s of probe-tone-only and 4-s with MOCR activation.

The magnitude and phase of the ear canal sound pressure at the probe frequency was extracted by first averaging the responses from a number of presentations (range 44–284) and then heterodyning that averaged waveform (Kim *et al.*, 2001; Guinan *et al.*, 2003). The change in the SFOAE due to MOCR activation was calculated by vector subtracting a baseline no-MOCR activation pressure (an average over 400 ms taken from a different section of the no-MOCR condition than the noise-measurement) from the pressures at each time sample (resolution=5 ms) during MOCR activation and deactivation. Since only the steady-state change was desired, only those time samples >1.5 s from MOCR activation (or deactivation) were used. This settle time is greater than four time constants of the overall MOCR "fast effect" onset or decay (Backus and Guinan, 2006).

Each experiment produced 800 data points (each having a magnitude and phase), 400 points representing the steady-state change in the SFOAE due to MOCR activation (i.e., the signal) and 400 points of measurement-noise which contained an unsuppressed SFOAE, i.e., no MOCR activation.

C. Theoretical methods

Computer simulations were used to compare the errors introduced by four different OAE magnitude estimators. For the simulations, a known signal with constant magnitude A and phase $\phi = \pi/4$ was embedded in circular Gaussian noise (CGN, see Sec. III A for a definition of this term) with known statistics. These simulated data were created for various signal-magnitude/noise-variance combinations using Matlab's "normrnd" function. The datasets were then analyzed via four different estimators and their biases, standard deviations (SDs), and total rms errors were computed along with their estimated 95% confidence intervals (CIs). Since different OAEs are typically measured and analyzed in different ways, the four estimation methods that were chosen depended upon the type of OAE being simulated.

1. Magnitude estimators for DPOAEs and SFOAEs

Distortion-product, stimulus-frequency, and spontaneous otoacoustic emissions (DPOAEs, SFOAEs, and SOAE) are analyzed in the frequency domain at a single frequency and can have constant magnitudes. Of these, DPOAEs and SFOAEs can be measured with reference to a stimulus and thereby give informative phase as well as magnitude data.² Since DPOAEs and SFOAEs occur at known frequencies, a heterodyne method of signal extraction is often used to produce hundreds of observations across time (Kim *et al.*, 2001; Guinan *et al.*, 2003). These observations can change with time, for instance if a nearby suppressor tone is introduced or if the medial olivocochlear (MOC) reflex is stimulated, but here only steady-state or nonchanging emissions are considered. The DPOAE/SFOAE data will be represented as $N = 1000$ complex-valued observations (with real and imaginary components). The SFOAE/DPOAE results take N into account by using a SNR measure that divides by the square root of N . These results are insensitive to N .

The following sections describe the four DPOAE/SFOAE magnitude estimators that were chosen to operate on this data.

*a. DPOAE/SFOAE conventional estimator.*³ The "conventional estimator" is the most commonly used estimator in OAE literature. It was calculated by first vector-averaging the $N = 1000$ complex-valued observations, \mathbf{M}_n , from a simulation, then taking the resulting magnitude. This calculation was termed the *conventional estimator* of the sought magnitude, A :

$$\hat{A}_{CE} = \left| \frac{1}{N} \sum_{n=1}^N \mathbf{M}_n \right|. \quad (1)$$

b. DPOAE/SFOAE modified conventional estimator. Repeated estimates of \hat{A}_{CE} from additional data produce a Rice distribution (see Appendix A). Due to the vector averaging in

Eq. (1), the averaged (real and imaginary) component variance, σ_N^2 , that enters as a parameter in this Rice distribution will be a factor of N smaller than the original component variance, σ^2 , of the observed \mathbf{M}_n .

$$\sigma_N^2 = \frac{\sigma^2}{N}. \quad (2)$$

A property of this Rice distribution is that the second moment is

$$E[M^2] = A^2 + 2\sigma_N^2. \quad (3)$$

This property was used to calculate a *modified conventional estimator* of A .⁴ The calculation was done using an estimate of the original component variance, $\hat{\sigma}^2$, calculated by taking the average of the simulated data's real and imaginary component variances.⁵

$$\hat{A}_{MCE} = \text{Re} \sqrt{\left| \frac{1}{N} \sum_{n=1}^N \mathbf{M}_n \right|^2 - 2 \frac{\hat{\sigma}^2}{N}}. \quad (4)$$

c. DPOAE/SFOAE maximum a posteriori estimator. The *maximum a posteriori* estimator used the conventional estimator value [Eq. (1)] and the knowledge that this value itself comes from a Rice distribution with parameters A and σ_N to produce an estimate of A .

The probability of observing a specific conventional estimate, A_{CE} , from N samples is

$$p_{A_{CE}}(A_{CE}|A, \sigma_N) = \frac{A_{CE}}{\sigma_N} e^{-(A_{CE}^2 + A^2)/2\sigma_N^2} \cdot I_0\left(\frac{A \cdot A_{CE}}{\sigma_N^2}\right), \quad (5)$$

where I_0 is a modified Bessel function of the first kind [see Appendix A for a proof and further description of Eq. (5)].

This probability distribution was reinterpreted as the probability of A for a calculated \hat{A}_{CE} by (1) allowing A to be the variable in Eq. (5), (2) evaluating the expression at \hat{A}_{CE} and $\hat{\sigma}_N$, and (3) assuming all values of A were initially equally likely (no *a priori* probability of A) and normalizing over all possible values of A . This produced a probability density function (PDF) of A for specific calculated \hat{A}_{CE} and $\hat{\sigma}_N^2$ [Fig. 1; Eq. (6)]:

$$p_A(A|\hat{A}_{CE}, \hat{\sigma}_N) = \frac{p_{A_{CE}}(A_{CE}|A, \sigma_N)}{\int_{A=0}^{\infty} p_{A_{CE}}(A_{CE}|A, \sigma_N) dA} \Bigg|_{\substack{A_{CE}=\hat{A}_{CE} \\ \sigma_N=\hat{\sigma}_N}}. \quad (6)$$

The *maximum a posteriori* estimator (or maximum likelihood estimator⁶) was taken as the most probable value of A given \hat{A}_{CE} and $\hat{\sigma}_N$:

$$\hat{A}_{MAPE} = \underset{A}{\text{argmax}}(p_A(A|\hat{A}_{CE}, \hat{\sigma}_N)). \quad (7)$$

d. DPOAE/SFOAE Bayesian estimator. Like the *maximum a posteriori* estimator the Bayesian estimator⁷ used the posterior probability of Eq. (6), i.e., $p_A(A|\hat{A}_{CE}, \hat{\sigma}_N)$. How-

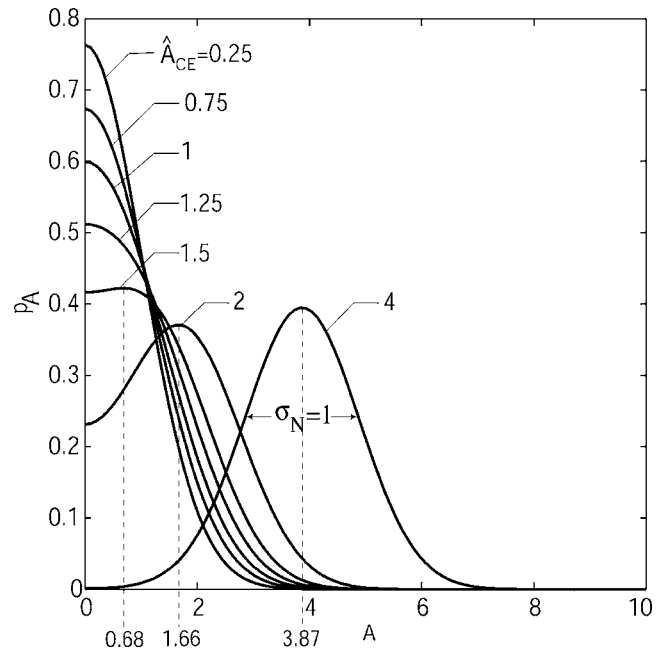


FIG. 1. A set of probability density functions (PDFs) that describe the probability that an OAE signal with magnitude A —embedded in circular Gaussian noise with standard deviation $=\sqrt{N} \cdot \sigma_N$ —would produce a specific conventional estimate (CE) from N observations. Probability density functions for seven different CEs are shown. The *maximum a posteriori* estimate of A for a given CE was the value of A that maximized the associated PDF; this value is pointed to by dotted lines where different from 0.

ever, instead of maximizing the probability of A , the Bayesian estimator minimized the total squared error between all the possible estimates and the true value of A . This strategy results in choosing the posterior conditional mean as the estimate (i.e., the mean from a curve in Fig. 1 rather than its maximum). Bayesian estimator values will differ from those of the *maximum a posteriori* estimator when the PDF (e.g., of Fig. 1) is asymmetric—at low SNRs:

$$\hat{A}_{BE} = \int_{A=0}^{\infty} A \cdot p_A(A|\hat{A}_{CE}, \hat{\sigma}_N) dA. \quad (8)$$

e. Posterior PDF 95% confidence interval estimation for the DPOAE/SFOAE maximum a posteriori and Bayesian estimators. A 95% confidence interval about \hat{A}_{MAPE} or \hat{A}_{BE} was calculated from the probability density function described by Eq. (6) (or as shown in Fig. 1) by finding the smallest interval, $[a, b]$, along A that produced a 95% probability of having generated the observed conventional estimate [the constraint of finding the smallest 95% CI is stated after the semicolon in Eq. (9)]:

$$0.95 = \int_a^b p_A(A|\hat{A}_{CE}, \hat{\sigma}_N) dA; \quad \min(b - a). \quad (9)$$

The interval was calculated iteratively. The interval calculation was the same for *maximum a posteriori* and Bayesian estimators because both estimates are based on the same PDF [i.e., Eq. (6) or Fig. 1]. The difference is that \hat{A}_{MAPE} and

\hat{A}_{BE} values occur at different points within this interval [see Fig. 6(b)].

f. Bootstrapped 95% confidence interval estimation for the DPOAE/SFOAE conventional estimator and modified conventional estimator. Since conventional estimator and modified conventional estimator calculations do not involve probability distributions, a simple bootstrap type of 95% confidence interval (CI) was done for these. The estimators themselves were used to generate their own bootstrapped CIs. For each, the general method was the same. The $N=1000$ complex-valued observations, \mathbf{M}_n , from a single simulated measurement were pooled; 1000 were selected from this pool at random with replacement (i.e., duplicates and omissions were allowed). An estimate (conventional or modified conventional) was calculated from this resampled data and the process was repeated 1000 times. This resulted in 1000 estimates of each type from a single simulation. From these the standard deviation of each estimate type was calculated. The 95% confidence interval was considered to be 1.96 standard deviations from original estimate. If the lower bound of the interval ended below 0 it was truncated to 0. This bootstrap procedure assumes the estimates are unbiased, but this was not the case (see Fig. 5).

2. Magnitude estimators for SOAEs

Spontaneous otoacoustic emissions (SOAEs) are measured in the absence of any stimulus so they have no phase reference, and their frequency is not precisely known. Without a phase reference, averaging successively acquired buffers of ear canal pressure in the time domain (point by point) reduces a SOAE signal as well as any noise. So, acquiring high-quality low SNR SOAE measurements is more laborious than for DPOAEs or SFOAEs. Furthermore, without precise knowledge of their frequency, the heterodyne method is less powerful. For these reasons, it makes sense to acquire long duration buffers and perform long duration FFTs when measuring SOAEs. But increasing SNR using the long FFT solution is limited by a human subject's inability to remain quiet for a long time. For these reasons, new approaches for estimating SOAE magnitudes could be helpful.

Here, simulated magnitude observations—which would, in the real world, be calculated from FFTs⁸ of serially acquired ear canal pressures—are used as the input to four estimators. This magnitude data will be represented as $N=20$ magnitude observations with unknown underlying real and imaginary components, containing the signal with magnitude A embedded in CGN.

The number of observations, $N=20$, was chosen because (1) the calculation of CIs by bootstrapping required multiple observations to be accurate and (2) the calculation of the Bayesian and maximum likelihood estimators required multiple observations to be accurate. In prior work, we had used 2.6 s for our FFT length (Guinan *et al.*, 2003; Backus and Guinan, 2006), so 20 observations would have corresponded to ~ 1 min of measurement time. Unlike DPOAEs and SFOAEs, the SOAE results are sensitive to N . Specifically, as N increases, the Bayesian and maximum likelihood estimators provide increasing benefits over the conventional and

modified conventional estimators at low SNRs. To illustrate this effect of N , an additional simulation is presented where $N=1000$ observations were used.

Since only magnitude data are available, the estimators are different for SOAEs than for DPOAEs or SFOAEs. The following sections describe the four SOAE magnitude estimators that were chosen to investigate.

a. SOAE conventional estimator (CE). For the SOAE conventional estimator, the N real-valued, magnitude-only observations, M_n , were averaged. The result was the conventional estimate of the magnitude A :

$$\hat{A}_{CE} = \frac{1}{N} \sum_{n=1}^N M_n. \quad (10)$$

b. SOAE modified conventional estimator. The distribution of observed SOAE magnitudes, M_n , produces a Rice distribution (see Appendix A). As was the case for DPOAEs and SFOAEs, the second moment of the Rice distribution (this time without the benefit of vector averaging),

$$E[M^2] = A^2 + 2\sigma^2, \quad (11)$$

was used to estimate A :

$$\hat{A}_{MCE} = \text{Re} \sqrt{\left(\frac{1}{N} \sum_{n=1}^N M_n\right)^2 - 2\hat{\sigma}^2}. \quad (12)$$

The SOAE modified conventional estimator of Eq. (12) is different from the one for DPOAEs/SFOAEs [Eq. (4)] because here only magnitudes are available. The benefit of vector averaging—apparent in Eq. (4) as the reduction of the variance term by a factor of N —is abolished.

Since the real and imaginary components cannot be separately observed, an additional noise measurement is required to estimate their component variance, σ^2 (see Appendix B). For this, an additional simulation as done where the signal magnitude was set to 0 and the measurement-noise variance, $\hat{\sigma}^2$, was estimated by the method described in Appendix B. In a real experiment this noise measurement could be done by using a suppressor tone to eliminate the SOAE or by measuring at a nearby frequency where there was no SOAE.

c. SOAE maximum likelihood estimator. The SOAE maximum likelihood estimator differed from the DPOAE/SFOAE maximum a posteriori estimator only in implementation. Instead of using the single conventional estimate and its associated PDF (Fig. 1), the SOAE maximum likelihood estimator was calculated by using all M_n observations and the estimated noise variance, $\hat{\sigma}^2$, to iteratively search for the Rice parameter, A , that maximized the likelihood of the observed M_n —essentially a curve to histogram fitting procedure:

$$p_{M_n}(M_n|A, \hat{\sigma}) = \frac{M_n}{\hat{\sigma}^2} e^{-(M_n^2 + A^2)/2\hat{\sigma}^2} I_0\left(\frac{AM_n}{\hat{\sigma}^2}\right), \quad (13)$$

$$\hat{A}_{MLE} = \underset{A}{\operatorname{argmax}} \prod_{n=1}^N p_{M_n}(M_n|A, \hat{\sigma}). \quad (14)$$

For faster computation, an equivalent was used (Sijbers *et al.*, 1998),

$$\hat{A}_{MLE} = \underset{A}{\operatorname{argmin}} \left(\frac{NA^2}{2\sigma^2} - \sum_{n=1}^N \ln I_0 \left(\frac{A \cdot M_n}{\sigma^2} \right) \right). \quad (15)$$

I_0 is a modified Bessel function of the first kind.

d. SOAE Bayesian estimator. The SOAE Bayesian estimator was calculated by iteratively updating posterior probabilities of A [Eq. (16)] with successive M_n observations:

$$p_A(A|M_n, \hat{\sigma}) = \frac{p_{M_n}(M_n|A, \hat{\sigma}) p_A(A|M_{n-1}, \hat{\sigma})}{\int_{A=0}^{\infty} p_{M_n}(M_n|A, \hat{\sigma}) p_A(A|M_{n-1}, \hat{\sigma}) dA}. \quad (16)$$

To start, when $n=1$, there was no prior probability term, $p_A(A|M_{n-1}, \hat{\sigma})$, and Eq. (16) devolved to the same calculation as Eq. (6). Thereafter, the posterior probability of A was updated with subsequent M_n observations until all observations were exhausted ($N=20$). The resulting PDF of A was used to estimate A by calculating the conditional posterior mean:⁹

$$\hat{A}_{BE} = \int_{A=0}^{\infty} A \cdot p_A(A|M_N, \hat{\sigma}) dA. \quad (17)$$

e. Posterior PDF 95% confidence interval estimation for the SOAE maximum likelihood estimator and Bayesian estimator. CIs for the maximum likelihood and Bayesian SOAE estimators were calculated using the same method described for the DPOAE/SFOAE *maximum a posteriori* and Bayesian estimators. The 95% confidence interval was calculated from the posterior probability density function [this time Eq. (16)] by iteratively finding the smallest interval, $[a, b]$, of A that produces a 95% probability of having generated the observed M_n .

f. Bootstrapped 95% confidence interval estimation for the SOAE conventional estimator and modified conventional estimator. CIs for the SOAE conventional and modified conventional estimators were calculated by the same bootstrap method described for the analogous DPOAE/SFOAE estimators except that the $N=20$ magnitude-only data were pooled from which 20 were taken with replacement. This was done 20 times and each time the estimators were calculated. The interval spanning 1.96 standard deviations of these estimates were used to place 95% CIs about the original estimates.

III. RESULTS

The results are in two parts. The first part discusses the nature of OAE measurement-noise by analyzing the noise from actual human experimental data. The second part compares four methods of estimating OAE magnitudes in terms

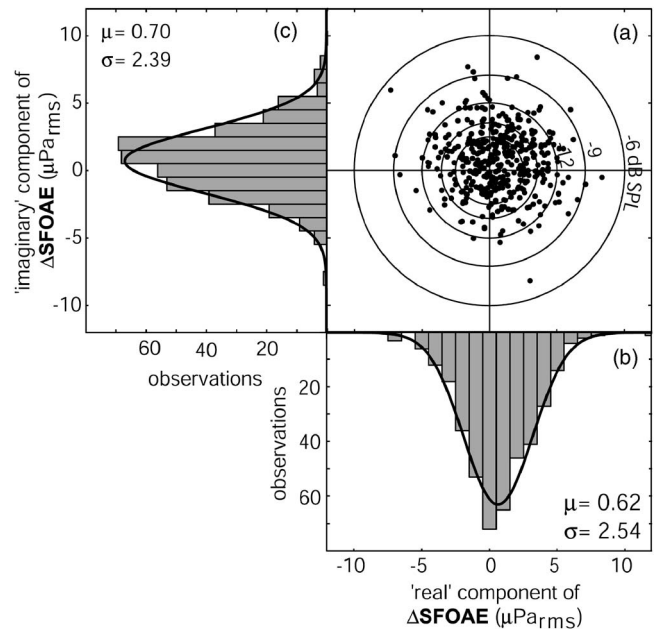


FIG. 2. Measured noise (dots) from a 1-kHz stimulus-frequency otoacoustic emission measurement (panel a). Histograms of the real (panel b) and imaginary (panel c) components of the measurement-noise show that both components have approximately Gaussian distributions. Curves are Gaussians made with statistics calculated from the data.

of their bias, variability, total rms error, and the accuracy of their estimated confidence intervals (CIs) by analyzing simulated data.

A. Properties of OAE measurement-noise

In practice, to measure an OAE magnitude, a number of ear-canal responses are recorded and typically averaged. The resulting averaged response is then “filtered” and/or averaged again across successive samples. With all this averaging, and if signal + measurement-noise is statistically invariant, the central limit theorem dictates that the averaged quantities will have a Gaussian distribution, regardless of whether the underlying noise or signal distributions are themselves Gaussian. The central limit theorem also insures that the mean of the result will be an unbiased estimate of the mean of the underlying signal + measurement-noise. The central limit theorem, however, does not guarantee bias free estimates for magnitude data because magnitude data are bounded, i.e., magnitudes cannot be <0 .

Figure 2 shows measurement-noise from an actual averaged response of an SFOAE measurement (Female subject, right ear, 284 averages). Each time sample (dots) represents the steady-state change in the emission’s magnitude and phase for the control condition, i.e., without any MOCR activation, relative to an earlier average no-MOCR baseline. The distribution of the real and imaginary components of the measurement-noise were approximately Gaussian with mean=0. A Komolgorov-Smirnov (KS) test¹⁰ was consistent with the idea that both the real and imaginary data could have been drawn from the same zero-mean Gaussian distribution (i.e., there was no reason to reject that null hypothesis; $P=0.73$). Furthermore, the real and imaginary components effectively shared the same standard deviation (2.54 vs.

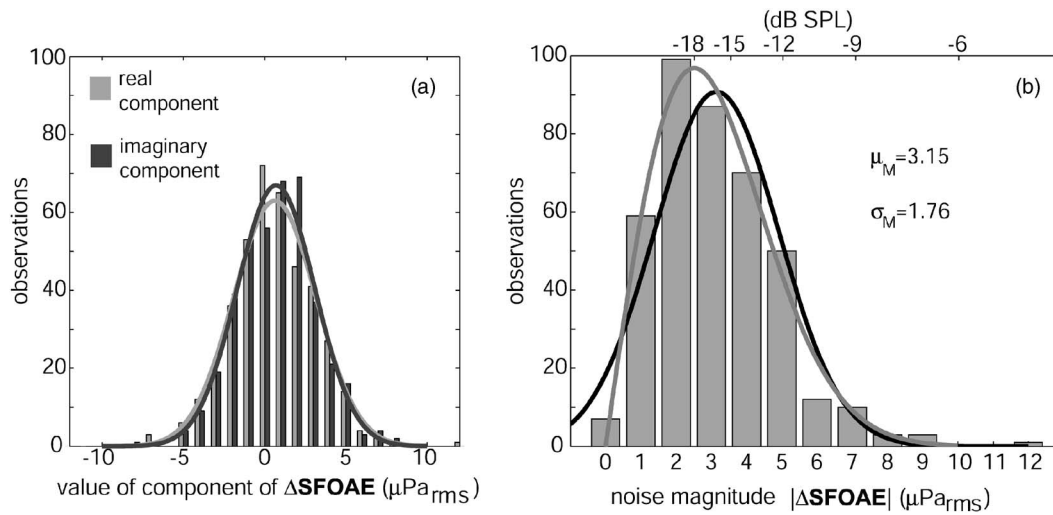


FIG. 3. Panel (a) shows histograms of the real (light) and imaginary (dark) components of measurement-noise from a set of SFOAE measurements (same data as Fig. 2). Two color-coded curves represent associated Gaussian functions created using the statistics from these data. The real and imaginary components of the measurement-noise effectively shared the same Gaussian distribution. Panel (b) shows a histogram of the magnitudes of the measurement-noise (from data in Fig. 2) overlaid with a Rayleigh (light curve) and a Gaussian (dark curve) function. The measurement-noise amplitude was better approximated by the Rayleigh distribution.

2.39) while their covariance terms were $<10\%$ ¹¹ (an arbitrary criteria) of the mean of their individual variances, suggesting that the same noise process influenced both channels relatively independently.

Magnitudes calculated from the complex data in Fig. 2, however, were neither zero-mean nor Gaussian; they were approximately Rayleigh (Fig. 3, also Appendix B). This is important because OAE magnitudes are often treated as though they are Gaussian distributed; but at low SNRs where noise dominates the response, they are not!

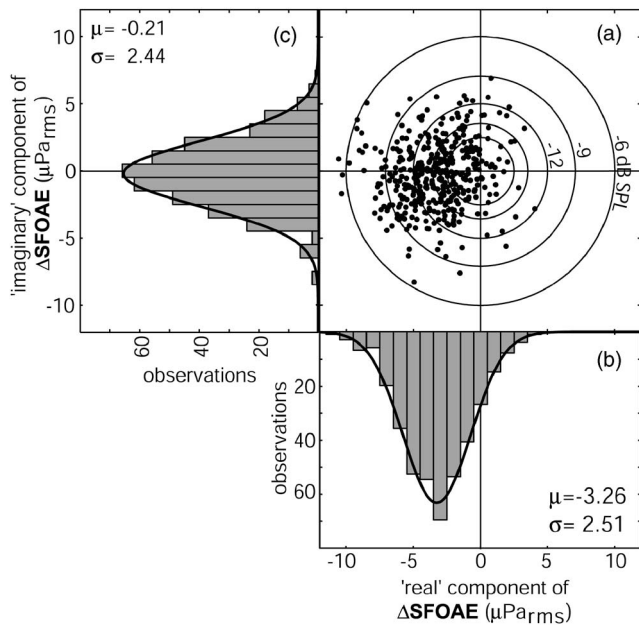


FIG. 4. Measured changes in a 1-kHz stimulus-frequency otoacoustic emission under MOCR activation with 60 dB SPL wideband contralateral acoustic stimulation (dots, panel a). Histograms of the real (panel b) and imaginary (panel c) components of the changes show that both components have approximately Gaussian distributions. Curves are Gaussians made with statistics from the data.

Figure 4 shows a result from the same experiment as Fig. 1 but for the case of a weak OAE signal embedded in the measurement-noise. In this case, the signal was a change induced in the SFOAE by the addition of 60 dB SPL wideband noise in the contralateral ear. Figure 4 shows that the distribution of signal + measurement-noise was again approximately Gaussian for both the real and imaginary components. A Shapiro-Wilks test confirms there was no evidence of either component being non-Gaussian ($P=0.93$ and 0.83 , respectively). Furthermore, measurement-noise during the “signal” (i.e., with MOCR activation in Fig 4) had effectively the same standard deviation (SD) as the measurement-noise in the absence of a “signal” (i.e., without MOCR activation in Fig. 3, total SD range 2.39–2.54 $\mu\text{Pa}_{\text{rms}}$). Thus the effect of MOCR activation was to change the SFOAE simply by shifting the means of the real and imaginary components.

These observations lead to four assumptions about the nature of OAE measurement-noise:

- (1) The noise is zero-mean and Gaussian for both the real and imaginary components.
- (2) The noise statistics are identical for both the real and imaginary components.
- (3) The noise contributes to the real and imaginary components independently.
- (4) The noise is statistically time invariant (e.g., it does not change with the signal).

Together these assumptions define *circular Gaussian noise* (CGN).

To determine whether CGN is a universal condition of OAE measurement-noise, 50 experiments, including the one above, were tested against the CGN assumptions. Forty-six of the 50 experiments were consistent by the KS-test¹⁰ with the assumption that the same zero-mean Gaussian noise adds to both the real and imaginary components; 45/50 had a covariance term $<10\%$ of the mean of the component vari-

ances, suggesting that the noise added to real and imaginary components relatively independently. Here I note that some correlation will always exist between real and imaginary components due to subject movement or other common artifacts and that this 10% criterion is arbitrary and depends on the number of observations.

An F -test across all 50 experiments and 25 subjects indicated that there was no significant difference ($P=0.63$) in component variance with and without MOCR activation. Since this test assumes normality and since the variances were not rigorously normally distributed, a KS-test was also used to compare the two variance distributions. The KS-test did not find any evidence that the variances with and without MOCR activation came from different distributions ($P=0.56$). Thus, MOCR activation does not appear to co-produce a new source of measurable noise nor does it appear to reduce the existing noise variance as previously reported (Maison *et al.*, 1997). In total, 42/50 experiments met criteria for all four CGN assumptions.

These data together with the central-limit theorem provide a basis for modeling OAE measurement-noise as CGN. All ensuing simulations were done using artificial CGN.

B. Comparison of DPOAE/SFOAE magnitude estimators

Four DPOAE/SFOAE estimators, a conventional estimator, a modified conventional estimator, a maximum *a posteriori* estimator, and a Bayesian estimator, were compared using simulated datasets. SNR, A/σ_N , was varied by changing σ_N over a range from 0.5 to 4.0 with A fixed at 1. At each SNR tested, the whole procedure was repeated 1000 times and the mean and standard deviations of the estimates were calculated (Fig. 5).

Figure 5 shows that the Bayesian estimator had the fastest convergence to true A with increasing SNR. The BE also had the smallest variability errors (or SDs) for SNR < 2.5 (bottom axis) and the smallest total rms error for $0.75 < \text{SNR} < 2.75$ —this corresponds to a more conventionally measured SNR, the ratio of signal to noise magnitudes, between 0.6 and 2.2 (top axis) or -4.5 to 6.8 dB. In terms of total error, the BE was the most reliable estimator for SNRs from 0.75 to 2.75. Above 2.75, the CE produced the least total rms error. Although the simulations used $N=1000$ observations, the results do not depend on N ,¹² because all DPOAE/SFOAE estimators vector averaged over all N observations as a first step in their calculations. By plotting the results versus A/σ_N the effect of N in this vector averaging was incorporated. Simulations done with $N=20$ observations (not shown) produced quantitatively similar results to those shown here done with $N=1000$.

The distributions of the different estimators (insets in Fig. 5) revealed that they produced different kinds of errors. For example, both the *maximum a posteriori* estimator and modified conventional estimator were sensitive to single terms in their calculations. Consider the modified conventional estimator: when the argument of its square root function is negative [Eq. (4)], the estimate is forced to zero. For the *maximum a posteriori* estimator, the single value at the

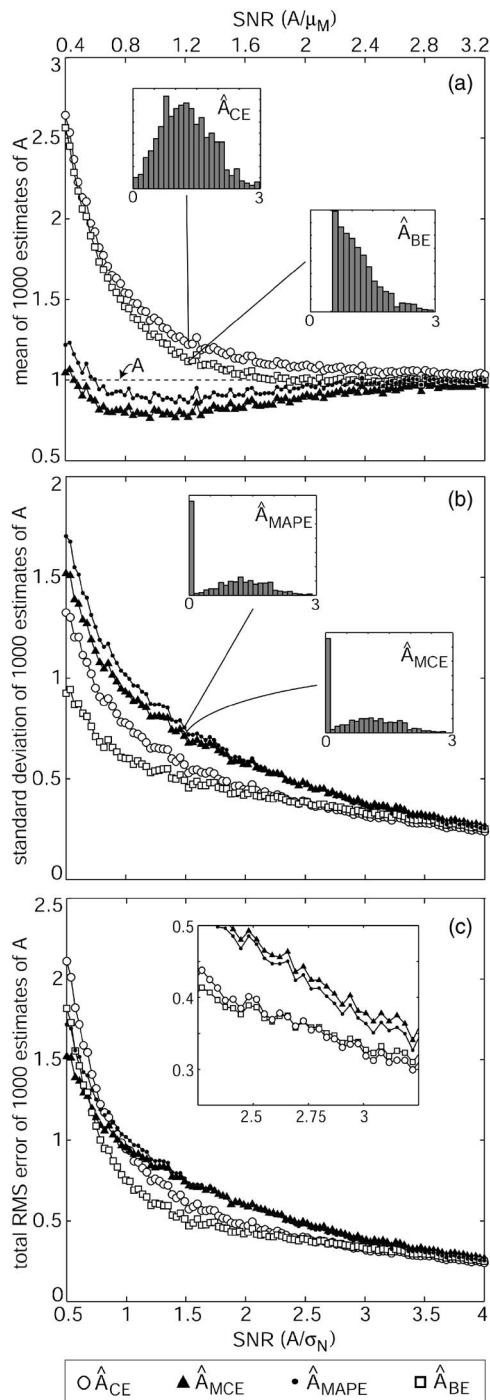


FIG. 5. Comparison of the bias (panel a), variability (panel b), and total rms error (panel c) introduced into noisy DPOAE/SFOAE signal magnitude estimates by four estimators: a conventional estimator (CE), a modified conventional estimator (MCE), a maximum *a posteriori* estimator (MAPE), and a Bayesian estimator (BE). The data were simulated. The lower SNR axis uses noise standard deviation while the upper SNR axis is the conventional ratio of signal to noise magnitudes. Histograms for each estimator are shown for lower SNR=1.5 (histograms are segregated two to a panel only to save space). The inset in panel (c) shows detail of the transition region about lower SNR=2.75, where the CE overtakes the BE in terms of total rms error.

maximum of the posterior PDF determines the estimate. Random fluctuations from the noisy simulations interacted with these sensitivities and when the SNR was near 1 there was a $\sim 50\%$ chance of producing an estimate of $A=0$ from either one (the modified conventional and *maximum a poste-*

riori estimators were the only estimators that produced estimates of 0). The Bayesian and conventional estimators, due to their averaging, did not have such sensitivities so they produced more consistent, lower SD, estimates. However, both of these regularly overestimated A for $\text{SNR} < 2$.

C. Comparison of confidence interval estimation methods for DPOAE/SFOAE

Confidence intervals (CIs) are used to indicate the (typically 95%) probability that the actual value of A lies within a specified range of its estimate, \hat{A} . To investigate the accuracy of CIs calculated from different estimation methods, simulations were done using 21 values of true A from 0 to 5.0 with σ_N fixed at 1. For each A value, 1000 CIs were calculated either by bootstrapping or by using the posterior PDF (see Sec. II B 1).

Figure 6 shows that CIs calculated via posterior probabilities (for the *maximum a posteriori* and Bayesian estimators) generated CIs within $\sim 3\%$ of 95% after 1000 simulations over the entire SNR range tested [Fig. 6(c)], CIs calculated for the conventional estimator by bootstrapping were within 3% for $\text{SNR} > 1$ but were less accurate at $\text{SNR} < 1$ due to bias errors. CIs for modified conventional estimator showed a dip at $\text{SNR} = 2$ (a region where the bias/variability error ratio was a maximum).

Bootstrapped CIs for an estimator were accurate when variability errors dominated the total rms error. However, when bias errors dominated, the bootstrapped CIs were themselves biased and thus were not accurate [e.g., the conventional estimator at $\text{SNR} < 0.8$ in Fig. 6(c)]. One advantage of CIs calculated with posterior probabilities is that they can remain accurate at very low SNRs [Fig. 6(c)], where the total rms error of an estimate is large. So even if an estimated value exceeds some total rms error criteria for reporting it, its posterior PDF calculated CI can still be reported. In other words, there may be times where it is appropriate to report only the confidence interval.

None of the CIs were symmetric about any estimate at low SNR. This was expected because magnitude estimates are bounded at 0 and therefore CIs cannot extend below 0. So CIs were not symmetric on this linear (e.g., $\mu\text{Pa}_{\text{rms}}$) scale nor would they be symmetric on a logarithmic decibel scale.

The results for the posterior PDF generated CIs for Bayesian and *maximum a posteriori* estimators suggest that PDFs calculated via Eq. (6) (like those shown in Fig. 1) are, on average, realistic representations of the probability distribution of DPOAE/SFOAE magnitudes when the assumptions of CGN are valid.

D. Comparison of SOAE magnitude estimators

Four SOAE estimators, a conventional estimator, a modified conventional estimator, a maximum likelihood estimator, and a Bayesian estimator, were compared using simulated datasets. SNR, A/σ_N , was varied by changing σ_N over a range from 1 to 0.071 with A fixed at 1, corresponding to a conventionally measured SNR, A/μ_M , from 0.36 to 2.5 (top axis in Fig. 7). A/μ_M is the commonly reported

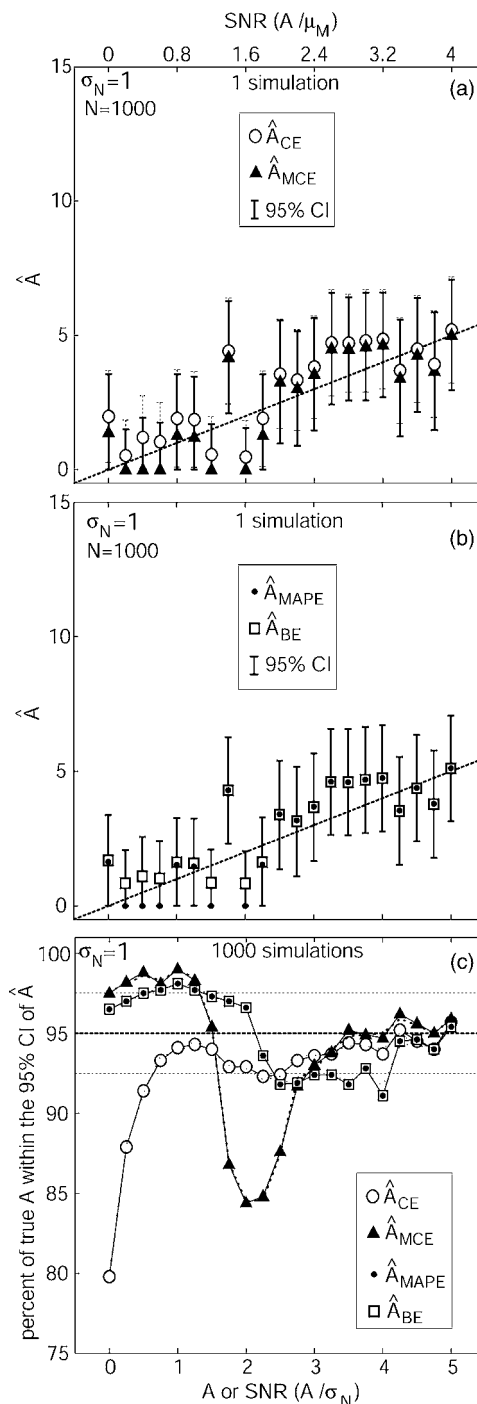


FIG. 6. Panel (a) shows one simulation for the conventional (CE) and modified conventional (MCE) DPOAE/SFOAE magnitude estimators versus SNR and their 95% confidence intervals (CIs) calculated by bootstrapping (CE CIs are dotted and MCE CIs are solid). The dotted line $y=x$ is drawn to facilitate the comparison of true A with the CIs. Panel (b) shows the same simulation recalculated using Bayesian (BE) and *maximum a posteriori* (MAPE) estimators. CIs for the BE and MAPE were both calculated using the same posterior probability density functions and were thus identical. Panel (c) shows the percentage of true A that fell within the 95% CI for each estimator after 1000 simulations (including the ones in panels a and b). Ideally this percentage would be 95% regardless of SNR, but the CIs themselves are estimated and contain errors.

SNR quantity for magnitudes, where μ_M is the mean measurement-noise magnitude; here A/σ_N was included as the lower axis to allow for comparison with DPOAE/SFOAE data (Figs. 5 and 6).

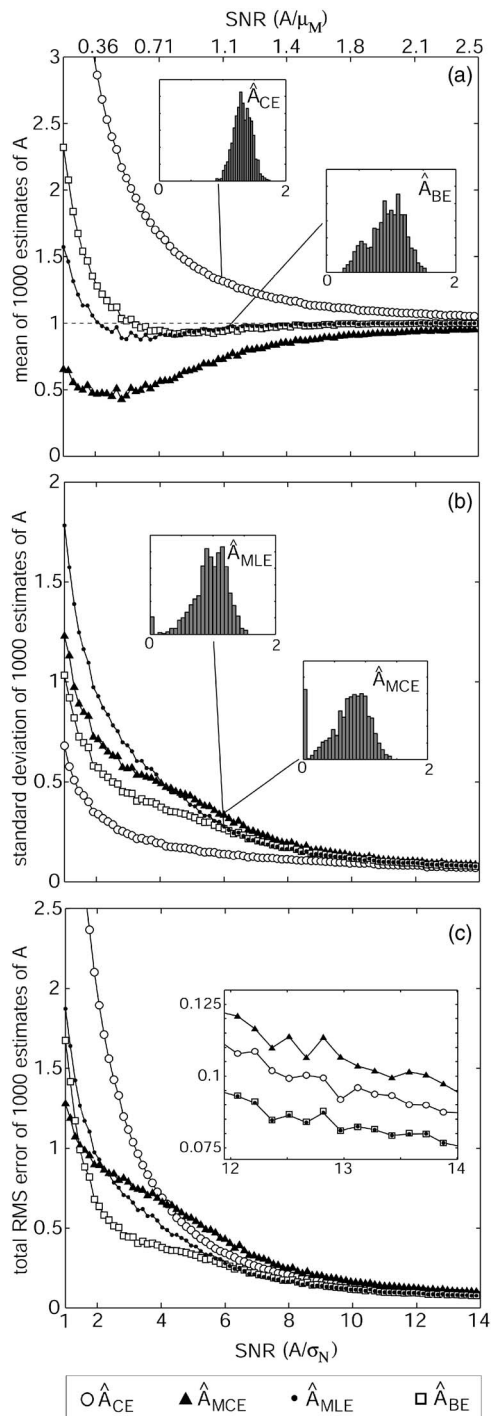


FIG. 7. Comparison of the bias (panel a), variability (panel b), and total rms error (panel c) introduced into noisy SOAE signal magnitude estimates of $N=20$ observations by four estimators: a conventional estimator (CE), a modified conventional estimator (MCE), a maximum likelihood estimator (MLE), and a Bayesian estimator (BE). The data were simulated. Histograms for each estimator are shown for $\text{SNR}=6$ (histograms are segregated two to a panel only to save space). The inset in panel c shows detail of the high SNR region, where the BE continues to outperform the CE.

At each SNR tested, the whole procedure was repeated 1000 times and the mean and standard deviations of the estimates were calculated (Fig. 7).

For SOAEs, the maximum likelihood and Bayesian estimators introduced less bias than the conventional or modified conventional ones. The improvement was due to using a

Rice PDF to model how magnitudes are distributed (see Appendix A). The Rice PDF model was more beneficial for SOAEs than for DPOAEs and SFOAEs because SOAEs contained *only* magnitude data and reducing the SNR by vector averaging was not possible. Unlike results for DPOAE/SFOAEs, the SOAE results *did* depend on the number of observations, N . The benefit of using a Rice PDF model increases as the number of magnitude observations, N , increases (see Fig. 9). This provides a method to continue to reduce the effective SNR for SOAE measurements beyond what single long-duration FFTs can do, providing it is possible to accurately estimate the noise variance and providing that the noise is CGN.

E. Comparison of confidence interval estimation methods for SOAEs

The SOAE conventional estimator suffered from bias dominated error at low SNR. The relatively low standard deviations (Fig. 7) made the biases inherent to this estimator blatant in its bootstrapped CIs (Fig. 8)—for DPOAE/SFOAEs, this bias error was obscured by larger variability errors. SOAE bias error can go undetected because magnitude values from multiple measurements are comfortably repeatable. However, repeatability alone does not guarantee a reliable estimator. Actually, it is less reliable than other estimators in the sense that one could miscalculate its confidence interval at low SNR due to consistent noise-related bias.

The Bayesian and maximum likelihood estimators' 95% CIs were estimated from the same posterior PDF [Eq. (16)] and this produced more accurate CIs, with a “percentage” that converged toward 95% before either the conventional or modified conventional estimators did (Fig. 8). To illustrate the effect of estimating noise variance, dotted curves [Fig. 8(c)] were added for the maximum likelihood, Bayesian, and modified conventional estimators from a simulation that used the known variance rather than an estimate of it. A comparison shows that estimating the variance caused a reduction of the accuracy of CIs particularly for maximum likelihood and Bayesian estimators. This was not surprising because these calculations relied heavily on the Rice PDF and its parameters. Even though estimating the noise variance decreased their accuracy, these estimators continued to outperform the conventional and modified conventional ones.

For a large number of observations, e.g., $N=1000$, the Bayesian and maximum likelihood estimators continued to be able to estimate SOAE magnitudes (and their PDFs) even when the conventional SNR was less than 1, i.e., $A/\mu_M < 1$ (Fig. 9). Thus the Rice PDF provides a method for improving SOAE magnitude estimation accuracy even when the individual measurement SNRs cannot be increased through any form of averaging (e.g., using the conventional estimator). CIs calculated from bootstrapping were misleading for high- N and low-SNR in the sense that the 95% CI never contained the true magnitude A (except when $A=0$, for which the modified conventional estimator always included it).

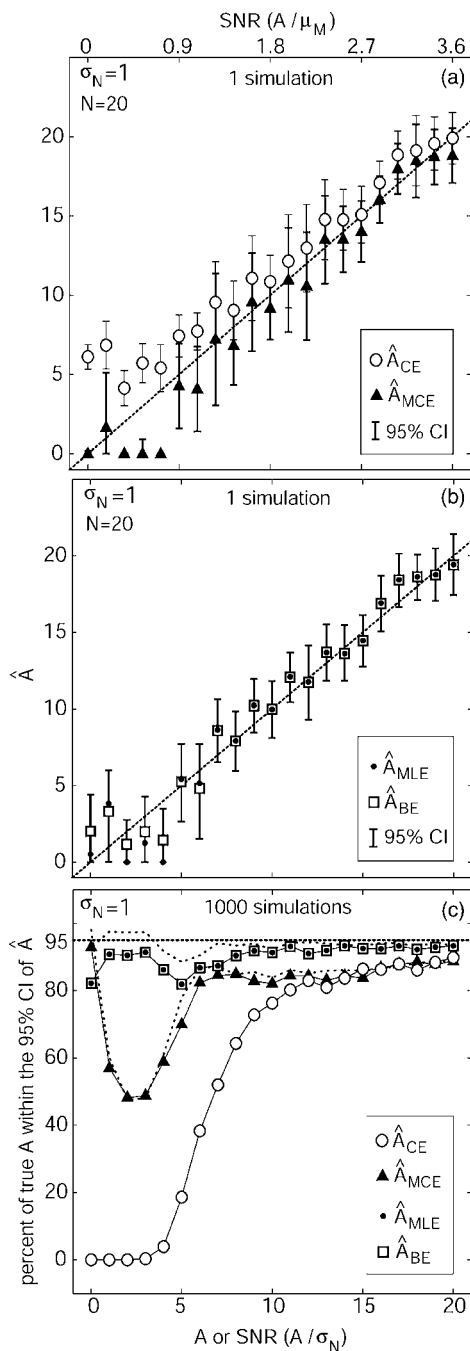


FIG. 8. Panel (a) shows one simulation for the conventional (CE) and modified conventional (MCE) SOAE magnitude estimators and their 95% confidence intervals (CIs) calculated by bootstrapping the $N=20$ magnitude observations. The dotted line $y=x$ is drawn to facilitate the comparison of true A with the CIs. Panel (b) shows the same simulation data recalculated using Bayesian (BE) and maximum likelihood (MLE) estimators. CIs for the BE and MLE were both calculated using the same posterior probability density function and were thus identical. Panel (c) shows the percentage of true A that fell within the 95% CI for each estimator after 1000 simulations (including the ones in panels a and b). Ideally this percentage would be 95% regardless of SNR, but, since CIs are estimated, errors in the CIs appear as deviations from 95%. The dotted curves in panel (c) show the result when the true measurement-noise variance was used (i.e., when it was not estimated).

IV. DISCUSSION

Calculating and reporting OAE magnitudes is a staple of OAE research, but bias in those reported values due to measurement-noise is not often discussed. The most common

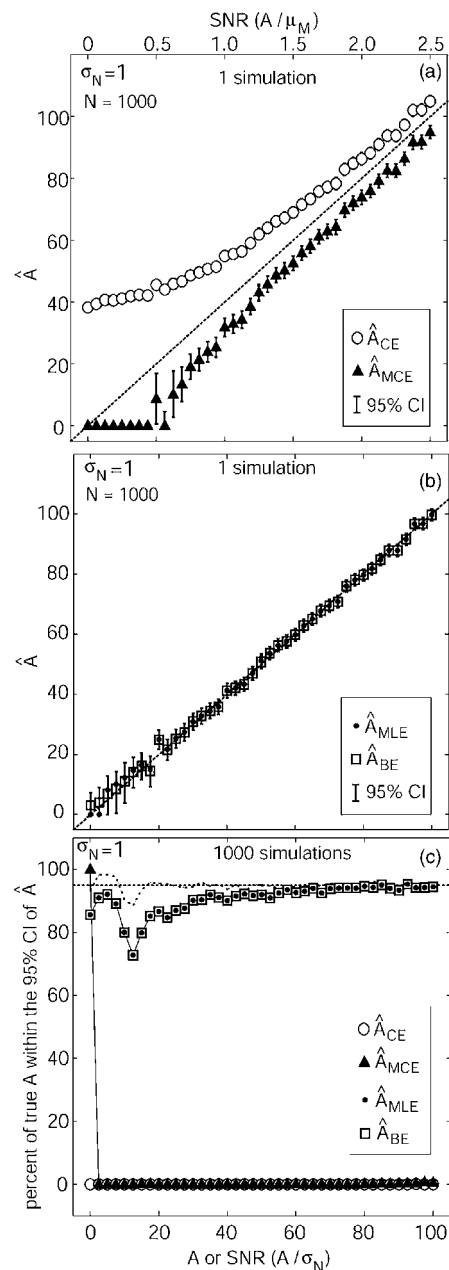


FIG. 9. Panel (a) shows one simulation for the conventional (CE) and modified conventional (MCE) SOAE magnitude estimators and their 95% confidence intervals (CIs) calculated by bootstrapping the $N=1000$ magnitude observations. The dotted line $y=x$ is drawn to facilitate the comparison of true A with the CIs. Panel (b) shows the same simulation data recalculated using Bayesian (BE) and maximum likelihood (MLE) estimators. CIs for the BE and MLE were both calculated using the same posterior probability density function and were thus identical. Panel (c) shows the percentage of true A that fell within the 95% CI for each estimator after 1000 simulations (including the ones in panels a and b). Ideally this percentage would be 95% regardless of SNR, but, since CIs are estimated, errors in the CIs appear as deviations from 95%. The dotted curve in panel (c) is the result for the BE/MLE CI when the known noise variance was used rather than its estimate; it shows how inaccuracies in those noise estimates contributed to inaccuracies in the CIs. The CE and MCE CIs calculated from bootstrapping were very inaccurate at $N=1000$ because the reduction in variance accompanying high numbers of observations exposes the bias inherent to the CE and MCE methods.

way to deal with the issue is to report OAE magnitudes only when the SNR is high and noise-related bias is low, but this solution is not always workable. For example, studies that aim to characterize an entire population must test all indi-

viduals, regardless of SNR. For these situations it behooves us to gain an understanding of how measurement-noise affects OAE magnitude estimates and how to minimize and account for those effects.

OAE measurement-noise was found to be well approximated by circular Gaussian noise (CGN) in 42/50 actual SFOAE experiments. The reason eight experiments failed the CGN criteria was unclear but could be due to artifacts such as subject movement. On the whole, it appears that a major component of OAE measurement-noise can be modeled as CGN, in other words as independent identically distributed zero-mean Gaussians adding to both the real and imaginary channels. The thrust of this paper was to address CGN and to find ways of dealing with it, particularly at low SNRs where it biases OAE magnitude estimates.

When DPOAE/SFOAE and SOAE magnitudes were estimated from simulated datasets in the presence of artificial CGN, the estimation method, regardless of type, introduced bias and variability errors due to the noise. It was often possible to trade bias error for variability error by choosing a different estimation method. For example, the DPOAE/SFOAE conventional estimator had larger bias errors but smaller variability errors at low SNR than did the maximum *a posteriori* estimator (Fig. 5). The ratio of bias/variability errors affected CIs calculated via bootstrapping—a high ratio reduced CI accuracy (Figs. 8 and 9). Thus, in situations where CIs are paramount and are calculated via bootstrapping, bias error should be preferentially reduced. However, since CIs calculated via posterior probabilities are not affected by this ratio and are more realistic, it makes sense to use posterior probabilities to calculate CIs at low SNR and this frees us to choose the estimator with the smallest combined bias and variability error, i.e., the smallest total rms (expected) error,

$$rms\ error = \sqrt{bias(\hat{A})^2 + standard\ deviation(\hat{A})^2}. \quad (18)$$

The total rms error provides a measure of the overall reliability of an estimator (Figs. 5 and 7) and can be used to discover the best estimator.

For DPOAE/SFOAEs, the Bayesian estimator (BE) had the lowest total rms error for SNRs (A/σ_N) between 0.75 and 2.75, but this improvement was small compared with the conventional estimator and does not warrant the extra assumptions and computational effort unless a probability distribution is sought. For SNRs >2.75 the conventional estimator (CE) had the lowest total rms error. These results did not depend upon the number of observations N .

Unlike DPOAEs and SFOAEs, SOAEs cannot be synchronously time averaged or vector averaged and this generally puts SOAEs at a SNR disadvantage. However, by using a Rice statistical model that describes how SOAE magnitude observations are expected to be distributed, the SNR can effectively be increased. For example, although the average SNR of individual SOAE observations might be $A/\mu_M=0.5$, analyzing $N=1000$ such observations via the Bayesian or maximum likelihood estimator can result in reasonable signal estimates (Fig. 9). This idea allowed the Bayesian estimator to produce the lowest total rms error at low SNR for

SOAEs, with an advantage that increased with increasing N . This technique provides a potential method for measuring very small SOAEs providing the Rice PDF model is workable, i.e., CGN assumptions are valid, and accurate noise estimates can be made.

A. Practical considerations for using the Bayesian, maximum a posteriori, or maximum likelihood estimators

In order to use the Bayesian, *maximum a posteriori*, or maximum likelihood estimator, the CGN assumptions that belie the Rice model need to be validated and the component noise variance, which enters as a parameter in the model, must be estimated. For DPOAEs and SFOAEs the component variance can be calculated directly from the real and imaginary channels. For SOAEs, however, estimating the component variance is harder as one does not have access to the real and imaginary channels. This necessitates an additional measurement, free from any SOAE signal, which can be used to calculate component variance (e.g., Appendix B). This “noise-only measurement” can be taken directly from original SOAE data by analyzing it at a nearby frequency where there is no SOAE (of course it may not always be possible to determine whether there is an SOAE there or not). Alternatively, it may be possible to use an entirely separate measurement wherein a suppressor tone is used to completely suppress the SOAE without altering the noise. Or it may be possible to briefly entrain the SOAE in order to extract the real and imaginary components.

B. “Drift” is a second source of measurement-noise

In addition to circular Gaussian noise, OAE measurements are subject to “drift,” a slowly varying change in the recorded emission that can be generated either by cochlear factors such as slowly changing MOCR activity or noncochlear factors such as an ear-phone slowly falling out or ear-canal temperature changing.

Drift can bias OAE magnitudes too. And drift bias can be large compared to CGN-generated bias. For example, in the noise data of Fig. 2 it is hard to see the amount of drift by eye, but an estimate of signal magnitude using the conventional estimator reveals a signal of $0.941\ \mu\text{Pa}_{\text{rms}}$ with a 95% CI of [0.71, 1.56]. This signal is greater than what would be expected from a conventional estimate of CGN alone, $0.56\ \mu\text{Pa}_{\text{rms}}$ —as estimated from a simulation using artificial pure CGN with the same noise variance. Therefore, for the data of Fig. 2 there is evidence of drift in addition to CGN, and this drift is a comparable contributor to the overall potential bias. It is, however, not known whether this drift was due to cochlear sources that should be considered a signal or a noise source like a changing ear probe fit.

The CGN assumption enables a distinction between CGN and drift and it enables one to quantify their individual contributions. This comparison can be done at even lower SNR than was done above. The problem then becomes to untangle the cochlear factors of drift from noncochlear ones.

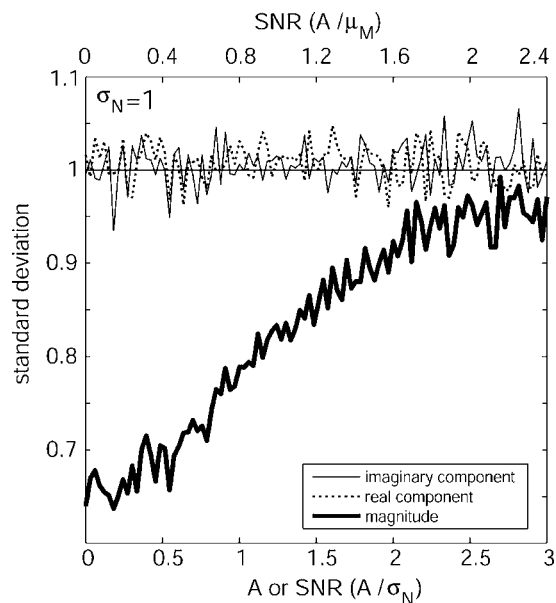


FIG. 10. Standard deviations (SDs) of the real and imaginary components and their associated magnitudes versus SNR from a simulation. The figure shows how magnitude SDs can decrease with decreasing signal strength even if component SD is held constant. The simulation is of a constant signal with variable magnitude A embedded in circular Gaussian noise with a fixed standard deviation, $\sigma_N=1$.

C. MOCR activation and cochlear noise sources

A paired t -test and a KS-test over all 50 experiments showed that MOCR activation did not change OAE component (real and imaginary) variances. This finding ostensibly conflicts with a previous report (Maison *et al.*, 1997) that showed that MOCR activation reduced the variance of the magnitudes of TEOAEs for subjects who had strong MOCR suppression. The discrepancy can be reconciled by a simulation that reveals that magnitude variances ought to decrease as the signal magnitude, A , approaches 0, even if the component variances do not change (Fig. 10). This is due to the boundary at 0 for magnitudes. This explanation seems complete considering Maison *et al.* found the magnitude variance reduction only for subjects who had the most MOCR induced suppression (i.e., A closer to 0). It appears that this effect is a measurement-noise effect and therefore MOCR activation does not necessarily change noise sources within the cochlea as suggested. Since measurement-noise comes from both biological sources and equipment, it is possible that Maison *et al.*'s finding is a result of equipment-related noise—noise that is not expected to vary with MOCR activation. In any event, there was no evidence here to support the notion that MOCR activation altered biological noise sources.

D. Advantages of using the Rice PDF

Single number OAE magnitude estimates are crude measures containing bias and variability errors. Repeating these calculations via bootstrapping allows one to add confidence intervals, but CIs calculated this way are subject to the same bias errors as the original estimates. The simulated results indicate that we can calculate OAE CIs more accurately us-

ing posterior probabilities from Rice PDFs even at low SNR when CGN measurement-noise assumptions are valid. This suggests that those posterior probabilities can be used to represent OAE magnitude data. This is a novel idea. There are times when it would make sense to work with probability density functions of OAE magnitudes, rather than the magnitude estimates themselves. For example, PDFs from many subjects can be combined to construct a continuous, as opposed to binned, population histogram. Or they can be used to make continuous probabilistic (rather than point by point deterministic) correlations between OAE magnitudes and psychophysical measures, perhaps enhancing the power of such correlations.

E. Implications for estimating transiently evoked OAEs

Measurement-noise bias is also a problem for estimating transient-evoked OAE magnitudes. The reason TEOAEs are not specifically addressed in this manuscript is that they decay—unlike DPOAEs, SFOAEs, and SOAEs, which can be constant. It is more appropriate to use wavelet analysis (or other time-restricted basis functions) than Fourier analysis for these rapidly decaying signals (Pasanen *et al.*, 1994; Jedrzejczak *et al.*, 2004). Another difference is that the concept of a “phase reference,” which is appropriate for pure tone stimuli, does not apply to the click stimuli used for TEOAEs where a “delay reference” is more apt. Even so, there are relevant ideas in this work for estimating TEOAE magnitudes.

At low stimulus levels TEOAEs appear to be a superposition of SFOAEs (Kalluri and Shera, 2004). These SFOAEs are evoked together by a single impulse stimulus and therefore there is no available phase reference although they can be synchronously averaged. Under these conditions, estimating a TEOAE magnitude can involve the same calculations as estimating the (average) magnitudes of a large number of (decaying) SOAEs and then combining them.¹³ TEOAEs suffer from the same problems as SOAEs, namely low SNR at some frequencies and no phase reference. It has also been observed that sometimes a single strong SFOAE or SOAE can dominate the TEOAE response. Given these characteristics, it is not surprising to find evidence of Rice-like distributions within reported TEOAE data (De Ceulaer *et al.*, 2001; Jedrzejczak *et al.*, 2005).

Currently, there are a number of magnitude estimation methods for TEOAEs, each of which has its own bias and variability error. While some of these methods use Fourier analysis to analyze separate frequency regions, none are motivated by an understanding of the measurement-noise issue presented here; adapting the idea from the SOAE method could help, particularly for constructing realistic confidence intervals and/or probability distributions.

F. Conclusion

All magnitude estimation methods, including those not presented here, introduce bias and variability errors due to noise—errors that often go unappreciated. These errors are largest at low SNRs and can lead to the misinterpretation of

data. Several OAE estimation methods put forth here can be used to estimate OAE magnitudes and their 95% CIs more reliability than is commonly done. Moreover, these methods can be used to better appreciate the effects of noise.

The most common way to estimate DPOAE/SFOAE magnitudes is by using simple averaging termed *conventional estimators* [Eqs. (1) and (10)]. Such conventional estimators work well when the SNR is large; however, at low SNR they introduce bias and variability errors that are not always appreciated. Here, three other estimators, (1) a modified conventional estimator, (2) a maximum *a posteriori* estimator or maximum likelihood estimator, and (3) a Bayesian estimator, were investigated to see whether the noise-induced errors could be reduced and/or quantified. The new estimators all use the knowledge that OAE measurements contain circular Gaussian noise (see Sec. III A) in order to model the noise + signal as a Rice distribution (Appendix A). Each one takes advantage of this knowledge in a different way. One of the new estimators, the Bayesian DPOAE/SFOAE estimator, did introduce less total rms error for SNRs <6.8 dB than the conventional estimator (Fig. 5). But, more importantly, modeling the noise allowed accurate probability distributions to be calculated [Eq. (6)]. By representing OAE magnitudes as probability distributions, the errors involved are quantified and exposed. This representation could be useful when low SNR measurements are unavoidable—where single-valued magnitude estimates are known to contain large errors.

For SOAEs, two of the new estimators, the maximum likelihood and Bayesian estimators, enabled accurate estimates at very low SNR where current techniques—e.g., using longer duration recordings and FFTs, and/or spectral fitting (van Dijk and Wit, 1990; Talmadge *et al.*, 1993)—fail to do so. This is achieved by optimally combining multiple magnitude observations; each of which may have an SNR <1 [Eq. (15) or (17)]. This works providing one can estimate the measurement-noise variance (see Appendix B).

The errors addressed here are those that arise from measurement-noise; however, “drift” is another source of error, one that can be equally large or larger. Drift can be cochlear in origin or it can arise from an artifact, such as the ear phone slowly falling out. Separate efforts should be made to monitor and control for drift.

ACKNOWLEDGMENTS

Thanks to John J. Guinan for his valuable input. This work was supported in part by NICDC R01 DC005977, P30 DC005209.

Nomenclature

BE	= Bayesian estimator
CAS	= contralateral acoustic stimulation
CE	= conventional estimator
CGN	= circular Gaussian noise (zero-mean Gaussian noise adding to the real and imaginary components independently)
CI	= 95% confidence interval
DPOAE	= distortion-product otoacoustic emission

FFT	= fast Fourier transform
KS-test	= Komolgorov-Smirnov statistical test
MAPE	= maximum <i>a posteriori</i> estimator
MCE	= modified conventional estimator
MLE	= maximum likelihood estimator
MOCR	= medial olivocochlear reflex
OAE	= otoacoustic emission
PDF	= probability density function
rms	= root-mean-square
SFOAE	= stimulus-frequency otoacoustic emission
SD	= standard deviation
SNR	= signal/noise ratio, A/σ , A/σ_N or A/μ_M
SOAE	= spontaneous otoacoustic emission
TEOAE	= transiently evoked otoacoustic emission
\mathbf{A}	= signal constant (complex valued)
A	= true magnitude of signal constant
\hat{A}	= estimated magnitude of signal constant
ϕ	= phase of signal constant
I	= imaginary component of measurement variable
I_0	= modified Bessel function of the first kind
\mathbf{M}	= measurement variable (complex valued)
M	= magnitude of measurement variable
μ	= mean of Gaussian distributed random variable (e.g., R and I)
μ_M	= mean of Rayleigh distributed magnitudes
R	= real component of measurement variable
σ	= component standard deviation (the SD of the real and imaginary components) of circular Gaussian measurement-noise
$\hat{\sigma}$	= estimated component standard deviation of circular Gaussian measurement-noise
σ_M	= standard deviation of Rayleigh distributed magnitudes
σ_N	= component standard deviation of circular Gaussian measurement-noise after synchronous vector averaging N observations
$\hat{\sigma}_N$	= estimated component standard deviation of circular Gaussian measurement-noise after synchronous vector averaging N observations
θ	= phase of measurement variable

APPENDIX A: PROOF OF THE APPLICABILITY OF THE RICE PDF TO THE CONDITION OF A CONSTANT COMPLEX-VALUED SIGNAL EMBEDDED IN CIRCULAR GAUSSIAN NOISE

The Rice distribution applies to a constant complex signal embedded in circular Gaussian noise (CGN). The proof of this has been addressed in the literature for magnetic resonance imaging, for example in the Ph.D. thesis of Jans Sijbers (Sijbers, 1998), and in textbooks, but it has not yet been applied to OAE data. The following proof is included to demonstrate the applicability of the Rice PDF to OAE data and to show the functional form of the Rice distribution and its shape (Fig. 11).

Assume a complex-valued constant signal, \mathbf{A} , has magnitude, A , and phase, ϕ . If this signal is measured in the

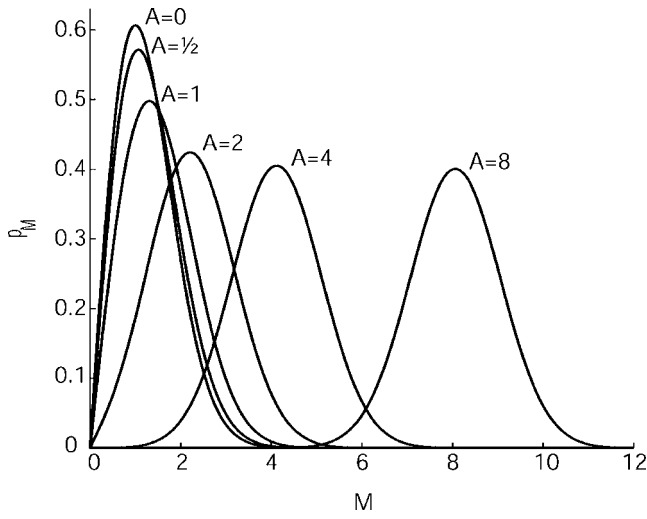


FIG. 11. Rice probability density functions (PDFs) for different values of parameter A . The distributions can be used to model the probability of observing magnitude M (linear units) from a DPOAE, SFOAE, or SOAE measurement containing an OAE signal with magnitude= A embedded in circular Gaussian noise with standard deviation=1.

presence of CGN with variance σ^2 , the probability distributions of the hypothetically measured real, R , and imaginary, I , components respectively,

$$p_R(R|\mathbf{A}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(R - \text{Re}\{\mathbf{A}\})^2/2\sigma^2}, \quad (\text{A1})$$

$$p_I(I|\mathbf{A}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(I - \text{Im}\{\mathbf{A}\})^2/2\sigma^2}, \quad (\text{A2})$$

are Gaussians (see Fig. 4) with means shifted from zero. Assuming the components are *independent*, their joint probability is

$$p_{R,I}(R, I|\mathbf{A}, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(R - \text{Re}\{\mathbf{A}\})^2/2\sigma^2} e^{-(I - \text{Im}\{\mathbf{A}\})^2/2\sigma^2}. \quad (\text{A3})$$

Rewriting Eq. (B3) in polar form allows the magnitude and phase of measurement variable \mathbf{M} (magnitude= M and phase= θ) to replace the real and imaginary components. Complex signal \mathbf{A} is also converted to polar notation:

$$p_{M,\theta}(M, \theta|A, \phi, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(M \cos \theta - A \cos \phi)^2/2\sigma^2} \times e^{-(M \sin \theta - A \sin \phi)^2/2\sigma^2}, \quad (\text{A4})$$

equivalently,

$$p_{M,\theta}(M, \theta|A, \phi, \sigma) = \frac{1}{2\pi\sigma^2} e^{-[M^2 + A^2 - 2AM \cos(\theta - \phi)]/2\sigma^2}. \quad (\text{A5})$$

Integrating over θ , using the arc-length unit of integration $M \cdot d\theta$ for polar coordinates, generates the probability of observing magnitude, M , between M and dM :

$$p_M(M|A, \phi, \sigma) = \int_0^{2\pi} \frac{1}{2\pi\sigma^2} e^{-[M^2 + A^2 - 2AM \cos(\theta - \phi)]/2\sigma^2} \times M d\theta. \quad (\text{A6})$$

$$= \frac{M}{2\pi\sigma^2} e^{-\frac{M^2 + A^2}{2\sigma^2}} \times \int_0^{2\pi} e^{-[-2AM \cos(\theta - \phi)]/2\sigma^2} d\theta. \quad (\text{A7})$$

When integrating over all θ the constant offset ϕ is not a factor, thus

$$p_M(M|A, \sigma) = \frac{M}{2\pi\sigma^2} e^{-(M^2 + A^2)/2\sigma^2} \int_0^{2\pi} e^{(AM \cos \theta)/\sigma^2} d\theta, \quad (\text{A8})$$

equivalently,

$$p_M(M|A, \sigma) = \frac{M}{\sigma^2} e^{-(M^2 + A^2)/2\sigma^2} I_0\left(\frac{AM}{\sigma^2}\right), \quad (\text{A9})$$

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos \theta} d\theta, \quad (\text{A10})$$

where I_0 is a modified Bessel function of the first kind. This is the probability that an observed magnitude M lies between M and dM given a signal with magnitude A is being measured in the presence of circular Gaussian noise with variance σ^2 . This probability distribution is called a Rice PDF (Fig. 11).

APPENDIX B: A METHOD FOR ESTIMATING THE UNDERLYING COMPONENT GAUSSIAN NOISE VARIANCE FOR SOAE MEASUREMENTS

Estimating the magnitudes of SOAEs (and TEOAEs) by using an estimator that invokes the Rice distribution (i.e., using a modified conventional, maximum likelihood or maximum *a posteriori* estimator) requires knowledge of the underlying variance of the circular Gaussian noise, which enters as a parameter in the Rice distribution (see Appendix A). However, since no external phase reference is available for SOAEs, the real and imaginary components are not directly measured and their component variance cannot be computed directly. A practical method for estimating the component variance is needed.

In the absence of a signal, i.e., $A=0$, the Rice distribution [Eq. (B9)] devolves to a Rayleigh distribution [Fig. 3(b)],

$$p_M(M|\sigma) = \frac{M}{\sigma^2} e^{-M^2/2\sigma^2}. \quad (\text{B1})$$

Two properties of the Rayleigh [Eqs. (B2) and (B3)] provide methods for calculating the underlying variance when only magnitude data is known:

$$\mu_M^2 = \frac{\pi}{2} \sigma^2, \quad (\text{B2})$$

$$\sigma_M^2 = \left(\frac{4 - \pi}{2} \right) \sigma^2, \quad (\text{B3})$$

$$\frac{\mu_M^2}{\sigma_M^2} = \frac{\pi}{4 - \pi}. \quad (\text{B4})$$

The underlying Gaussian variance, σ^2 , can be calculated from either the mean of the magnitude data in the absence of a signal [from Eq. (B2)] or the variance of that data [from Eq. (B3)]. The extent to which these two calculations agree [the extent to which Eq. (B4)] is true] gives an indication of the validity of the initial assumptions: (1) that no signal is present and (2) the measurement is taking place in zero-mean equal-variance Gaussian noise. A chi square test could also be used to determine if the distribution is consistent with a Rayleigh. Practically, the needed SOAE “no signal” measurement can be made either from a nearby frequency with no SOAE or perhaps by measuring while suppressing the SOAE with a tone.

As an example, pretend the data in Fig. 3(b) came from an SOAE noise measurement, i.e., no SOAE present. Equation (B4) left side is 3.19 versus the standard Rayleigh value of 3.66 (right side) and the two estimates of the noise standard deviation would be 2.51 from Eq. (B2) and 2.69 from Eq. (B3). Their average 2.60 would be a reasonable estimate. The actual calculated values from the real and imaginary components were 2.54 and 2.39, respectively (shown in Fig. 2).

In this way, underlying Gaussian component variances can be practically estimated for use in one of the Rice-based SOAE magnitude estimators (see Sec. II B 2).

APPENDIX C: PROOF THAT THE 2D MAXIMUM LIKELIHOOD ESTIMATOR FOR DPOAE/SFOAE MAGNITUDES IS EQUIVALENT TO THE CONVENTIONAL ESTIMATOR

The 2D maximum likelihood estimator maximizes the likelihood of the complex-valued observations, $\mathbf{M}_n = R_n + iI_n$, over the unknown model parameters, A , ϕ :

$$L = \prod_{n=1}^N \frac{1}{2\pi\sigma^2} e^{-(R_n - A \cos \phi)^2 / 2\sigma^2} e^{-(I_n - A \sin \phi)^2 / 2\sigma^2}. \quad (\text{C1})$$

The problem reduces to maximizing Λ :

$$\Lambda = \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N R_n^2 + I_n^2 - \dots - 2R_n A \cos \phi - 2I_n A \sin \phi + A^2, \quad (\text{C2})$$

$$\hat{A}_{2DMLE} = \operatorname{argmax}_A(\Lambda). \quad (\text{C3})$$

Taking the derivative with respect to A ,

$$\frac{d\Lambda}{dA} = \frac{1}{\sigma^2} \sum_{n=1}^N R_n \cos \phi + I_n \sin \phi - A. \quad (\text{C4})$$

Setting the derivative to 0 and solving for A ,

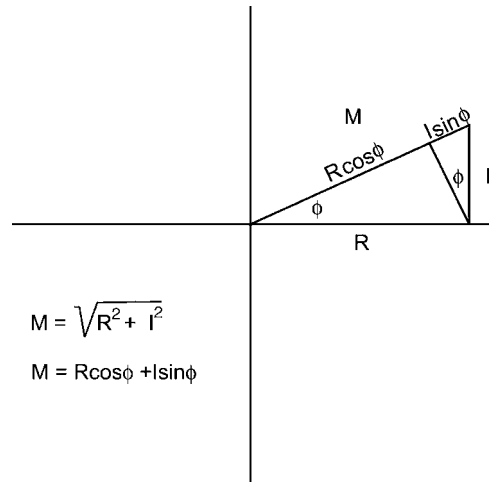


FIG. 12. A geometric argument showing that the magnitude of a hypotenuse, M , can be calculated with or without knowledge of the angle ϕ .

$$\hat{A}_{2DMLE} = \frac{1}{N} \sum_{n=1}^N R_n \cos \phi + I_n \sin \phi. \quad (\text{C5})$$

A geometric argument (Fig. 12) shows that this calculation is equivalent to the conventional estimator,

$$\hat{A}_{CE} = \left| \frac{1}{N} \sum_{n=1}^N \mathbf{M}_n \right|. \quad (\text{C6})$$

¹The stimulus did not need to be bilateral, but bilateral stimulation was used for comparison with prior work.

²It is possible to “entrain” pre-existing SOAEs with an external stimulus and thereby provide an external phase reference (Van Dijk and Wit, 1988). It is a matter of semantics whether this emission is considered a SOAE or SFOAE. Here such entrained SOAEs are considered SFOAEs.

³The conventional estimator is equivalent to using a 2D maximum likelihood estimator to estimate parameter A from the N complex data points (see Appendix C).

⁴Although an estimate of A^2 using Eq. (3) is unbiased, the square root function renders the corresponding estimate of A biased.

⁵There was no noticeable difference in the result from using the estimated variance, $\hat{\sigma}^2$, versus the true variance, σ^2 (Fig. 8). This was because the $N=1000$ samples were enough to provide an accurate variance estimate.

⁶In this case, the *maximum a posteriori* estimator is equivalent to the maximum likelihood estimator. The *maximum a posteriori* estimator formulates the solution via Rice PDFs and allows for the inclusion of *a priori* information about A . The maximum likelihood estimator also invokes a Rice PDF but does not have the capability to include *a priori* information. Since no *a priori* information is used here, these estimators are equivalent. In other words, the maximum likelihood estimator is a subset of the *maximum a posteriori* estimator; they are equivalent when all values of A are *a priori* considered equally likely.

⁷Like the *maximum a posteriori* estimator, the Bayesian estimator could have included prior information about A . For example, information from a prior population study could be included. Here, only information about A from the actual observations was included.

⁸The optimal FFT length for these buffers is a complicated function of factors including the subject’s ability to sit quietly. That optimization is best done by trial and error.

⁹The maximum of Eq. (16), after all M_n had updated it, was a *maximum a posteriori* estimate of A that was numerically identical to the maximum likelihood estimate from Eqs. (14) and (15).

¹⁰For the KS-test, the real and imaginary data were compared with the same zero-mean Gaussian whose standard deviation (SD) was the mean of the component SDs. P values from the two (real and imaginary) KS-tests were multiplied to give a joint P value. A joint $P < 0.05$ indicates that the real

and imaginary components were *not* likely to have come from this same Gaussian distribution.

¹¹A covariance term whose magnitude is small compared with the individual variances indicates independence only for the special case of Gaussian distributions. The covariance matrix was

$$\text{cov}(R, I) = \begin{bmatrix} 6.4 & -0.5 \\ -0.5 & 5.7 \end{bmatrix}.$$

¹²There is a second-order dependence on N that comes from the improved accuracy of the estimate of σ_N with increasing N . This effect can be considered negligible for $N > 20$.

¹³It is more appropriate to combine energy terms (magnitude squared) as energy (not magnitude) is a conserved quantity. It also makes sense to forgo the combining and leave the frequency specificity inherent to this analysis intact.

Backus, B. C., and Guinan, J. J. (2006). "Time course of the human medial olivocochlear acoustic reflex," *J. Acoust. Soc. Am.*

De Ceulaer, G., Yperman, M., Daemers, K., Van Driessche, K., Somers, T., Offeciers, F. E., and Govaerts, P. J. (2001). "Contralateral suppression of transient evoked otoacoustic emissions: normative data for a clinical test set-up," *Otol. Neurotol.* **22**, 350–355.

Guinan, J. J., Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). "Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," *J. Assoc. Res. Otolaryngol.* **4**, 521–540.

Jedrzejczak, W. W., Blinowska, K. J., and Konopka, K. (2005). "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise," *Hear. Res.* **205**, 249–255.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (2004). "Identification of otoacoustic emissions components

by means of adaptive approximations," *J. Acoust. Soc. Am.* **115**, 2148–2158.

Kalluri, R., and Shera, C. A. (2004). "The Relationship Between TEOAEs and SFOAEs at Low Stimulus Levels," in *Proceedings of the ARO Mid-winter Meeting*, New Orleans, LA, Vol. **27**, pp. 180–181.

Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.

Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (2001). "Adaptation of distortion product otoacoustic emission in humans," *J. Assoc. Res. Otolaryngol.* **2**, 31–40.

Maison, S., Micheyl, C., Chays, A., and Collet, L. (1997). "Medial olivocochlear system stabilizes active cochlear micromechanical properties in humans," *Hear. Res.* **113**, 89–98.

Pasanen, E. G., Travis, J. D., and Thornhill, R. J. (1994). "Wavelet-type analysis of transient-evoked otoacoustic emissions," *Biomed. Sci. Instrum.* **30**, 75–80.

Sijbers, J. (1998). "Signal and Noise Estimation from Magnetic Resonance Images," in *Natuurkunde*, vol. Doctor in de Wetenschappen (Universitaire Instelling Antwerpen, Antwerpen), p. 95.

Sijbers, J., den Dekker, A. J., Scheunders, P., and Van Dyck, D. (1998). "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Trans. Med. Imaging* **17**, 357–361.

Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (1993). "New off-line method for detecting spontaneous otoacoustic emissions in human subjects," *Hear. Res.* **71**, 170–182.

Van Dijk, P., and Wit, H. P. (1988). "Phase-lock of spontaneous oto-acoustic emissions to a cubic difference tone," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (London Academic, London), pp. 101–105.

van Dijk, P., and Wit, H. P. (1990). "Amplitude and frequency fluctuations of spontaneous otoacoustic emissions," *J. Acoust. Soc. Am.* **88**, 1779–1793.

Modeling binaural loudness

Brian C. J. Moore^{a)} and Brian R. Glasberg^{b)}

*Department of Experimental Psychology, University of Cambridge, Downing Street,
Cambridge CB2 3EB, England*

(Received 13 July 2006; revised 6 December 2006; accepted 6 December 2006)

A survey of data on the perception of binaurally presented sounds indicates that loudness summation across ears is less than perfect; a diotic sound is less than twice as loud as the same sound presented monaurally. The loudness model proposed by Moore *et al.* [J. Audio Eng. Soc. **45**, 224–240 (1997)] determines the loudness of binaural stimuli by a simple summation of loudness across ears. It is described here how the model can be modified so as to give more accurate predictions of the loudness of binaurally presented sounds, including cases where the sounds at the two ears differ in level, frequency or both. The modification is based on the idea that there are inhibitory interactions between the internal representations of the signals at the two ears, such that a signal at the left ear inhibits (reduces) the loudness evoked by a signal at the right ear, and vice versa. The inhibition is assumed to spread across frequency channels. The modified model gives reasonably accurate predictions of a variety of data on the loudness of binaural stimuli, including data obtained using loudness scaling and loudness matching procedures. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2431331]

PACS number(s): 43.66.Cb, 43.66.Pn, 43.66.Ba, 43.66.Yw [AJO]

Pages: 1604–1612

I. INTRODUCTION

It is well established that if a sound is presented to both ears, the perceived loudness is greater than when the sound is presented to one ear only. This effect is often called “binaural loudness summation.” Some early data appeared to be consistent with the idea that the loudness evoked by presentation of a sound to both ears corresponded to a simple sum of the loudness (in sones) evoked by the sound at each ear separately (Fletcher and Munson, 1933; 1937; Hellman and Zwislocki, 1963; Levelt *et al.*, 1972; Marks, 1978; Algom *et al.*, 1989a). In the case of diotic presentation (the same sound at each ear), such a rule would mean that the binaural loudness was simply double the monaural loudness. For midfrequency pure tones at moderate sound levels, a doubling of loudness is produced by approximately a 10-dB increase in sound level (Stevens, 1957; Hellman, 1976). Thus, for equal loudness, a monaural tone would have to be 10 dB higher in level than a diotic tone. Again, some early data were roughly consistent with this prediction (Hellman and Zwislocki, 1963), although Reynolds and Stevens (1957) found a 10-dB monaural-diotic difference only at high sound levels; the difference was smaller at low levels. For simplicity, the assumption of simple summation of loudness across ears was adopted for the model of loudness developed by Moore *et al.* (1997), which forms the basis for a recent ANSI standard (ANSI, 2005).

More recent data, and some older data, have suggested less-than-perfect summation of loudness across ears (Irwin, 1965; Scharf, 1969; Scharf and Fishken, 1970; Gigerenzer and Strube, 1983; Algom *et al.*, 1989b; Zwicker and Zwicker, 1991). For example, Scharf and Fishken (1970)

presented data on loudness scaling (magnitude estimation and magnitude production) for sounds presented monaurally and diotically. They found diotic/monaural loudness ratios between 1.4 and 1.7. On the basis of magnitude estimation experiments, Zwicker and Zwicker (1991) proposed that the diotic/monaural loudness ratio was about 1.5. Culling and Edmonds (2007) reported that for medium overall levels, a monaural noise band had to have a level about 5–6 dB higher than for the same band presented diotically to achieve equal loudness. An effect of similar magnitude was reported by Scharf (1969); a dichotic tone pair (with one tone at each ear, each having the same loudness) had a loudness level that was typically 5–7 phons greater than the loudness level of either of the components presented monaurally. Since a 5-dB increase in level corresponds to about a 1.4-fold increase in loudness, while a 7-dB increase corresponds to a 1.7-fold increase, these results imply a diotic/monaural loudness ratio between 1.4 and 1.7. Whilby *et al.* (2006) reported that the level difference between equally loud monaural and diotic tones varied with level, being about 5 dB for low and high levels, and about 8–9 dB for medium levels. The decrease at high levels appears to be inconsistent with the results of Reynolds and Stevens (1960), who found an increase at high levels. In any case, the data of Whilby *et al.* suggest that the diotic/monaural loudness ratio is about 1.8.

Recently, Sivonen and Ellermeier (2006) described the results of loudness-matching experiments for 1/3-octave wide bands of noise presented from various locations within an anechoic chamber. The subjects were required to adjust the level of a sound from a nonfrontal direction so as to match the loudness of the same sound presented from a frontal direction. The relative sound level at the two ears varied over a wide range with the position of the sound source, especially for sounds centered at high frequencies. Their results were not consistent with the idea that loudness was

^{a)}Electronic mail: bcjm@cam.ac.uk

^{b)}Electronic mail: bgl2@cam.ac.uk

simply summed across ears. Rather, the results suggested much less-than-perfect loudness summation. Based on an assessment of how well various models could account for their results, they proposed that the difference in level required for equal loudness of monaural and diotic sounds was only about 3 dB, corresponding to a diotic/monaural loudness ratio of 1.22.

Comparable data were reported by Keen (1972) for headphone presentation. He presented a reference diotic stimulus with a level of 70 dB SPL at each ear, followed by a test stimulus with a fixed level in one ear (say the right) and an adjustable level in the other ear (say the left). Subjects were required to match the overall loudness of the reference and test stimuli. When the stimulus was a 1-kHz pure tone, some subjects could not achieve a match when the level of the test tone in the right ear was fixed at 76 dB SPL, since even when the level of the adjustable tone in the left ear was made very low, the test stimulus appeared louder than the reference stimulus. This suggests that the difference in level between equally loud monaural and diotic tones is less than 6 dB for some subjects, which is not compatible with simple summation of loudness across ears.

Clearly, there are large differences across studies in the reported amount of binaural loudness summation, and in the reported difference in level between equally loud monaural and diotic sounds. These differences across studies probably arose partly from differences in methodology and perhaps partly from individual differences across subjects. However, the results of most studies, especially those using modern psychophysical methods, lead to the conclusion that binaural loudness summation is less than perfect, and that the diotic/monaural loudness ratio is less than 2. In this paper, we describe how the loudness model proposed by Moore *et al.* (1997) can be modified so as to predict less-than-perfect loudness summation across ears. The modification is based upon the assumption that a strong input to one ear can inhibit the internal response evoked by a weaker input to the other ear. This assumption has been used in some models of sound localization and binaural unmasking (Breebaart *et al.*, 2001). In the context of binaural loudness perception, the concept has been called contralateral binaural inhibition (Gigerenzer and Strube, 1983). Gigerenzer and Strube presented loudness-perception data supporting this concept, but they did not attempt to model the form of the binaural inhibition.

II. IMPLEMENTATION OF BINAURAL INHIBITION

The revised model described here is based on the model described by Moore *et al.* (1997), which was in turn based on concepts derived from earlier models (Fletcher and Munson, 1933; Zwicker and Scharf, 1965). For details, the reader is referred to the earlier papers and to Glasberg and Moore (2006). The model of Moore *et al.* (1997) consists of the following stages:

- (a) A fixed filter representing transfer through the outer ear. This filter can differ depending on the listening conditions (free field, diffuse field, or headphone presentation).
- (b) A fixed filter representing transfer through the middle ear.
- (c) Calculation of an excitation pattern from the physical

spectrum (Moore and Glasberg, 1983; Glasberg and Moore, 1990). The excitation pattern is calculated on an ERB_N number scale, where ERB_N stands for the equivalent rectangular bandwidth of the auditory filter, as determined using young normally hearing listeners at moderate sound levels (Glasberg and Moore, 1990; Moore, 2003). The excitation level is calculated at intervals of 0.1 ERB_N .

(d) Transformation of the excitation pattern to a specific loudness pattern. Specific loudness is a kind of loudness density and has units sones/ ERB_N .

(e) Determination of the area under the specific loudness pattern, which gives the overall loudness for a given ear in sones.

Stages (a)–(e) are calculated separately for each ear. In the original version of the model, the loudness was then simply summed across ears for a binaurally presented sound. In the revised model, stages (a)–(d) remain the same [except for a change of one constant in stage (d), as described later], but subsequent stages differ.

Let $S_L(E)$ and $S_R(E)$ be the specific loudnesses evoked at the left and right ears, respectively, at a given ERB_N number E . It is assumed that there are inhibitory interactions between the two ears, such that a signal at the left ear inhibits (reduces) the loudness evoked by a signal at the right ear, and vice versa. It is assumed further that these inhibitory interactions are relatively broadly tuned. For example, a large value of $S_L(E)$ at a specific E value E_k may produce inhibition in the right ear over a wide range of E values around E_k . This assumption is required to explain the finding of Scharf (1969) that the loudness of a tone in one ear was reduced by a tone of the same loudness in the other ear even when the two tones were separated in frequency by several ERB_N . However, when the two tones were very widely separated in frequency, no such loudness reduction was found, so the inhibition must be tuned to some extent.

To implement the broad tuning of the inhibition, the specific loudness pattern at each ear is initially smeared or smoothed by convolution with a weighting function. Several weighting functions were tried, including a pair of back-to-back exponentials, a pair of back-to-back rounded exponentials, and a Gaussian function, and it was found that a Gaussian function gave the best fit to the data of Scharf (1969). Hence a Gaussian function (on an ERB_N number scale) was adopted. The function is described by

$$W(g) = \exp(-Bg)^2, \quad (1)$$

where g is the normalized deviation from E , i.e., $(E - \Delta E)/E$, and B is a parameter (broadening factor) determining the degree of spread of inhibition along the E scale. The value of B was set to 0.08. The smoothed specific loudness patterns at the left and right ears are denoted $S_L(E)_{\text{smoothed}}$ and $S_R(E)_{\text{smoothed}}$.

Let $\text{INHIB}_L(E)$ denote the effective factor by which the specific loudness evoked by the signal at the left ear is reduced after inhibition produced by the signal at the right ear. Let $\text{INHIB}_R(E)$ denote the effective factor by which the specific loudness evoked by the signal at the right ear is reduced after inhibition produced by the signal at the left ear. We

assume that the values of $\text{INHIB}_L(E)$ and $\text{INHIB}_R(E)$ at a given E value are determined by the relative values of $S_L(E)_{\text{smoothed}}$ and $S_R(E)_{\text{smoothed}}$. The value of $\text{INHIB}_L(E)$ should approach 1 when the ratio $S_R(E)_{\text{smoothed}}/S_L(E)_{\text{smoothed}}$ is small, and should increase as the ratio increases. The value of $\text{INHIB}_R(E)$ should vary in a similar way as a function of $S_L(E)_{\text{smoothed}}/S_R(E)_{\text{smoothed}}$. The inhibition can be modeled by the following equations:

$$\text{INHIB}_L(E) = 2/[1 + \{\text{sech}(S_R(E)_{\text{smoothed}}/S_L(E)_{\text{smoothed}})\}^p], \quad (2)$$

$$\text{INHIB}_R(E) = 2/[1 + \{\text{sech}(S_L(E)_{\text{smoothed}}/S_R(E)_{\text{smoothed}})\}^p], \quad (3)$$

where p is a constant ($p=1.5978$). To prevent problems associated with dividing by zero when $S_L(E)_{\text{smoothed}}$ or $S_R(E)_{\text{smoothed}}$ are zero, a small number (10^{-13}) is added to the values of $S_L(E)_{\text{smoothed}}$ and $S_R(E)_{\text{smoothed}}$ prior to entering them in Eqs. (2) and (3).

The gain values calculated using Eqs. (2) and (3) are applied to the original specific loudness values in each ear to give inhibited specific loudness values. Specifically, the value of $S_L(E)$ is divided by $\text{INHIB}_L(E)$ and the value of $S_R(E)$ is divided by $\text{INHIB}_R(E)$. The loudness for each ear is then calculated by summing the inhibited specific loudness values over center frequency on an ERB_N number scale, and the overall binaural loudness is obtained by summing the (inhibited) loudness values across the two ears.

The overall form of Eqs. (2) and (3), and the value of the constant p , were chosen so as to satisfy several constraints:

(1) When $S_R(E)_{\text{smoothed}}$ is increased from a very small value, keeping $S_L(E)_{\text{smoothed}}$ fixed, the decrease in inhibited specific loudness at the left ear should be smaller than the increase in inhibited specific loudness at the right ear. This is necessary to ensure that the binaural loudness increases as $S_R(E)_{\text{smoothed}}$ increases, in accord with empirical data, which show that adding a sound to one ear, with a fixed sound in the other ear, always results in an increase in loudness (Zwicker and Zwicker, 1991).

(2) When $S_R(E)_{\text{smoothed}}=S_L(E)_{\text{smoothed}}$, INHIB_L and INHIB_R should each be equal to $4/3$, so that the inhibited specific loudness at each ear is 0.75 times the uninhibited value, leading to a binaural specific loudness of $1.5S_R(=1.5S_L)$. The factor of 1.5 was chosen to give a reasonable fit to a large variety of data on loudness magnitude estimation and loudness matching of monaural and binaural stimuli. Of course, given the variability in data across studies, some data will be fitted better than others. The value chosen is consistent with loudness scaling data showing that a signal presented diotically is approximately 1.5 times as loud as that same signal presented monaurally, as shown, for example by Scharf and Fishken (1970) and Zwicker and Zwicker (1991). The value chosen is also consistent with the fact that, at least in recent data (Culling and Edmonds, 2007; Whilby *et al.*, 2006), a diotic signal is matched in loudness by a corresponding monaural signal that is typically about 5–8 dB higher in level; a 6-dB increase in level corresponds to approximately a 1.5-fold change in loudness at medium sound levels.

(3) The values of INHIB_L and INHIB_R should not increase too rapidly with increases in the ratios S_R/S_L and S_L/S_R ,

respectively. This is necessary to account for the finding that, even when the sound level at one ear is 20 dB lower than the sound level at the other ear, the overall loudness is slightly greater (by a factor of about 1.1) than when only the higher level sound is presented (Zwicker and Zwicker, 1991; Sivonen and Ellermeier, 2006).

In the loudness model described by Moore *et al.* (1997), the equations relating specific loudness to excitation include a constant C , with value equal to 0.047. Its value was chosen so that the loudness of a 1-kHz sinusoid presented binaurally in free field with frontal incidence at 40 dB SPL was exactly one sone. With the modifications to include binaural inhibition, as described above, the loudness of a binaural sound is reduced by a factor of 0.75 relative to that calculated using the original model. To restore the loudness of a 1-kHz binaural 40-dB SPL tone to one sone, the value of C in the modified model was increased by a factor of $1/0.75$ to 0.06267. With this new value of C , the modified model makes the same predictions as the original model for all diotically presented stimuli (although monaural stimuli are predicted to be slightly louder by the modified model). Thus, for diotic stimuli, predictions of changes in loudness with level and bandwidth and of equal-loudness contours remain the same as before. In what follows, we focus on predictions that are different for the original and modified models.

III. PREDICTIONS OF THE MODIFIED MODEL

A. Data of Scharf (1969) and of Algom *et al.* (1989a)

Scharf (1969) asked subjects to make loudness matches between two successively presented stimuli: (1) a single sinusoid presented to one ear with frequency f_1 ; (2) a sinusoid with frequency f_1 , presented to the same ear as in (1), together with an equally loud sinusoid with frequency f_2 , presented to the opposite ear. The separation between f_1 and f_2 was varied. For small separations, subjects reported hearing the dichotic sound (stimulus 2) as a single sound image, but for larger separations subjects reported hearing two separate tones, one at each ear. On trials where two separate tones were heard, subjects were asked to match the loudness of the two successive tones that were heard in the same ear. For example, if f_1 was 1720 Hz and was presented to the left ear, and f_2 was 2320 Hz and was presented to the right ear, then subjects would be asked to match the loudness of the single 1720-Hz tone in stimulus (1) with that of the 1720-Hz tone heard in the left ear in stimulus (2). The match was made either by adjusting the level of the single tone in stimulus (1), or by adjusting the level of both tones together in stimulus (2).

Scharf (1969) conducted the experiment using center frequencies of 500, 1000, and 2000 Hz. In a preliminary experiment, the component tones to be compared for a given center frequency were all adjusted to have the same loudness as a tone at that center frequency with a level of either 20, 50, or 80 dB SL. Scharf did not specify the values of the absolute thresholds of the subjects, so, for simplicity, we assumed that the component tones had loudness levels of 20, 50, or 80 phons. The first step in fitting Scharf's data was to

use the modified loudness model to calculate the level of each monaural component tone required to give a loudness level of 20, 50, or 80 phons. Then, the conditions of Scharf's experiment were simulated. For example, the input to the model was specified as either a monaural 1720-Hz tone with a loudness level of 50 phons, or that tone together with a 2320-Hz tone with a loudness level of 50 phons presented to the opposite ear. For one type of match, the level of the monaural tone was adjusted (reduced) until its calculated loudness level equaled the loudness level calculated for the same tone when presented in the presence of the 2320-Hz tone in the opposite ear; note that the calculated loudness here is for the ear receiving the 1720-Hz tone alone. We denote the amount of adjustment ΔL_1 . For the other type of match, both tones in the dichotic stimulus were adjusted (increased) in level by the same amount until the calculated loudness level of the 1720-Hz tone was 50 phons. We denote the amount of this adjustment ΔL_2 . The mean of the values of ΔL_1 and ΔL_2 was taken as the predicted level difference required for equal loudness of the 1720-Hz tone when presented alone and when presented together with the 2320-Hz tone in the opposite ear. This whole calculation was then repeated with the role of the two tones swapped, i.e., the single tone was at 2320 Hz instead of 1720 Hz. This gave two level adjustments, ΔL_3 and ΔL_4 , whose mean was taken as the predicted level difference required for equal loudness of the 2320-Hz tone when presented alone and when presented together with the 1720-Hz tone in the opposite ear. The mean of all four level adjustments, ΔL_1 , ΔL_2 , ΔL_3 , and ΔL_4 , was taken as the value to be compared with Scharf's data. This mean is denoted ΔL_{mean} . The predicted and obtained values of ΔL_{mean} are compared in Fig. 1. Generally, there is a very good correspondence between the obtained and predicted values, including the decrease in the values of ΔL_{mean} with increasing frequency separation of the two tones and the decrease in the values of ΔL_{mean} at the lowest loudness level. Almost all of the predicted values lie within the range of the error bars (interquartile ranges) shown by Scharf (1969).

Scharf (1969) also conducted an experiment in which subjects were asked to make matches of the overall loudness of a dichotic tone pair (different frequencies at the two ears, each component tone of the same loudness when presented monaurally) and a diotic tone pair (same level and frequency at the two ears). He reported that the overall loudness of the dichotic pair was independent of the frequency separation of the two tones. The model predicts that the loudness of a dichotic tone pair is approximately invariant with frequency separation for moderate frequency separations (for example, for a center frequency of 2000 Hz, the loudness level is predicted to increase by only 0.6 phon when the frequency separation of the dichotic tones is increased from 0 to 1220 Hz), but predicts an increase in loudness level of about 5 phons when the frequency separation is very large. This discrepancy may be connected with the fact that, for large frequency separations, Scharf's subjects reported hearing two separate tones, making it difficult to judge the overall loudness. When the frequency separation was large, on some trials, subjects probably judged the loudness of just one of the tones, rather

than judging the overall loudness of the two tones, even though they were instructed to do the latter. This is consistent with the large increase in within- and across-subject variability that Scharf found for large frequency separations and with Scharf's report that "summation was greatest for those subjects who fused most often."

Algom *et al.* (1989a) also obtained loudness judgments for monaural tones and dichotic tone pairs, but using magnitude estimation rather than loudness matching. When the dichotic tones were closely spaced in frequency (1000 and 1100 Hz), their overall loudness was less than predicted from summing the loudnesses of the monaural tones. Averaged across the levels used, the difference corresponded to a factor of 0.77 (their Fig. 2). The model predicts a factor of 0.75. When the dichotic tones were widely spaced in frequency (2000 and 5000 Hz), their overall loudness was about a factor of 0.90 less than predicted from summing the loudnesses of the monaural tones (their Fig. 7). The model predicts a factor of 0.85. Overall, the predictions of the model are highly consistent with the data of Algom *et al.*

B. Data of Zwicker and Zwicker (1991)

Zwicker and Zwicker (1991) investigated the loudness of interaurally correlated narrow- and broadband noises presented via free-field equalized headphones, using a loudness estimation paradigm. Two different reference sounds (anchors) were used, one presented monaurally and one presented diotically, and the anchor for a given set of judgments was assigned a loudness of 100 units. In one experiment, the spectra of the noises were the same at the two ears, the level at one ear was fixed, and the interaural level difference ΔL was varied ($\Delta L=0, -4, -10, -20$, and $-\infty$ dB). The noises used included 1/3-oct wide noises centered around 250, 710, or 2000 Hz, and broadband red (-10 dB/oct), pink (-3 dB/oct), and blue ($+10$ dB/oct) noises. The pattern of the results was almost independent of the center frequency, level, and spectral shape of the noise (although no levels below 40 dB SPL were used). The model correctly predicts results that are almost independent of the center frequency or spectral shape of the noise. Table I compares the predicted loudness judgments generated using a 1/3-oct wide band of noise centered at 2000 Hz with a reference level of 70 dB SPL with the median results obtained by Zwicker and Zwicker using a monaural sound as anchor (100 units of loudness). There is an excellent correspondence between the obtained and predicted judgments.

The model gives somewhat less accurate predictions of the results of a second experiment of Zwicker and Zwicker (1991) in which sounds with nonoverlapping spectra were presented to the two ears via free-field equalized earphones. The basic input signal was "uniformly exciting noise" (UEN), with the same intensity in each critical band (Zwicker and Fastl, 1999) and with an overall level of 60 dB SPL. This noise was initially filtered into 24 bands, each one critical band (one Bark) wide (Zwicker and Fastl, 1999). The bands were then grouped into wider composite bands (consisting of 1, 2, 4, or 12 successive bands) and each composite band was presented either to one ear or the other. For ex-

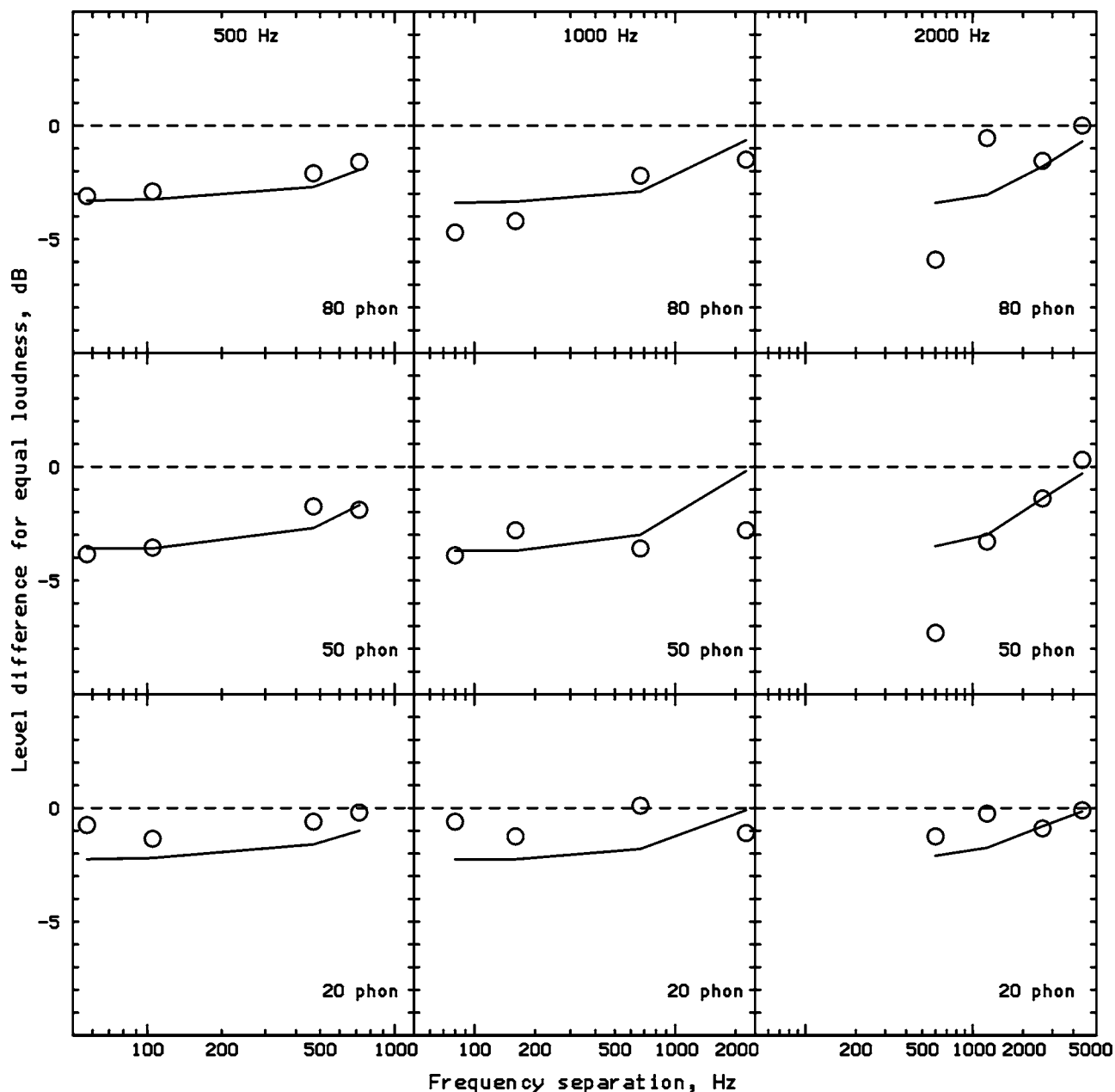


FIG. 1. Comparison of results obtained by Scharf (1969) (circles) with prediction of the modified loudness model (solid lines). The ordinate shows the level difference required for equal loudness of a tone presented monaurally, and of that same tone when presented simultaneously with a tone of different frequency in the opposite ear. The abscissa shows the frequency difference between the two tones, which were symmetrically placed around 500, 1000, or 2000 Hz (left to right columns) on a bark scale (Zwicker, 1961). Results and predictions are presented for three loudness levels of the component tones, 20, 50, and 80 phon, shown by the three rows.

TABLE I. Comparison of the predicted loudness of stimuli used by Zwicker and Zwicker (1991) with the median obtained loudness judgments, using a monaural sound as anchor (which was assigned 100 units of loudness). The sound had the same spectral shape in each ear. The sound at one ear was fixed in level, while the sound in the other ear had a relative level ΔL of 0, -4, -10, -20, and $-\infty$ dB. The obtained loudness judgments are the median across all stimulus types. Predictions were generated for a 1/3-octave-wide noise centered at 2 kHz with a fixed level in one ear of 70 dB SPL.

ΔL	0	-4	-10	-20	$-\infty$
Obtained loudness	154	139	123	110	100
Predicted loudness	150	134	121	111	100

ample, the frequency bands from 0–12 Barks would be presented to the left ear, and the bands from 13–24 Barks would be presented to the right ear. Or, the bands from 1–4, 9–12, and 17–20 Barks would be presented to the left ear and the bands from 5–8, 13–16, and 21–24 Barks would be presented to the right ear. Note that the overall level in each ear was 57 dB SPL (3 dB below the level of the unfiltered noise). The loudness of the dichotically presented sounds was judged (in separate blocks of trials) relative to either a monaural or a diotic standard consisting of all 24 bands, which was assigned 100 units of loudness.

The free-field equalizer used with the headphones employed by Zwicker and Zwicker (1991) resulted in an effective

TABLE II. Predicted loudness of stimuli used by Zwicker and Zwicker (1991), in which 1-critical-band-wide bands of noise (band number from 1 to 24) were presented in various patterns across the two ears (see text for details). Predicted loudness is given both in sones, and relative to the loudness of a diotic reference or monaural reference; in each case the reference was assigned 100 units of loudness. The median loudness judgments obtained by Zwicker and Zwicker (1991) are also shown.

Bands in left ear	Bands in right ear	Predicted loudness sones	Predicted loudness re diotic standard	Obtained loudness re diotic standard	Predicted loudness re monaural standard	Obtained loudness re monaural standard
1-12	13-24	10.6	65	90	97	120
1-4, 9-12, 17-20	5-8, 13-16, 21-24	10.3	63	85	94	115
1-2, 5-6, 9-10, 13-14, 17-18, 21-22	3-4, 7-8, 11-12, 15-16, 19-20, 23-24	11.8	72	80	108	108
1,3,5,7,9,11,13,15,17,19,21,23	2,4,6,8,10,12,14,16,18,20,22	16.9	79	75	118	105

tive frequency response at the eardrum somewhat different from that assumed for free-field presentation in the loudness model used here; specifically the response at the eardrum for the former was lower in the frequency range around 3 kHz. To allow for this, we calculated the effective transfer function of the earphones-plus-equalizer used by Zwicker and Zwicker (1991), based on data presented in Zwicker and Fastl (1999), and we used this transfer function to generate predictions of the modified model.

The predicted and obtained loudness judgments are compared in Table II. The predicted and obtained judgments are not greatly different; the largest discrepancy between the two corresponds to a factor of 1.38 in loudness (in sones), which corresponds to a change in level of about 4 dB for a broadband sound. However, for the obtained judgements the loudness decreases somewhat as the number of sub-bands composing the composite bands increases (moving from the top to bottom of the table), while the predictions show the opposite pattern. The reason for the discrepancy is not clear. It is noteworthy, however, that some aspects of the obtained loudness judgments appear strange. When all of the bands (1-24) are presented monaurally, the predicted loudness (10.9 sones) is a factor of 1.5 smaller than when they are presented diotically (16.4 sones), and this is consistent with the ratings obtained by Zwicker and Zwicker (1991). When bands 1-12 are presented to one ear only, the predicted loudness is 5.7 sones, while when bands 13-24 are presented to one ear only the predicted loudness is 5.8 sones. Since the smoothed specific loudness pattern evoked by bands 1-12 overlaps only moderately with the pattern evoked by bands 13-24, the predicted loudness when bands 1-12 are presented to one ear and bands 13-24 are presented to the other ear (10.6 sones) is only slightly less than the sum of the monaural loudness for the two composite bands ($5.7+5.8=11.5$). Also, the predicted loudness for dichotic presentation (10.6 sones) is similar to the predicted loudness for monaural presentation of bands 1-24 (10.9 sones). These predictions all seem very reasonable. Yet, the loudness ratings obtained by Zwicker and Zwicker (1991) suggest that the predicted loudness when bands 1-12 are presented to one ear and bands 13-24 are presented to the other ear is 20% greater than when all bands are presented to one ear, with no signal at the other ear. It is difficult to understand why the loudness

should be 20% greater in the former case when each of the two composite bands covers a wide frequency range and the bands do not overlap in frequency.

Overall, the model gives accurate predictions of the data obtained by Zwicker and Zwicker (1991) for cases where the spectra at the two ears were the same, and moderately accurate predictions for cases where the spectra at the two ears were markedly different (a situation that would not occur often in everyday life).

C. Data of Sivonen and Ellermeier (2006)

Sivonen and Ellermeier (2006) obtained loudness matches between sounds presented from a frontal direction and sounds presented from various locations within an anechoic chamber. The stimuli were 1/3-octave wide bands of noise with various center frequencies. The relative sound level at the two ears varied over a wide range with the position of the sound source, especially for the highest center frequency (5 kHz). The level of the sound from the nonfrontal direction was adjusted to achieve a loudness match. To predict the data of Sivonen and Ellermeier (2006) the signals reaching each ear for the fixed sound from the frontal direction were specified as input to the model and the binaural loudness was calculated. Denote the value obtained in this way $Sone_{BIN}$. Then, the input to the model was specified as corresponding to one of the sounds from a nonfrontal direction. This signal had a relative level at the two ears as specified by Sivonen and Ellermeier, and its overall level was adjusted until the model predicted a loudness equal to $Sone_{BIN}$. This input level was taken as the predicted matching level of the nonfrontal sound. The results were expressed in terms of the difference in level between the fixed and adjustable sound at the point of equal loudness, where level is measured in the absence of a listener at the center position of the listener's head. This difference was referred to by Sivonen and Ellermeier as "directional loudness sensitivity" (DLS).

Table III compares the obtained and predicted DLS values for the stimuli centered at 5000 Hz, for which the variations in interaural level were largest. Results are shown only for the higher level used. There is a reasonably good correspondence between the obtained and predicted DLS values,

TABLE III. Loudness matches obtained by Sivonen and Ellermeier (2006) for noise bands centered at 5000 Hz, and predicted loudness matches. The first column shows the azimuth of the sound source. The next two columns show the predicted and obtained “directional loudness sensitivity” (DLS) values; see text for details.

Azimuth, degrees	Predicted DLS, dB	Obtained DLS, dB
30	1.4	2.2
60	2.1	3.3
90	0.4	2.1
135	-5.9	-4.3
180	-3.9	-3.7

although there is a tendency for the predicted values to be slightly below the obtained values. The largest discrepancy is 1.7 dB.

D. Data of Keen (1972)

Keen (1972) presented a reference diotic stimulus with a level of 70 dB SPL at each ear, followed by a test stimulus with a fixed level in one ear and an adjustable level in the other ear. Subjects were required to match the overall loudness of the reference and test stimuli. We present here a comparison of the predictions of the model with the data obtained by Keen for a 1-kHz pure tone. It was assumed that the levels specified by Keen refer to levels at the eardrum. The predictions were generated in the following way. First, the loudness of the diotic reference stimulus (70 dB SPL per ear) was calculated; the obtained value was 6.82 sones. Then, a dichotic stimulus was used as input to the model, with the level fixed for one ear (at 70 dB+ ΔL) and a variable level for the other ear. The variable level was adjusted until the model predicted a binaural loudness of 6.82 sones. The variable level obtained in this way was taken as the level required to achieve a loudness match between the reference and test stimuli. Keen (1972) obtained results for eight normally hearing subjects, using $\Delta L=2, 4, 6,$ and 8 dB. However, several of the subjects were not able to make matches for the two largest values of ΔL , since the test stimulus appeared louder than the reference stimulus for all settings of the variable stimulus in the left ear. This is consistent with the predictions of the model, since a monaural 1-kHz tone with a level of 76 dB SPL has a calculated loudness of 6.78 sones, which is almost the same as the loudness of a diotic 1-kHz tone with a level of 70 dB SPL. For the two smaller values of ΔL , the matching levels in the ear receiving the adjustable tone were 67.8 and 61.0 dB SPL, for $\Delta L=2$ and 4 dB, respectively. The corresponding values predicted by the model were 67.2 and 59.8 dB SPL. Clearly, the agreement is very good.

E. Level difference for equal loudness of diotic and monaural sounds

To calculate the level difference required for equal loudness of diotic and monaural sounds, ΔLDM , the model was first used to generate the predicted loudness of a 1-kHz tone presented diotically in free field with frontal incidence, for

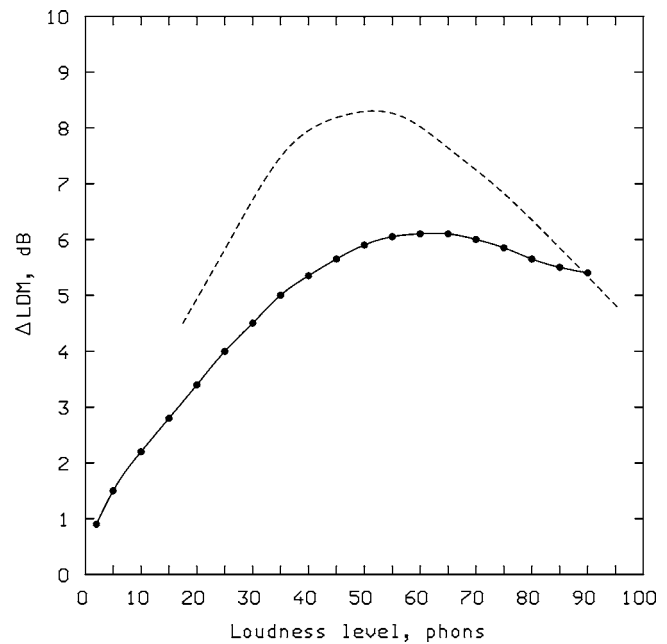


FIG. 2. The level difference required for equal loudness of a diotic and monaural 1-kHz tone, ΔLDM , for a tone presented in free field with frontal incidence. The value of ΔLDM is plotted as a function of the level of the diotic tone, which is equal to its loudness level in phons. For comparison, the dashed line shows the experimental results of Whilby *et al.* (2006).

levels from 2 dB SPL (corresponding to absolute threshold) up to 90 dB SPL. In this case, the predicted loudness level in phons is equal to the level of the tone in dB SPL. The model was then used to determine what level of a monaural 1-kHz tone, presented under the same conditions, would give the same loudness levels. For example, a diotic tone at 40 dB SPL gave a loudness level of 40 phons, and a monaural tone of 45.3 dB SPL was required to give equal loudness. Hence, in this case, the value of ΔLDM is 5.3 dB. Figure 2 shows the value of ΔLDM plotted as a function of the loudness level of the diotic tone. The value of ΔLDM is highest for levels of 60–70 dB SPL, and becomes somewhat smaller at high levels, and markedly smaller at very low levels. The figure also shows the results of Whilby *et al.* (2006) based on a smooth curve drawn by eye through their mean data. The pattern of the predictions is similar to the pattern of results found by Whilby *et al.* (2006), except that the value of ΔLDM found by Whilby *et al.* is slightly greater than predicted by the model, especially for medium levels. However, the curve predicted by the model falls within the range of the error bars (showing across-subject standard deviations) plotted by Whilby *et al.*

The value of ΔLDM varies with level because the slope of the function relating loudness in sones to level in dB SPL (the loudness function) varies with level; the steeper the slope, the smaller the value of ΔLDM . To fit the data of Whilby *et al.* (2006) more accurately, it would be necessary to modify the model so that the loudness function was somewhat flatter at midrange levels. There are empirical data consistent with such a flattening (Florentine *et al.*, 1996; Buus *et al.*, 1997). However, other recent data show midlevel values of ΔLDM close to the values predicted by the model (Cull-

ing and Edmonds, 2007), so it is not clear at present whether a modification of the model to incorporate this is justified.

F. Prediction of absolute thresholds

The model can be used to predict absolute threshold using the assumption that the absolute threshold corresponds to a fixed, low loudness of 0.003 sones or a loudness level of 2 phons; 0.003 sones was found to be the predicted loudness for a diotically presented 1-kHz tone at the absolute threshold. The earlier version of the model (Moore *et al.*, 1997) predicted that the absolute threshold for diotic presentation was about 2 dB lower than the absolute threshold of the same sound for monaural presentation. The revised version of the model described here predicts only a 1-dB difference between diotic and monaural thresholds. Both values are within the range reported in empirical studies (Fletcher and Munson, 1933; Pollack, 1948; Hellman and Zwillocki, 1963), although such data are complicated by the fact that the absolute threshold is rarely exactly the same at the two ears of an individual, whereas the model predictions are based on an ideal case of exactly the same sensitivity at the two ears.

IV. LIMITATIONS OF THE MODEL

The modified loudness model described here, like the earlier version of the model, is based on excitation patterns and specific loudness patterns calculated from the power spectrum of the sound. Hence, the predictions of the model are not affected by changes in the phase spectrum of complex sounds, even though it is known that the phase spectrum can influence loudness (Gockel *et al.*, 2002). Also, the predictions of the model are not affected by changes in the interaural correlation of sounds, even though this does have a small effect on loudness; anticorrelated noises are slightly less loud than correlated noises and the effect is typically equivalent to a change in level of 2 dB or less (Culling and Edmonds, 2007). To make the model sensitive to the phase spectrum and the interaural correlation, it would be necessary to have a very different “front end,” with a realistic time-domain auditory filter bank. However, that is beyond the scope of the present paper.

The model also does not take into account relatively “high-level” processes that might influence loudness perception. For example, judgments of the loudness of sound sources can be influenced by the perceived distance of the sources (Zahorik and Wightman, 2001), even when the cues for distance are visual (Mershon *et al.*, 1981). Also, judgments of the loudness of sounds such as speech might be influenced by the perceived vocal effort, as conveyed by the spectral shape and other acoustical characteristics of speech (ANSI, 1997; Brungart and Scott, 2001). Again, such effects are beyond the scope of the present paper.

V. SUMMARY

We have presented a revised version of a loudness model, intended to give more accurate predictions of the loudness of binaural (including dichotic) stimuli. The revised model is based on the idea of inhibitory interactions between

the outputs of the two ears. The inhibition is assumed to be broadly tuned. The model parameters are chosen so that a sound presented diotically is 1.5 times as loud as the same sound presented monaurally. The model gives reasonably accurate predictions of a large variety of data on binaural loudness perception.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Michael Stone for mathematical tips and Ville Sivonen and Wolfgang Ellermeier for providing their data in numerical form. We also thank Andrew Oxenham, Mary Florentine, and Ville Sivonen for helpful comments on an earlier version of this paper.

- Algom, D., Ben-Aharon, B., and Cohen-Raz, L. (1989a). “Dichotic, diotic, and monaural summation of loudness: a comprehensive analysis of composition and psychophysical functions,” *Percept. Psychophys.* **46**, 567–578.
- Algom, D., Rubin, A., and Cohen-Raz, L. (1989b). “Binaural and temporal integration of the loudness of tones and noises,” *Percept. Psychophys.* **46**, 155–166.
- ANSI (1997). *ANSI S3.5-1997, Methods for the calculation of the speech intelligibility index* (American National Standards Institute, New York).
- ANSI (2005). *ANSI S3.4-2005. Procedure for the Computation of Loudness of Steady Sounds* (American National Standards Institute, New York).
- Breebaart, J., van de Par, S., and Kohlrausch, A. (2001). “Binaural processing model based on contralateral inhibition. I. Model structure,” *J. Acoust. Soc. Am.* **110**, 1074–1088.
- Brungart, D. S., and Scott, K. R. (2001). “The effects of production and presentation level on the auditory distance perception of speech,” *J. Acoust. Soc. Am.* **110**, 425–440.
- Buus, S., Florentine, M., and Poulsen, T. (1997). “Temporal integration of loudness, loudness discrimination, and the form of the loudness function,” *J. Acoust. Soc. Am.* **101**, 669–680.
- Culling, J. F., and Edmonds, B. A. (2007). “Interaural correlation and loudness,” in *Hearing—From Basic Research to Applications*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, New York, in press).
- Fletcher, H., and Munson, W. A. (1933). “Loudness, its definition, measurement and calculation,” *J. Acoust. Soc. Am.* **5**, 82–108.
- Fletcher, H., and Munson, W. A. (1937). “Relation between loudness and masking,” *J. Acoust. Soc. Am.* **9**, 1–10.
- Florentine, M., Buus, S., and Poulsen, T. (1996). “Temporal integration of loudness as a function of level,” *J. Acoust. Soc. Am.* **99**, 1633–1644.
- Gigerenzer, G., and Strube, G. (1983). “Are there limits to binaural additivity of loudness?,” *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 126–136.
- Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data,” *Hear. Res.* **47**, 103–138.
- Glasberg, B. R., and Moore, B. C. J. (2006). “Prediction of absolute thresholds and equal-loudness contours using a modified loudness model,” *J. Acoust. Soc. Am.* **120**, 585–588.
- Gockel, H., Moore, B. C. J., and Patterson, R. D. (2002). “Influence of component phase on the loudness of complex tones,” *Acust. Acta Acust.* **88**, 369–377.
- Gockel, H., Moore, B. C. J., Patterson, R. D., and Meddis, R. (2003). “Louder sounds can produce less forward masking: Effects of component phase in complex tones,” *J. Acoust. Soc. Am.* **114**, 978–990.
- Hellman, R. P. (1976). “Growth of loudness at 1000 and 3000 Hz,” *J. Acoust. Soc. Am.* **60**, 672–679.
- Hellman, R. P., and Zwillocki, J. J. (1963). “Monaural loudness summation at 1000 cps and interaural summation,” *J. Acoust. Soc. Am.* **35**, 856–865.
- Irwin, R. J. (1965). “Binaural summation of thermal noises of equal and unequal power in each ear,” *Am. J. Psychol.* **78**, 57–65.
- Keen, K. (1972). “Preservation of constant loudness with interaural amplitude asymmetry,” *J. Acoust. Soc. Am.* **52**, 1193–1196.
- Levelt, W. J., Riemersma, J. B., and Bunt, A. A. (1972). “Binaural additivity of loudness,” *Br. J. Math. Stat. Psychol.* **25**, 51–68.
- Marks, L. E. (1978). “Binaural summation of the loudness of pure tones,” *J. Acoust. Soc. Am.* **64**, 107–113.

- Mershon, D. H., Desaulniers, D. H., Kiefer, S. A., Amerson, T. L., and Mills, J. T. (1981). "Perceived loudness and visually-determined auditory distance," *Perception* **10**, 531–543.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, San Diego).
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.* **45**, 224–240.
- Pollack, I. (1948). "Monaural and binaural threshold sensitivity for tones and for white noise," *J. Acoust. Soc. Am.* **20**, 52–57.
- Reynolds, G. S., and Stevens, S. S. (1960). "Binaural summation of loudness," *J. Acoust. Soc. Am.* **32**, 1337–1344.
- Scharf, B. (1969). "Dichotic summation of loudness," *J. Acoust. Soc. Am.* **45**, 1193–1205.
- Scharf, B., and Fishken, D. (1970). "Binaural summation of loudness reconsidered," *J. Exp. Psychol.* **86**, 374–379.
- Sivonen, V. P., and Ellermeier, W. (2006). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation," *J. Acoust. Soc. Am.* **119**, 2965–2980.
- Stevens, S. S. (1957). "On the psychophysical law," *Psychol. Rev.* **64**, 153–181.
- Whilby, S., Florentine, M., Wagner, E., and Marozeau, J. (2006). "Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing," *J. Acoust. Soc. Am.* **119**, 3931–3939.
- Zahorik, P., and Wightman, F. L. (2001). "Loudness constancy with varying sound source distance," *Nat. Neurosci.* **4**, 78–83.
- Zwicker, E. (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," *J. Acoust. Soc. Am.* **33**, 248.
- Zwicker, E., and Fastl, H. (1999). *Psychoacoustics—Facts and Models*, 2nd ed. (Springer-Verlag, Berlin).
- Zwicker, E., and Scharf, B. (1965). "A model of loudness summation," *Psychol. Rev.* **72**, 3–26.
- Zwicker, E., and Zwicker, U. T. (1991). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," *J. Acoust. Soc. Am.* **89**, 756–764.

Hearing loss from interrupted, intermittent, and time varying Gaussian noise exposures: The applicability of the equal energy hypothesis

Wei Qiu,^{a)} Bob Davis, and Roger P. Hamernik

Auditory Research Laboratory, State University of New York, 107 Beaumont Hall, Plattsburgh, New York 12901

(Received 11 August 2006; revised 19 December 2006; accepted 20 December 2006)

Eight groups of chinchillas ($N=74$) were exposed to various equivalent energy [100 or 106 dB(A) sound pressure level (SPL)] noise exposure paradigms. Six groups received an interrupted, intermittent, time varying (IITV) Gaussian noise exposure that lasted 8 h/d, 5 d/week for 3 weeks. The exposures modeled an idealized workweek. At each level, three different temporal patterns of Gaussian IITV noise were used. The 100 dB(A) IITV exposure had a dB range of 90–108 dB SPL while the range of the 106 dB(A) IITV exposure was 80–115 dB SPL. Two reference groups were exposed to a uniform 100 or 106 dB(A) SPL noise, 24 h/d for 5 days. Each reference group and the three corresponding IITV groups comprised a set of equivalent energy exposures. Evoked potentials were used to estimate hearing thresholds and surface preparation histology quantified sensory cell populations. All six groups exposed to the IITV noise showed threshold toughening effects of up to 40 dB. All IITV exposures produced hearing and sensory cell loss that was similar to their respective equivalent energy reference group. These results indicate that for Gaussian noise the equal energy hypothesis for noise-induced hearing loss is an acceptable unifying principle.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2434692]

PACS number(s): 43.66.Ed, 43.64.Wn, 43.50.Qp [BLM]

Pages: 1613–1620

I. INTRODUCTION

Understanding the relation between noise-induced hearing loss and the noise exposure variables for the purpose of hearing conservation practice becomes more complex when the exposure is interrupted, intermittent and time varying (IITV). This complexity could be reduced if the equal energy hypothesis (EEH) could be shown to be valid for IITV exposures. While there are exposure conditions where the EEH has been shown to be valid (Dolan *et al.*, 1976; Eldredge *et al.*, 1959; Atherley and Martin, 1971; Clark, 1991) there are also numerous examples where it does not hold (e.g., Fredelius *et al.*, 1987; Mills *et al.*, 1981; Clark *et al.*, 1987; Ward and Turner, 1982; Campo and Lataye, 1992; Ward, 1991; Lataye and Campo, 1996; Patuzzi, 1998; Pourbakht and Yamasoba, 2003). While most industrial noise exposures are typically of the IITV type there does not appear to be any hearing loss data from experimental exposures designed to replicate some of the essential features of an industrial noise environment. Ward (1991, p. 164) defined an IITV exposure as: “An exposure that involves long periods of effective quiet (several hours) will be defined as being *interrupted*, one with short break (a few seconds to an hour) will be characterized as *intermittent*, and one in which the level varies but remains above effective quiet is simply *time varying*.” Typically exposures that are interrupted and/or intermittent produce lower levels of hearing loss than do continuous exposures (Clark and Bohne, 1992; Patuzzi, 1998; Campo and Lataye,

1992; Fredelius and Wersall, 1992; Pourbakht and Yamasoba, 2003). This seems reasonable since during the quiet periods the ear has some time to recover from any threshold shift. The EEH predicts no reduction in hearing loss for interrupted exposures and thus its application to estimating the hazards of an IITV exposure to hearing would be conservative.

While considerable data exist on interrupted and intermittent exposure paradigms there is virtually no data on long-term, time varying exposures. One of the earliest studies of interrupted/intermittent exposures that produced a permanent threshold shift (PTS) can be traced back to the work of Miller *et al.* (1963) who showed that threshold shifts following daily exposures could decrease despite the recurring exposure. Since that time a number of experiments (Clark *et al.*, 1987; Subramaniam *et al.*, 1991; Boettcher, 1993; Hamernik *et al.*, 1994) have confirmed this phenomenon, which has been referred to as a toughening effect. While an interrupted exposure, that produces toughening, has often been shown to make the cochlea less susceptible to PTS from a subsequent traumatic exposure, other studies (Hamernik and Ahroon, 1998, 1999; Ahroon and Hamernik, 1999) including the results found in Miller *et al.* (1963), have shown that the cochlea, although toughened by an interrupted noise exposure, is not protected from the noise that produced the toughening when compared to an equivalent energy uninterrupted control group.

A phenomenon similar to the toughening effect was found by Canlon *et al.* (1988). They showed that exposure of subjects to low level noise prior to a more traumatic exposure could have the effect of reducing the PTS from the

^{a)}Author to whom correspondence should be addressed. Electronic mail: wei.qiu@plattsburgh.edu

higher level exposure. The low level exposure was referred to as a conditioning or priming exposure and the reduction in PTS as the conditioning effect. Evidence is accumulating that both the conditioning and toughening effects are the result of the upregulation of enzymes involved in the control of reactive oxygen and nitrogen species known to increase following noise exposure (Bobbin *et al.*, 1995; Jacono *et al.*, 1998; Ohlemiller *et al.*, 1999) and to be toxic to sensory and supporting cells (Yamane *et al.*, 1995; Ohinataab *et al.*, 2000) in the cochlea.

For an individual, the acoustic environment both on and off the job is IITV in character and the biochemical processes underlying both the toughening and conditioning effects are likely to be continuously active. Thus, the IITV exposures of the industrial environment should elicit toughening effects while the lower level sounds of, for example, transportation, entertainment, etc., may elicit conditioning effects. Despite these inherent protective reactions to noise, hearing loss in the industrial population continues to increase.

The objective of the experiments reported here was to compare the effects on hearing of IITV exposures, modeled after an idealized three-week work schedule, with those of a uniform equivalent energy continuous exposure in order to establish the extent to which the EEH applies to Gaussian IITV exposures. A by-product of the interrupted and intermittent features of the experimental design was the ability to acquire data on the toughening phenomenon and the extent of its effect on noise-induced trauma.

II. METHODS

A. Auditory evoked potential

Seventy-four chinchillas were used as subjects. Each animal was anesthetized [IM injection of Ketamine (35 mg/kg) and Xylazine (1 mg/kg)] and made monaural by the surgical destruction of the left cochlea. During this procedure a bipolar electrode was implanted, under stereotaxic control, into the left inferior colliculus and the electrode plug cemented to the skull for the recording of auditory evoked potentials (Henderson *et al.*, 1973; Salvi *et al.*, 1982). The auditory evoked potential (AEP) was used to estimate pure tone thresholds. The animals were awake during testing and restrained in a yoke-like apparatus to maintain the animal's head in a constant position within the calibrated sound field. AEPs were collected to 20 ms tone bursts (5 ms rise/fall time) presented at a rate of 10/s. Each sampled wave form was analyzed for large-amplitude artifact, and, if present, the sample was rejected from the average and another sample taken. Averaged AEPs were obtained from 250 presentations of the 20 ms signal. Thresholds were measured using an intensity series with 5 dB steps at octave intervals from 0.5 to 16 kHz. Threshold was defined to be one-half step size (2.5 dB) below the lowest intensity that showed a "response" consistent with the responses seen at higher intensities. Additional details of the experimental methods may be found in Ahroon *et al.* (1993).

B. Experimental protocol

Following a two-week postsurgical recovery, three AEP preexposure audiograms were obtained (on different days) on each animal at octave intervals between 0.5 and 16.0 kHz. If the mean of the three audiograms fell beyond one standard deviation (s.d.) of laboratory norms (Hamernik and Qiu, 2000), in the direction of poorer thresholds at more than one test frequency, the animal was rejected. The animals were randomly assigned to one of eight experimental groups with 8–16 animals/group. Two groups were exposed to a uniform, continuous exposure for five days at 100 or 106 dB(A) sound pressure level (SPL). Each of these two groups constituted an equal energy reference group for the IITV exposures. Six groups received a 19-day IITV exposure as detailed below.

The animals were exposed four (or less) at a time to one of the noise conditions detailed below. During exposure, animals were given free access to food and water and were rotated through a bank of six cages daily. The SPLs, across cages, in the middle of each cage, varied within less than ± 1 dB. For the five-day uniform, continuous reference exposures, animals were removed daily for less than 0.5 h for AEP testing. The mean of the five audiograms thus obtained defined asymptotic threshold (AT). For the six IITV groups the animals were tested at the end of the daily exposure on days 1, 2, and 3 and 17, 18, and 19. The difference between the threshold measured following the first day (T_1) and the mean of the thresholds measured following the last three days (T_{17-19}) of the exposure was accepted as an estimate of threshold recovery or toughening (T_r) [i.e., $T_r = (T_1) - (T_{17-19})$]. Thirty days following the last exposure day for all exposure paradigms, three more audiograms were collected on different days and the mean used to define permanently shifted thresholds or permanent threshold shift (PTS).

C. Histology

Following the last AEP test protocol, each animal was euthanized under anesthesia and the right auditory bulla removed and opened to gain access to the cochlea for perfusion. Fixation solution consisting of 2.5% glutaraldehyde in veronal acetate buffer (final pH=7.3) was perfused through the cochlea. After 12–24 h of fixation the cochlea was post-fixed in 1% OsO₄ in veronal acetate buffer. Surface preparation mounts of the entire organ of Corti were prepared (Engstrom *et al.*, 1966) and inner hair cell (IHC) and outer hair cell (OHC) populations were plotted as a function of frequency and location using the frequency-place map of Eldredge *et al.* (1981). Missing cells were identified by the presence of a characteristic phalangeal scar. For purposes of this presentation, sensory cell population data are presented as group averages (in percent missing) taken over octave band lengths of the cochlea centered on the primary AEP test frequencies and as the group mean total number of IHCs or OHCs missing.

D. Noise measurement and analyses

During the exposures the noise field was monitored with a Larson Davis 814 sound level meter equipped with a

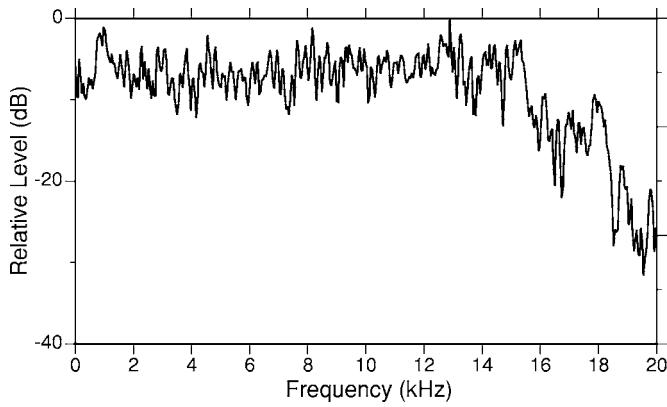


FIG. 1. The relative spectrum level of the unweighted 100 dB(A) SPL Gaussian noise. The 106 dB(A) SPL Gaussian noise had an unweighted spectrum that was qualitatively similar to the 100 dB(A) SPL noise.

1/2 in. microphone. The acoustic signal produced by the Electro-Voice Xi-1152/94 speaker system was transduced by a Brüel and Kjær 1/2 inch microphone (Model 4134), amplified by a Brüel and Kjær (Model 2610) measuring amplifier and fed to a Windows PC-based analysis system. The signal was sampled at 48 kHz with a recording duration of 5.5 min. The design and digital generation of the acoustic signal is detailed in Hsueh and Hamernik (1990, 1991).

E. Noise exposures

Each exposure had in common the same flat spectrum between 0.125 and approximately 15.0 kHz shown in Fig. 1. Experimental group numbers are concisely defined in Table I. The uniform uninterrupted exposures, groups G43 and G94, were presented at 100 and 106 dB(A) SPL, respectively, and served as reference groups for the IITV exposures. All exposures at a given level had approximately the same total energy. The two reference groups [G43:

100 dB(A) and G94: 106 dB(A)] were exposed for 24 h/d for 5 d. These exposures were interrupted once daily for less than a half hour for AEP testing. The six IITV exposures were structured to model an idealized three-week work shift. Each daily exposure consisted of two 4.25 h periods with a 1 h break in between. Each 4.25 h exposure was interrupted for 15 min and each five-day sequence was separated by a two-day break. The time varying SPL for each of the IITV exposures followed one of three different “roughly Gaussian” styled profiles. The SPL for IITV Groups G83, 84 and 85 varied from 90 to 108 dB; for Groups G86, 87 and 88 the SPL varied from 80 to 115 dB. Thus, each of the three IITV exposures at a given energy level differed only in their patterns of the time varying SPL. A schematic of the six IITV exposure paradigms and the group numbers assigned to each is shown in Fig. 2. It should be noted that there was no *a priori* reason for the choice of the Gaussian sound level patterns shown in Fig. 2. An example of the noise level recorded for one of the IITV exposures (G84) over one day is shown in Fig. 3. Animals were returned to the colony following the final hour of each daily exposure.

F. Statistical analysis

The permanent threshold shift (PTS) and sensory cell losses in octave-band lengths of the cochlea were compared among the groups of animals for each noise exposure level [100 and 106 dB(A) SPL] using a two-way, mixed model analysis of variance (ANOVA) with repeated measures on one factor (frequency). Analysis of PTS and sensory cell losses was performed using a completely within-subject analysis. The probability of a type 1 error was set at 0.05. Statistically significant main effects of frequency were expected and found in all of the following analyses because of the frequency-specific nature of the audibility curve of the

TABLE I. Experimental groups defined.

Group No.	Equivalent energy SPL dB(A)	SPL profile	Exposure conditions
G-43	100	—	uniform SPL, 24h/d for 5d
G-85	100		IITV, 90 to 108 dB over 19d
G-83	100		IITV, 90 to 108 dB over 19d
G-84	100		IITV, 90 to 108 dB over 19d
G-94	106	—	uniform SPL, 24h/d for 5d
G-88	106		IITV, 80 to 115 dB over 19d
G-86	106		IITV, 80 to 115 dB over 19d
G-87	106		IITV, 80 to 115 dB over 19d

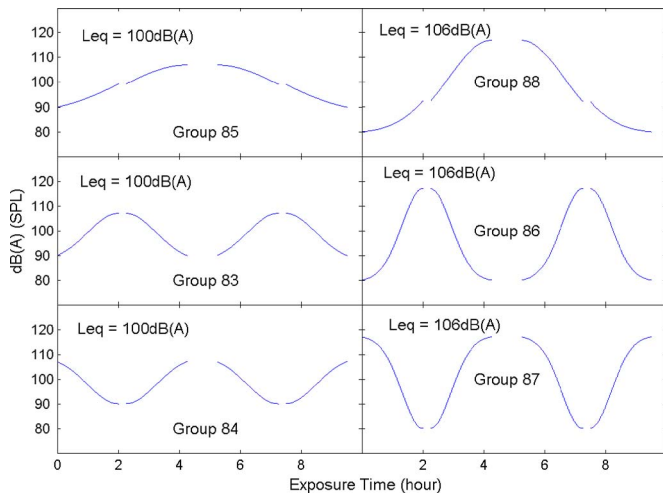


FIG. 2. A sketch of the SPL variations for the intermittent, interrupted and time varying noise exposure paradigm for the 100 and 106 dB(A) SPL groups. Each exposure sequence lasted for 19 days. Each daily exposure consisted of two 4.25 h periods with a 1 h break in between. Each 4.25 h exposure was interrupted for 15 min and each five-day sequence was separated by two-day break.

chinchilla and the noise exposure stimulus. For this reason main effects of frequency are not addressed in the presentation of the results.

III. RESULTS AND DISCUSSION

The group mean preexposure thresholds are shown as symbols in Fig. 4. The shaded area represents ± 1 s.d. of laboratory norms based on 1572 chinchillas. Only group G43 exhibited thresholds that were slightly lower (< 10 dB) than the other seven groups. Figures 5 and 6 show complete group mean data sets for the eight exposure conditions. The permanent threshold shifts (PTS) and the mean threshold recovery (T_r), i.e., toughening, developed during the IITV exposures are shown shaded in the upper panels of these figures. The lower panels show the IHC and OHC loss distribution in percent. Group mean total number of lost OHCs and IHCs along with standard errors of the mean (s.e.) are indicated. The bars on the data points in each figure represent one s.e. If a bar is not present the s.e. was less than the size of the symbol. The results for the two uniform and uninterrupted

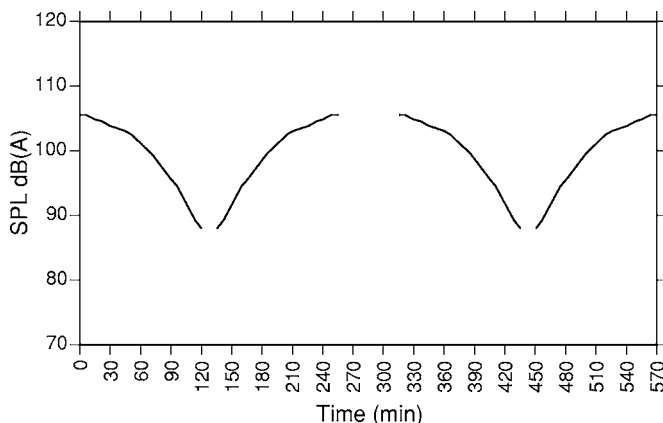


FIG. 3. An example of the actual SPL variations for the intermittent, interrupted and time varying noise exposure paradigm for group G84.

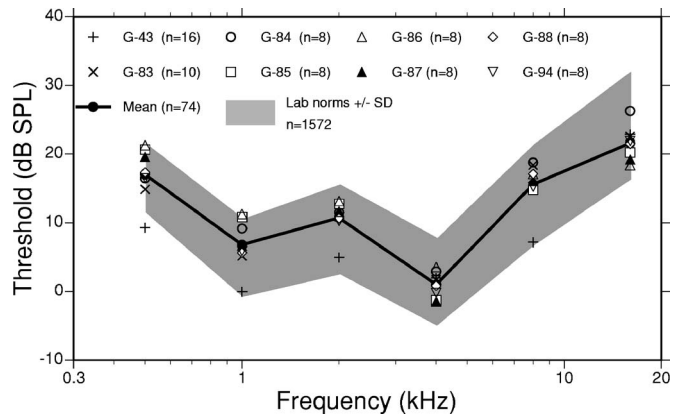


FIG. 4. Mean thresholds for the six experimental interrupted, intermittent and time varying exposure groups (G83, G84, G85, G86, G87, G88) and two uninterrupted reference groups (G43 and G 94) prior to noise exposure. The shaded area represents the mean preexposure thresholds ± 1 s.d. from the laboratory norm based on 1572 chinchillas. The bold line represents the mean preexposure thresholds of the 74 chinchillas included in this study.

reference exposures (G43 and G94) are shown in panel A of each figure. Visual inspection of these figures suggests the following. (1) All the IITV exposures produced a toughening effect that could approach 40 dB at some frequencies (see Fig. 6(G)). There was a noticeable frequency effect on T_r that was inversely related to the pattern of PTS. T_r was large at the lower frequencies where there was less PTS while T_r was small or zero at the highest frequencies where PTS was large. Although the IITV exposures showed a substantial toughening effect, this toughening did not result in a reduced PTS or an appreciable reduction in sensory cell loss when compared to the uninterrupted reference exposures. The issue of whether or not the toughened cochlea is protected from the noise that produced the toughening is not clear. Some exposure paradigms (Clark *et al.*, 1987) have yielded protection while others (Hamernik and Ahroon, 1998) have not. Since experimental paradigms differ considerably, comparisons across studies are difficult. (2) The frequency profile and magnitude of the PTS for the reference and IITV exposures, at both levels, are similar. (3) The distribution of IHC and OHC loss across equivalent energy groups is surprisingly similar.

The group mean PTS and cell loss for the four equivalent energy 100 dB (A) SPL exposures is compared in Fig. 7. The PTS across all four groups varied from about 5 dB at the lower frequencies to about 30 dB at the higher frequencies. The two-way ANOVA indicated that there was a statistically significant main effect of group ($F=6.1, df=3/222$) and no interaction between group and frequency ($F=0.92, df=15/222$). A one-way ANOVA as well as a posthoc analysis (Tukey test) revealed a statistically significant difference between two IITV groups (G83 and G84) and reference group G43 at the lower test frequencies of 0.5 and 1.0 kHz. These differences were less than 10 dB. All three IITV groups also showed significantly more PTS (10–15 dB) at 2.0 kHz compared to the reference group G43. Analysis of IHC loss across the four groups showed no main effect of group ($F=1.51, df=3/224$) and no interaction of group and frequency ($F=1.56, df=21/224$). The IHC loss was relatively small

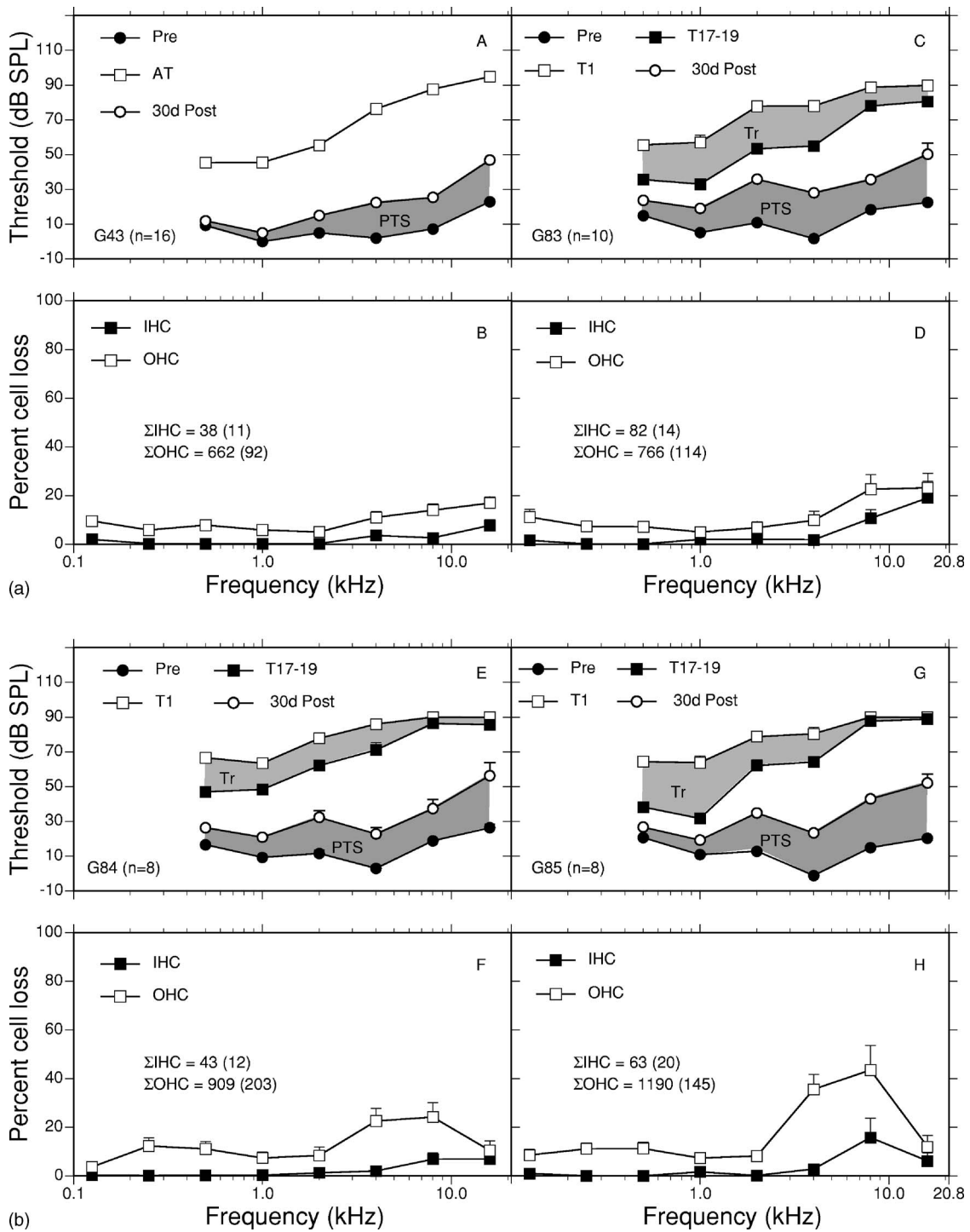


FIG. 5. Summary data from the four 100 dB(A) SPL exposures. Panel (A) shows the group mean asymptotic threshold (AT) and permanent postexposure thresholds. Panel (B) shows the group mean outer and inner sensory cell (IHC and OHC) losses (lower panel) for the uninterrupted reference group G43. Panels (C, D), (E, F), and (G, H) show a similar presentation of data for the three IITV equivalent energy exposures, G83, G84, and G85, respectively. Instead of AT, the group mean thresholds measured immediately following exposure on day 1 (T_1) and the group mean thresholds measured immediately following exposure on days 17, 18 and 19 (T_{17-19}) are shown. Toughening (T_r) and permanent threshold shifts (PTS) are shown shaded. The Σ indicates the group mean total number of OHCs or IHCs missing and () indicates the standard error.

and very near normal for frequencies up to about 4.0 kHz. At the higher frequencies (8 and 16 kHz) there were losses of 5% and 20% respectively. The OHCs showed a broad loss of 20–50% between 2.0 and 16 kHz. There was a significant main effect of group ($F=4.85, df=3/224$) and an interaction of group and frequency ($F=3.16, df=21/224$). At only 4 and 8 kHz was there a statistically significant greater amount of

OHC loss for the IITV group G85 relative to the reference group G43. For the three IITV exposures there was a significant difference only at 4 kHz where group G 85 showed about 22% more OHC loss than did group G83. Thus, among these four groups there were small differences in PTS at the low frequencies where there were no significant differences in OHC and IHC losses and while there were differences in

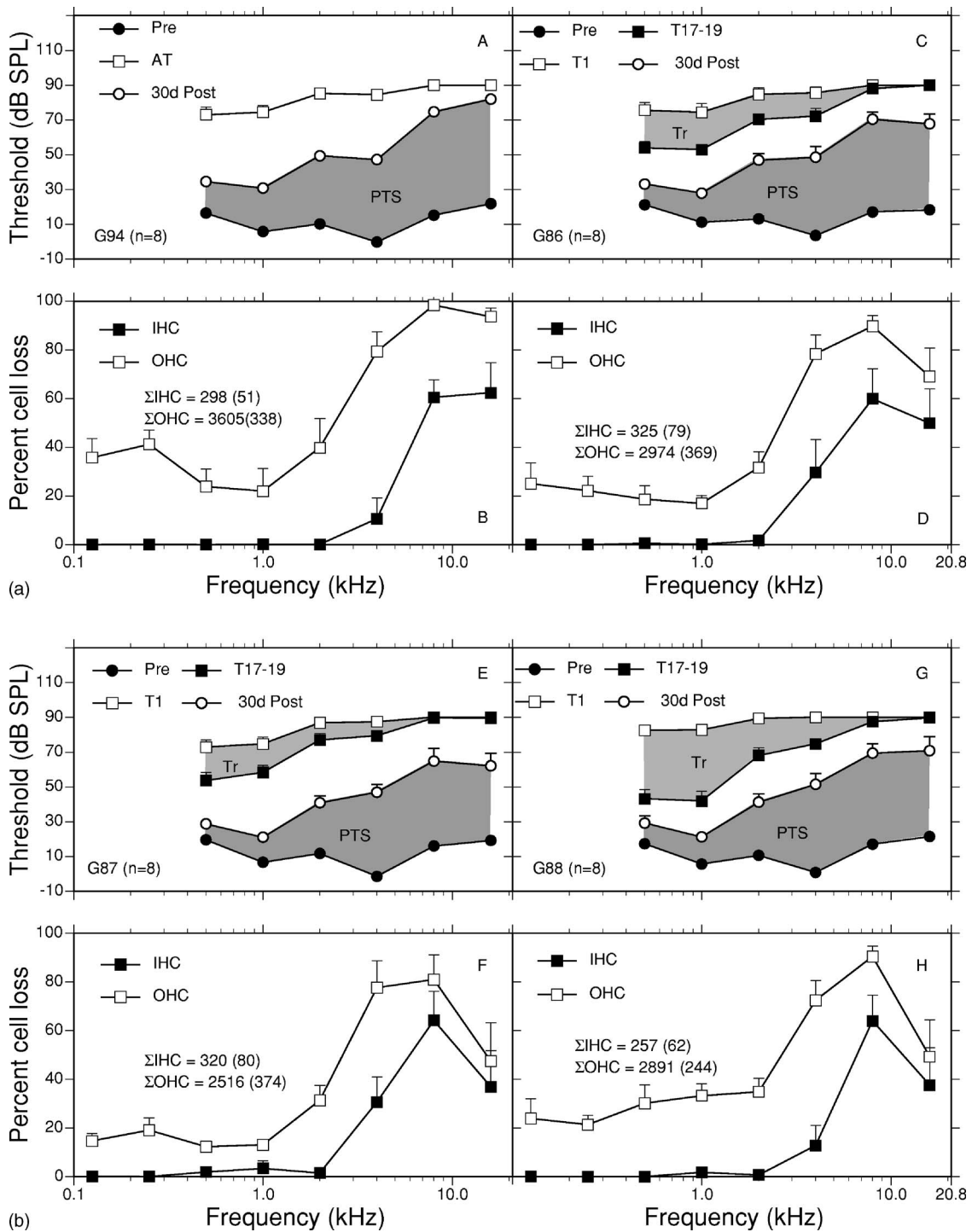


FIG. 6. Summary data from the four 106 dB(A) SPL exposures. Panels (A) shows the group mean asymptotic thresholds (AT) and permanent postexposure thresholds. Panel (B) shows the group mean outer and inner sensory cell (IHC, OHC) losses for the uninterrupted reference group G94. Panels (C, D), (E, F), and (G, H) show a similar presentation of data for the three IITV equivalent energy exposures, G86, G87, and G88, respectively. Instead of AT, the group mean thresholds measured immediately following exposure on day 1 (T_1) and the group mean thresholds measured immediately following exposure on days 17, 18 and 19 (T_{17-19}) are shown. Toughening (T_r) and permanent threshold shifts (PTS) are shown shaded. The Σ indicates the group mean total number of OHCs or IHCs missing and () indicates the standard error.

OHC loss at 4 and 8 kHz there was no significant difference in PTS. Based on the above it is clear that the toughened cochlea showed no protective effects. This result is consistent with our previous report (Hamernik and Ahroon, 1998) that showed no protective effect from a 10 to 30 dB T_r produced by long-term, high-level, narrow band impact noise

exposures in the chinchilla. These data further suggest that the EEH is a reasonable approach to the evaluation of Gaussian IITV long duration noise exposures.

The results of the four equivalent energy 106 dB(A) SPL set of exposures are summarized in Fig. 8. The two-way ANOVA showed that for PTS there was a main effect of

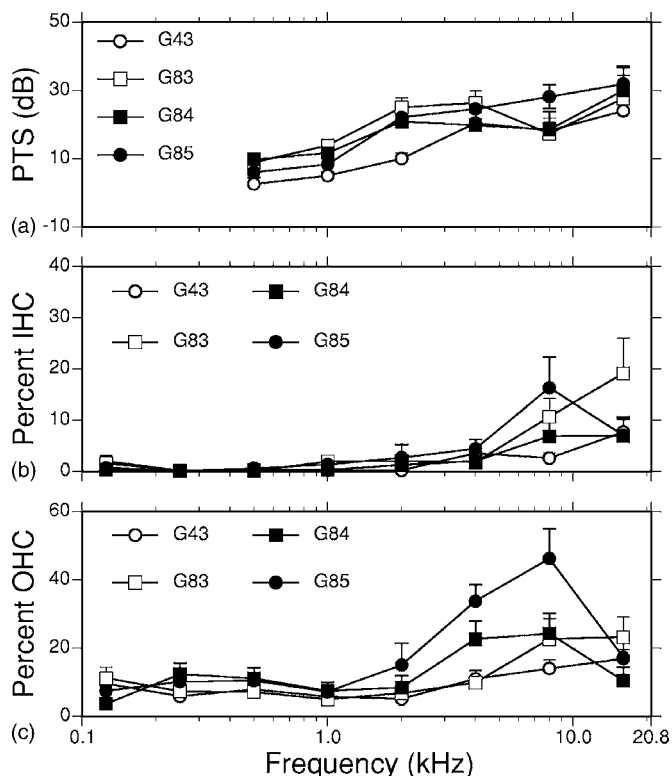


FIG. 7. A comparison of the PTS and cell loss data for the four 100 dB(A) SPL exposures. (a) The permanent threshold shifts (PTS). (b) The percent inner hair cell (IHC) loss. (c) The percent outer hair cell (OHC) loss.

group ($F=4.35, df=3/168$) and no interaction of group and frequency ($F=0.35, df=15/168$). The one-way ANOVA and Tukey tests indicated significantly less PTS (~ 10 dB) at only one frequency (1.0 kHz) for each IITV group (G86, G87 and G88) relative to the uninterrupted reference group (G94). There was no significant main effect of group ($F=0.28, df=3/224$) and no interaction of group and frequency ($F=0.68, df=21/224$) for IHC loss. The IHC population was very near normal between 0.125 and 2.0 kHz but showed losses of 30–60% from 4.0 to 16 kHz. There was a broadly distributed loss of OHCs across the entire cochlea in all four 106 dB(A) SPL groups with severe losses above 2.0 kHz. There was a significant main effect of group ($F=6.58, df=3/224$) and no interaction of group and frequency ($F=1.07, df=21/224$). There were no statistically significant differences in OHC loss among the three IITV groups and between each IITV and the reference group G94 with the exception of the mean percent OHC loss at 0.250 and 16 kHz where the uninterrupted reference group G94 showed more OHC loss compared to the three IITV groups. The similar amount of PTS and sensory cell loss across most of the test frequencies for the IITV and reference groups indicates that the toughening effects did not provide protection from the 106 dB(A) SPL IITV noise exposures. At this higher level, where the hearing and sensory cell loss is substantial, the EEH appears to be an acceptable unifying principle. This is consistent with the results of the 100 dB(A) exposure.

The results for both the 100 and 106 dB(A) SPL IITV exposures contrast with several experiments which have

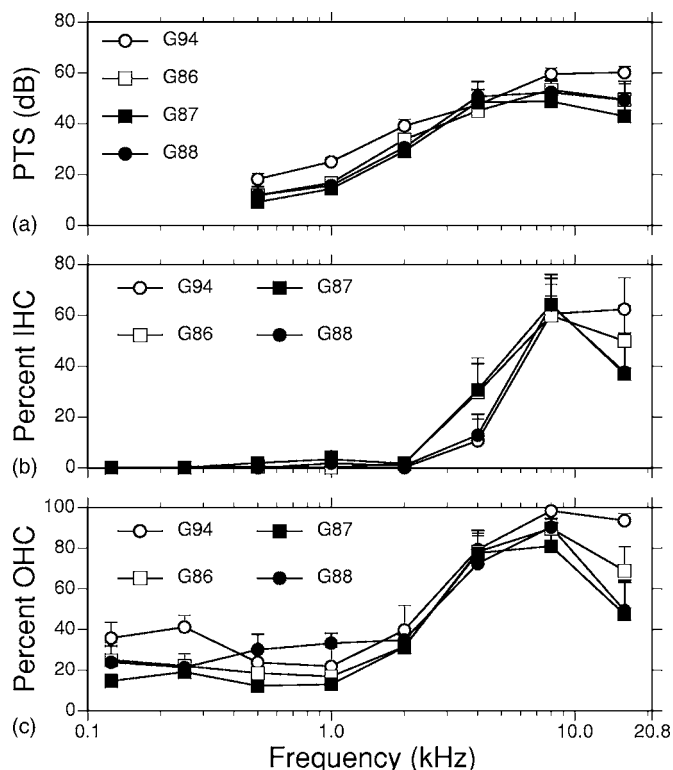


FIG. 8. A comparison of the PTS and cell loss data for the four 106 dB(A) SPL exposures. (a) The permanent threshold shifts (PTS). (b) The percent inner hair cell (IHC) loss. (c) The percent outer hair cell (OHC) loss.

demonstrated that intermittent noise exposures produce significantly smaller PTS and sensory cell loss than energy equivalent continuous noise (Patuzzi, 1998; Campo and Lataye, 1992; Fredelius and Wersall, 1992; Pourbakht and Yamasoba, 2003). These referenced studies were, however, of much shorter duration than the three-week IITV exposure in this study and used very different noise (pure tones, band limited noise) that was presented at a constant SPL. In a study involving 135 chinchillas exposed to high level impact noise on an interrupted and uninterrupted schedule Hamernik and Ahroon (1998) showed that there was no correlation between the amount of toughening and the extent of PTS produced by the toughening noise. They concluded that a toughened cochlea is not protected from the noise that produced the toughening. In the present paper, while a few statistically significant effects for both PTS and OHC loss were observed between the IITV and the uninterrupted Gaussian exposure groups, the relatively small magnitude and inconsistent appearance of these differences support the application of the EEH to Gaussian IITV long duration exposures.

IV. CONCLUSION

Despite a toughening effect found for the IITV exposures, the findings from the present study indicate that steady, uninterrupted Gaussian exposures at 100 or 106 dB(A) SPL produce similar effects on hearing as Gaussian IITV exposures of equivalent energy. Under the conditions of exposure reported here, the EEH, that forms the basis of current damage-risk criteria for noise exposure, appears to be insensitive to the IITV features of a Gaussian

noise exposure that follows an “on time” schedule of an ideal workweek. Thus this study, comparing the effects of continuous uniform exposures with IITV exposures of equivalent energy, suggests that the EEH can be used as a principal for estimating hearing loss from Gaussian IITV exposures. However, industrial noise environments are often non-Gaussian in character and there is evidence (Hamernik *et al.*, 2003) that variables other than energy affect the development of hearing loss from non-Gaussian exposures.

ACKNOWLEDGMENTS

This work was supported by Grant No. 1-R01-OH02317 from the National Institute for Occupational Safety and Health. The technical assistance of George A. Turrentine, Ann Johnson and Adam Bouchard is greatly appreciated. In conducting this research the investigators adhered to the “Guide for the Care and Use of Laboratory Animals,” prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council [DHHS Publication No. (NIH) 86-23, revised 1985].

- Ahroon, W. A., Hamernik, R. P., and Davis, R. I. (1993). “Complex noise exposures: An energy analysis,” *J. Acoust. Soc. Am.* **93**, 997–1006.
- Ahroon, W. A., and Hamernik, R. P. (1999). “Noise-induced hearing loss in the noise-toughened auditory system,” *Hear. Res.* **129**, 101–110.
- Ahroon, W. A., and Hamernik, R. P. (2000). “The effects of interrupted noise exposures on the noise-damaged cochlea,” *Hear. Res.* **143**, 103–109.
- Atherley, G.R.C., and Martin, A. M. (1971). “Equivalent-continuous noise level as a measure of injury from impact and impulse noise,” *Ann. Occup. Hyg.* **14**, 11–28.
- Bobbin, R. P., Fallon, M., LeBlanc, C., and Baber, A. (1995). “Evidence that glutathione is the unidentified amine (Unk 2.5) released by high potassium into cochlear fluids,” *Hear. Res.* **87**, 49–54.
- Boettcher, F. A. (1993). “Auditory brain-stem response correlates of resistance to noise-induced hearing loss in the Mongolian gerbil,” *J. Acoust. Soc. Am.* **94**, 3207–3214.
- Campo, P., and Lataye, R. R. (1992). *Noise Induced Hearing Loss*, edited by A. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik, (Mosby Year Book, St. Louis), pp. 456–466.
- Canlon, B., Borg, E., and Flock, A. (1988). “Protection against noise trauma by pre-exposure to a low-level acoustic stimulus,” *Hear. Res.* **34**, 197–200.
- Clark, W. W., Bohne, B. A., and Boettcher, F. A. (1987). “Effects of periodic rest on hearing loss and cochlear damage following exposure to noise,” *J. Acoust. Soc. Am.* **82**, 1253–1264.
- Clark, W. W. (1991). “Recent studies of temporary threshold shift (TTS) and permanent threshold shift (PTS) in animals,” *J. Acoust. Soc. Am.* **90**, 155–163.
- Clark, W. W., and Bohne, B. A. (1992). *Noise-Induced Hearing Loss*, edited by A. Dancer, D. Henderson, R. J. Salvi and R. P. Hamernik, (Mosby Year Book, St. Louis), pp. 445–455.
- Dolan, T. R., Murphy, R. J., and Ades, H. W. (1976). “A comparison of the permanent deleterious effects of intense noise on the chinchilla resulting from either continuous or intermittent exposure,” in *Effects of Noise on Hearing*, edited by D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. H. Mills (Raven, New York), pp. 327–340.
- Eldredge, D. H., Covell, W. P., and Gannon, R. P. (1959). “Acoustic trauma following intermittent exposure to tones,” *Ann. Otol. Rhinol. Laryngol.* **68**, 723–733.
- Eldredge, D. H., Miller, J. D., and Bohne, B. A. (1981). “A frequency-position map for the chinchilla cochlea,” *J. Acoust. Soc. Am.* **69**, 1091–1095.
- Engstrom, H., Ades, H. W., and Andersson, A. (1966). *Structural Pattern of the Organ of Corti* (Almqvist and Wiksell, Stockholm, Sweden).
- Fredelius, L., Johansson, B., Bagger-Sjoberg, D., and Wersall, J. (1987). “Qualitative and quantitative changes in the guinea pig organ of Corti after pure tone acoustic overstimulation,” *Hear. Res.* **30**, 157–167.
- Fredelius, L., and Wersall, J. (1992). “Haircell damage after continuous and interrupted pure tone overstimulation: A scanning electron microscopic study in the guinea pig,” *Hear. Res.* **62**, 194–198.
- Hamernik, R. P., Ahroon, W. A., Davis, R., and Lei, S. F. (1994). “Hearing threshold shift from repeated 6-h daily exposure to impact noise,” *J. Acoust. Soc. Am.* **95**, 444–453.
- Hamernik, R. P., and Ahroon, W. A. (1998). “Interrupted noise exposures: Threshold shift dynamics and permanent effects,” *J. Acoust. Soc. Am.* **103**, 3478–3488.
- Hamernik, R. P., and Ahroon, W. A. (1999). “Sound-induced priming of the chinchilla auditory system,” *Hear. Res.* **137**, 127–136.
- Hamernik, R. P., and Qiu, W. (2000). “Correlation among evoked potential thresholds, otoacoustic emissions and cochlear sensory cell loss following various noise exposures in the chinchilla,” *Hear. Res.* **150**, 245–257.
- Hamernik, R. P., Qiu, W., and Davis, B. (2003). “The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric,” *J. Acoust. Soc. Am.* **114**, 386–395.
- Henderson, D., Hamernik, R. P., Woodford, C., Sitler, R. W., and Salvi, R. J. (1973). “Evoked response audibility curve of the chinchilla,” *J. Acoust. Soc. Am.* **54**, 1099–1101.
- Hsueh, K. D., and Hamernik, R. P. (1991). “Performance characteristics of a phase domain approach to random noise synthesis,” *Noise Control Eng. J.* **18**–32.
- Hsueh, K. D., and Hamernik, R. P. (1990). “A generalized approach to random noise synthesis: Theory and computer simulation,” *J. Acoust. Soc. Am.* **87**, 1207–1217.
- Jacono, A. A., Hu, B., Kopke, R. D., Henderson, D., Van DeWater, T. R., and Steinman, H. M. (1998). “Changes in cochlear antioxidant enzyme activity after sound conditioning and noise exposure in the chinchilla,” *Hear. Res.* **117**, 31–38.
- Lataye, R., and Campo, P. (1996). “Applicability of the Leq as a damage-risk criterion: An animal experiment,” *J. Acoust. Soc. Am.* **99**, 1621–1632.
- Miller, J. D., Watson, C. S., and Covell, W. P. (1963). “Deafening effects of noise on the cat,” *Acta Suppl.* **176**, 1–91.
- Mills, J. H., Gilbert, R. M., and Adkins, W. Y. (1981). “Temporary threshold shift produced by wideband noise,” *J. Acoust. Soc. Am.* **70**, 390–396.
- Ohinataab, Y., Yamasobac, T., Schachta, J., and Millera, J. M. (2000). “Glutathione limits noise-induced hearing loss,” *Hear. Res.* **146**, 28–34.
- Ohlemiller, K. K., Wright, J. S., and Dugan, L. L. (1999). “Early elevation of cochlear reactive oxygen species following noise exposure,” *Audiol. Neuro-Otol.* **4**, 229–236.
- Patuzzi, R. (1998). “Exponential onset and recovery of temporary threshold shift after loud sound: Evidence for long term inactivation of mechano-electrical transduction channels,” *Hear. Res.* **125**, 17–38.
- Pourbakht, A., and Yamasoba, T. (2003). “Cochlear damage caused by continuous and intermittent noise exposure,” *Hear. Res.* **178**, 70–78.
- Salvi, R. J., Ahroon, W. A., Perry, J., Gunnarson, A., and Henderson, D. (1982). “Psychophysical and evoked-response tuning curves in the chinchilla,” *Am. J. Otolaryngol.* **3**, 408–416.
- Subramaniam, M., Campo, P., and Henderson, D. (1991). “The effect of exposure level on the development of progressive resistance to noise,” *Hear. Res.* **52**, 181–188.
- Ward, W. D., and Turner, C. W. (1982). *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. J. Salvi (Raven, New York), pp. 423–435.
- Ward, W. D. (1991). “The role of intermittence in PTS,” *J. Acoust. Soc. Am.* **90**, 164–169.
- Yamane, H., Nakai, Y., Takayama, M., Konishi, K., Iguchi, H., and Nakagawa, T. (1995). “The emergence of free radicals after acoustic trauma and strial blood flow,” *Acta Oto-Laryngol., Suppl.* **519**, 87–92.

Across-frequency pitch discrimination interference between complex tones containing resolved harmonics

Christophe Micheyl^{a)} and Andrew J. Oxenham^{b)}

Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 7 September 2005; revised 29 November 2006; accepted 7 December 2006)

Pitch discrimination interference (PDI) refers to an impairment in the ability to discriminate changes in the fundamental frequency (F_0) of a target harmonic complex, caused by another harmonic complex (the interferer) presented simultaneously in a remote spectral region. So far, PDI has been demonstrated for target complexes filtered into a higher spectral region than the interferer and containing no peripherally resolved harmonics in their passband. Here, it is shown that PDI also occurs when the target harmonic complex contains resolved harmonics in its passband (experiment 1). PDI was also observed when the target was filtered into a lower spectral region than that of the interferer (experiment 2), revealing that differences in relative harmonic dominance and pitch salience between the simultaneous target and the interferer, as confirmed using pitch matches (experiment 3), do not entirely explain PDI. When the target was in the higher spectral region, and the F_0 separation between the target and the interferer was around 7% or 10%, dramatic PDI effects were observed despite the relatively large F_0 separation between the two sequential targets (14%–20%). Overall, the results suggest that PDI is more general than previously thought, and is not limited to targets consisting only of unresolved harmonics. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431334]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Dc, 43.66.Ba [JHG]

Pages: 1621–1631

I. INTRODUCTION

Harmonic complex tones usually evoke a pitch that corresponds to their fundamental frequency (F_0). Most current theories and models of pitch perception assume that the auditory system determines the pitch of a complex tone by integrating information across a relatively wide frequency range. In some cases the integration is based on estimates of the frequencies of individual resolved harmonics (e.g., Goldstein, 1973); in other cases it is based on a “summary autocorrelation function” that sums the periodicities present across all frequency channels (e.g., Licklider, 1951; Meddis and O’Mard, 1997).

In cases where multiple harmonic sound sources are simultaneously present, the auditory system seems able to selectively integrate information from components that fit the harmonic series of the target source for the purpose of estimating its pitch, at least when peripheral interactions do not lead to masking of the target components (e.g., Beerends and Houtsma, 1989). Various models have been designed to account for this ability to extract concurrent pitches when the complexes occupy overlapping spectral regions (e.g., Beerends and Houtsma, 1989; Meddis and Hewitt, 1992; de Cheveigné and Kawahara, 1999). If the sources occupy distinct spectral regions, so that peripheral interactions between the two groups of harmonics are insignificant, it has generally been assumed that the pitch of one source should not interfere with the pitch of the other.

Contrary to this prevailing view, Gockel *et al.* (2004, 2005a) recently showed that listeners’ ability to discriminate small changes in the F_0 of a “target” group of harmonics filtered into a given spectral region could be severely disrupted by the presence of an “interferer” group of harmonics, which was filtered into another spectral region and had a constant F_0 . Note that this effect, which was termed pitch discrimination interference (PDI), occurred despite the fact that the target and the interferer were filtered into remote, nonoverlapping spectral regions, and listeners were instructed to focus on the target sound. This across-frequency PDI phenomenon is superficially reminiscent of another form of interference in pitch discrimination, where a listener’s ability to discriminate differences in the F_0 ’s of two successive target complexes is impaired by the presentation, before and after each target, of an interferer complex filtered into the same spectral region (Carlyon, 1996; Micheyl and Carlyon, 1998; Gockel *et al.*, 1999). Although in the latter case F_0 discrimination is impaired only when the interferer occupies the same spectral region as the target complex, in both cases the interference occurs between harmonic complexes that do not significantly interact within the auditory periphery: in one case the interference is between two complexes filtered *simultaneously* into *different* spectral regions, whereas in the other, it is between two complexes filtered *sequentially* into the *same* spectral region.

The discovery of PDI effects has potentially important implications for theories and models of pitch perception, some of which are discussed by Gockel *et al.* (2004, 2005a). However, it is not yet clear how general or relevant the effect of PDI is for everyday perception. Gockel *et al.* studied the case where the target contained only harmonics numbered above the 16th, which are believed to be

^{a)} Author to whom correspondence should be addressed. Electronic mail: cmicheyl@umn.edu

^{b)} Electronic mail: oxenham@umn.edu

unresolved in the auditory periphery and produce a relatively weak pitch percept (Plomp, 1964; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). In contrast, most sources in the environment also contain lower-order resolved harmonics, which produce a much more salient pitch than unresolved harmonics, and which appear to be important for the segregation of concurrent harmonic sources differing in F_0 (Carlyon, 1996; Carlyon *et al.*, 2002; Qin and Oxenham, 2005). Thus, an important question is whether PDI also occurs when the target contains peripherally resolved harmonics. If the relative pitch salience of target and interferer governs the amount of PDI observed, as suggested by Gockel *et al.* (2004), one might expect target complexes that contain resolved harmonics in their passband to be far less susceptible to PDI than target complexes that contain only unresolved harmonics. On the other hand, even resolved harmonics affect the overall pitch of a complex to differing degrees (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985; Dai, 2000; Gockel *et al.*, 2005b). If the relative dominance of individual resolved harmonics plays a role in PDI, substantial PDI might be obtained between two groups of resolved harmonics if the harmonics in the interferer group are more dominant in determining pitch than those of the target group. The first aim of the present study was therefore to test whether PDI generalizes to conditions in which the target complex contains resolved harmonics.

II. EXPERIMENT 1: PITCH DISCRIMINATION INTERFERENCE BETWEEN GROUPS OF RESOLVED HARMONICS

A. Methods

1. Stimuli

The basic stimuli in this study were harmonic complex tones with a nominal F_0 of 250 Hz. Each complex was passed through a digital bandpass filter (8th-order Butterworth) with fixed 3-dB cutoff frequencies of 125–625 Hz (lower spectral region) or 1375–1875 Hz (higher spectral region). These stimulus parameters were chosen, following Shackleton and Carlyon (1994) and Gockel *et al.* (2004), such that the complexes always contained peripherally resolved harmonics in their passbands. As a rule of thumb, it is usually considered that harmonics below about the 10th are resolved; for the lowest average target F_0 tested in this experiment (3 semitones below 250 Hz), the higher spectral region contained the 7th, 8th, and 9th harmonics between the 10-dB cutoff points of the filter.

The interferer F_0 , relative to which the two target F_0 's on a trial were set, was drawn randomly from a 6-semitone-wide uniform distribution centered on (i.e., ± 3 semitones around) 250 Hz. The large rove range, which was roughly twice that used by Gockel *et al.* (2004; 2005a), helped to ensure that listeners could not profitably base their judgments on the absolute F_0 of the target in any one interval, and instead had to compare the F_0 's of the targets in the two observation intervals presented within each trial. To test whether the use of a larger roving range had a significant impact on performance, and to facilitate comparisons with

the results of Gockel *et al.* (2004, 2005a), we retested the listeners in one condition of the first experiment using a smaller roving range (3 semitones, which is approximately 19% of the lowest F_0 in the range).

Each harmonic complex was 400 ms in duration, including 20-ms raised-cosine on and off ramps. The harmonics all started in 0-deg (sine) phase, and those not attenuated by the bandpass filter had a level of 45 dB SPL. The harmonics were summed and presented in a background of pink noise with a spectrum level of 12 dB ($re: 20 \mu\text{Pa}$) at 1 kHz [compared to approximately 15 dB SPL in Gockel *et al.*'s (2004) study], and an upper cutoff frequency of 20 kHz. The purpose of this noise was twofold: (1) to prevent listeners from relying on within-channel cues that could have arisen from interactions of components from both spectral regions within auditory filters centered between the two regions; and (2) to mask any possible distortion products. The noise was turned on 300 ms before the onset of the first tone and off 300 ms after the offset of the second tone on each trial, giving it a total duration of 1.9 s. On each trial, a fresh noise sample was obtained by cutting a randomly selected 1.9-s portion of a 10-s burst of pink noise that was generated at the start of each experimental block.

The stimuli were played out with 24-bit resolution at a 50-kHz sampling rate via a LynxStudio LynxOne sound card. After being passed to a headphone buffer (TDT HB6, Tucker Davis Technologies), they were delivered to the listener's left ear through the left earpiece of a Sennheiser HD 580 headset.

2. Procedure

This experiment used a two-interval, two-alternative (2I-2AFC), constant-stimuli procedure. On each trial, two "target" tones were presented successively, separated by a 500-ms interstimulus interval. One of the two complexes had an F_0 of $F_{0\text{REF}} + \Delta F_0/2$, and the other an F_0 of $F_{0\text{REF}} - \Delta F_0/2$, where $F_{0\text{REF}}$ was the average F_0 of the two targets and ΔF_0 was equal to 0.4375%, 0.875%, 1.745%, 3.5%, 7%, 14%, 20%, or 40% of $F_{0\text{REF}}$, depending on the condition being tested. The order of presentation of the lower- and higher- F_0 tones was randomly chosen on each trial with equal probability. The listener's task was to indicate which interval contained the target with the higher F_0 . Visual feedback was provided after each trial.

The target complexes, which listeners had to discriminate, were either presented alone or gated synchronously with another group of harmonics, which constituted the interferer. The F_0 of the interferer was arithmetically centered between the F_0 's of the two targets ($F_{0\text{REF}}$) and was always the same in both intervals of a trial, so that it provided no relevant information for the task at hand. In this first experiment, the target complex was always filtered into the higher spectral region (1375–1875 Hz), and the interferer was always in the lower region (125–625 Hz). The d' values reported in this article are based on 300 trials per listener per condition. The different test conditions resulting from the combination of the ΔF_0 and target absent/present factors were tested in blocks of 50 trials. Within a given block, the stimulus parameters remained constant. The testing order was pseudorandom, with the constraint that groups of blocks

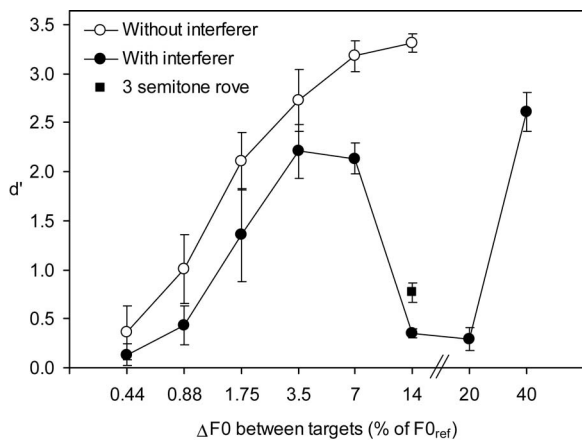


FIG. 1. $F0$ discrimination performance for a target complex filtered between 1375 and 1875 Hz as a function of the $F0$ difference between the two target tones, $\Delta F0$. The latter is expressed in percent of the reference $F0$ around which the two target $F0$'s were symmetrically placed. Empty symbols show performance for the target alone; solid symbols show performance in the presence of a simultaneous interferer complex, which was filtered into a lower spectral region (125–625 Hz). The solid square corresponds to performance measured using a smaller across-trial $F0$ roving range of 3 semitones (approximately 19%), instead of 6 semitones (approximately 41%). The error bars indicate standard errors of mean across listeners.

in which the interferer was absent were run before groups of blocks with the interferer, so that listeners were given the opportunity to listen to the target in isolation before they were tested with the interferer.

3. Listeners

Five listeners took part in this experiment. All had normal hearing, defined by absolute pure-tone thresholds lower than 15 dB HL at octave frequencies between 250 and 8000 Hz, as measured using a Madsen AC30 audiometer and TDH49 earphones. Listeners were given the opportunity to practice the task with and without interferers, until their performance remained stable across blocks of trials. In addition, four of the five listeners already had over 30 h of practice in experiments involving pitch perception, using harmonic complexes similar to those used here.

B. Results

The results of this experiment are shown in Fig. 1. In the absence of the interferer (open symbols) performance (d') increased monotonically with $\Delta F0$, as expected. At the smallest $\Delta F0$ tested (0.44%), the mean d' was not significantly different from zero [one-sample t-test, $t(4)=1.29$, $p=0.265$], indicating chance performance. At the 14% $\Delta F0$, performance was essentially at ceiling.¹ Preliminary results showed that performance remained at ceiling at $\Delta F0$'s larger than 14% in the no-interferer condition. Therefore, performance at the larger $\Delta F0$'s of 20% and 40% was not measured formally.

Performance in the presence of the interferer varied non-monotonically with $\Delta F0$, increasing from not significantly different from chance at the 0.44% [$t(4)=0.136$, $p=0.302$] and 0.875% [$t(4)=0.432$, $p=0.092$] $\Delta F0$'s, up to well above chance at the 3.5% and 7% $\Delta F0$'s, and then showing a marked dip in performance at $\Delta F0$'s of 14% and 20%. In

fact, at a $\Delta F0$ of 20%, performance was not significantly above chance [$t(4)=0.292$, $p=0.062$]. At the largest $\Delta F0$ tested (40%), performance in the presence of the interferer was well above chance, but still below ceiling.

In order to test whether the interferer had a significant effect overall, a two-way repeated-measures analysis of variance (RMANOVA) was performed, with the interferer (present/absent) and $\Delta F0$ as within-subject factors, and the d' values from each subject in each condition as the dependent variable. The results showed significant main effects of interferer [$F(1,4)=321.92$, $p<0.001$] and $\Delta F0$ [$F(4,7)=25.54$, $p=0.001$], and an interaction between these two factors [$F(7,28)=28.91$, $p<0.001$]². Planned comparisons (multiple paired t-tests) further revealed a significant interferer effect ($p<0.05$) at all $\Delta F0$'s except for 0.44% (where performance was at chance with and without an interferer) and, unexpectedly, for 3.5%.

The solid square in Fig. 1 represents the average performance measured in the same listeners using a smaller across-trial $F0$ roving range of 3 semitones (approximately 19%), instead of 6 semitones (approximately 41%), with the interferer present. As can be seen, the results of this control condition confirm the previous observation of poor performance at the 14% $\Delta F0$, indicating that this effect was not simply due to our use of a large $F0$ roving range. As expected, performance with the smaller rove range in the absence of the interferer remained near ceiling and is not shown here.

C. Discussion

The results demonstrate significant interference effects in pitch discrimination between two groups of resolved harmonics filtered into nonoverlapping spectral regions. This finding complements the earlier results of Gockel *et al.* (2004, 2005a), which demonstrated PDI with target complexes containing only unresolved harmonics. These results raise two important questions, one general and one specific: First, why does PDI occur between two groups of resolved harmonics? Second, why did performance in the presence of the interferer vary nonmonotonically with $\Delta F0$, with a marked dip in performance at $\Delta F0$'s of 14%–20%?

Gockel *et al.* (2004) suggested that PDI was caused by the more salient pitch of the interferer's resolved harmonics "swamping" the weaker pitch produced by the unresolved harmonics of the target. This explanation cannot account for the present results because both the target and the interferer complexes consisted primarily of resolved harmonics. However, it is still conceivable that the interferer evoked a more salient or dominant pitch than the target, because it was filtered into a lower spectral region, and thus contained lower-numbered harmonics in its passband. Consistent with this hypothesis, Gockel *et al.* (2005b) measured the dominance of individual harmonics for an $F0$ of 250 Hz (corresponding to that used here), and found that, for 200-ms-long complexes, harmonics 1 to 4 were the most dominant. Dai (2000) also found that for $F0$'s of 200 and 300 Hz (comparable to the 250-Hz $F0$ used here), harmonics 1 through 4 were most dominant, and that the contribution of harmonics above the 6th to the overall pitch was not significant. Here, the inter-

ferer, being filtered between 125 and 625 Hz, typically contained the first three harmonics, while the target, being filtered between 1325 and 1875 Hz, typically contained harmonics between the 5th and the 8th. If the PDI effects observed in this experiment can be accounted for in terms of the relative dominance of harmonics, then exchanging the spectral regions of the target and interferer should lead to a large reduction or elimination of PDI, because then the target would occupy the region with the most dominant harmonics. This prediction was tested in experiment 2.

The answer to the second question, regarding the dip in performance at $\Delta F0$'s of about 14%–20%, may be related to the relationships between individual harmonics in the two targets. Generally, harmonics between 5 and 8 were present in the target passband. The frequency differences between harmonics 5 and 6 (20%), 6 and 7 (17%), and 7 and 8 (14%) correspond to the $F0$ differences at which performance dipped. As the $F0$ ratio between two identically bandpass-filtered harmonic complexes approaches the frequency ratio between consecutive harmonics in the stimulus spectrum, the frequencies of the harmonics of rank n and $n+1$ in the higher- $F0$ complex become close to those of the harmonics of rank $n+1$ and $n+2$ in the lower- $F0$ complex. This has two consequences. The first is that the frequency distance between the harmonics of the two complexes reaches a local minimum. This is illustrated in the top panel of Fig. 2, which shows the mean absolute frequency distance between closest-neighbor components across the two targets, as a function of $\Delta F0$. As can be seen, this distance first increases with $\Delta F0$ below 7%, but then decreases to reach a local minimum at the 14% $\Delta F0$. The depth and width of this “trough” depend to a large extent on which components are included in the calculation, and how these components are weighted in computing the mean frequency shift. The data shown in Fig. 2 were obtained by considering only components that fell within the 10-dB bandwidth of the higher-region filter, and assuming for simplicity that these components received equal weights. With different assumptions, more or less marked decreases in the mean distance can be observed at the 14% and 20% $\Delta F0$'s.

The second consequence is that, as $\Delta F0$ exceeds half the percentage difference between the two harmonics in the target passband, the pairs of harmonics whose frequencies are most similar between the two targets have different ranks. This produces a decoupling between the direction of the shift in $F0$ and the shifts in the frequencies of individual harmonics. For instance, when $F0_{\text{REF}}$ was 250 Hz and $\Delta F0$ was 14%, the higher- $F0$ target contained harmonics with frequencies of 1605 Hz (6th harmonic) and 1873 Hz (7th harmonic), while the lower- $F0$ target contained harmonics with frequencies of 1628 Hz (7th harmonic) and 1860 Hz (8th harmonic). Thus, a listener basing his/her decisions on the direction of the frequency shifts of individual harmonics in the higher-region passband, or on the average frequency of these harmonics (i.e., the spectral centroid of the target), would perform least accurately for $\Delta F0$'s of 14%–20%. This is illustrated in the middle panel of Fig. 2, which shows the proportion of correct responses achieved by an “ideal” (virtual but nonoptimal) observer whose decisions are based on

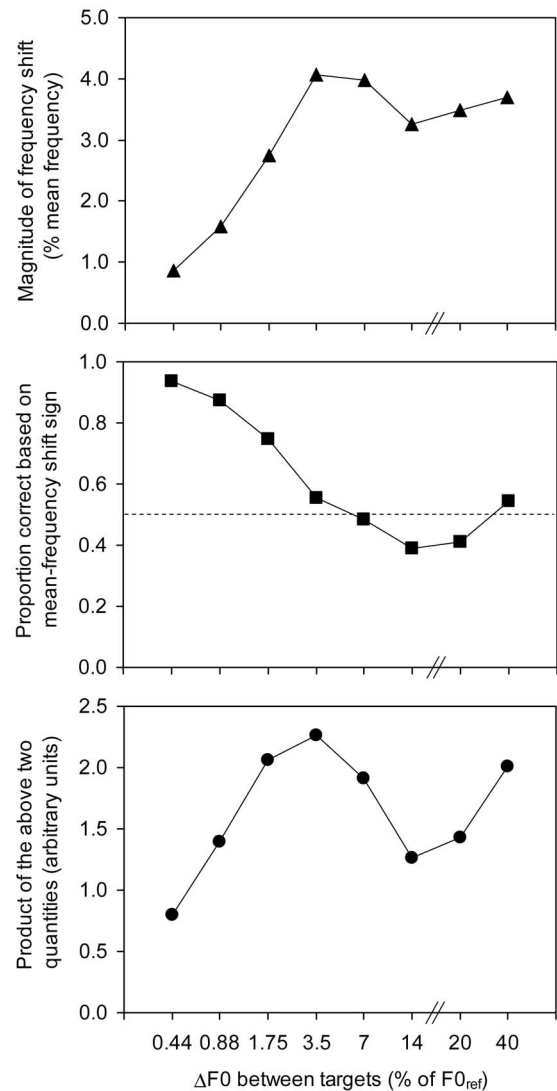


FIG. 2. Variations in the magnitude and sign of the mean frequency distance between closest-neighbor components across the two higher-region targets as a function of the $\Delta F0$ between them. The upper panel shows the absolute value of the difference between the mean frequencies of two sets of harmonics falling inside the 10-dB bandwidth of the higher-region filter, one set having an $F0$ of $F0_{\text{REF}} - \Delta F0/2$, and the other, an $F0$ of $F0_{\text{REF}} + \Delta F0/2$. $F0_{\text{REF}}$ was varied over a 6-semitone range, with a uniform distribution, as in the main conditions of experiment 1. The middle panel shows the proportion of cases in which the frequency difference was the same as that of the $F0$ shift between the first and second targets. This corresponds to the proportion of correct responses for an observer basing his decisions on the sign of the mean frequency difference between the two sets of target components. Chance performance, corresponding to a probability of 0.5, is indicated by the vertical dashed line. The bottom panel simply shows the product of the two above quantities, reflecting a combination of information on both the magnitude and the sign of the shift in the mean frequency of the target components.

the direction of the shift in the mean frequency of the harmonics in the higher region. As can be seen, this proportion reaches a minimum for $\Delta F0$'s of 14%–20%.

The predicted proportions of correct responses shown in the middle panel of Fig. 2 are based on simple calculations that do not take into account a number of factors that could affect the performance of real listeners. Clearly, these predictions are not entirely consistent with the experimental data

shown in Fig. 1. For instance, at $\Delta F0$'s of 14%–20%, the predicted proportion of correct responses falls below the 0.5 chance line, suggesting that, at these $\Delta F0$'s, an observer basing decisions solely on the mean spectral positions of the components in the passband would perform *below* chance. The important point here is that the trend for worst performance at $\Delta F0$'s of 14%–20% is qualitatively consistent with the data in Fig. 1. More importantly, the predictions shown in the middle panel of Fig. 2 were derived assuming perfect encoding of the component frequencies. The performance of real listeners is undoubtedly limited by imperfections in the ability to encode stimulus frequencies, and should therefore be expected to increase with the magnitude of the frequency difference between the two observation intervals plotted in the upper panel of Fig. 2. As shown in the bottom panel of Fig. 2, taking both the magnitude and the sign of the frequency shifts into account (which was done here in a crude way, by multiplying the quantity in the top panel by that in the middle panel), one obtains predictions that show the same general qualitative trends as the data in Fig. 1, with an increase in performance with $\Delta F0$ up to 7% and a dip in performance around 14%–20%.

These considerations suggest a first possible explanation for the dip in performance observed at the 14%–20% $\Delta F0$'s in the presence of the interferer. According to this explanation, the interferer may have forced (or at least encouraged) listeners to rely on the spectral positions of the individual harmonics or on the overall spectral location, rather than on the $F0$ of the target. Earlier studies have shown that in pitch discrimination tasks with small numbers of harmonics, many listeners rely on changes in the spectral position of the harmonics, rather than on changes in the $F0$ *per se* when no feedback is provided (Smootenburg, 1970; Laguitton *et al.*, 1998; Schneider *et al.*, 2005). Here, this listening strategy may have been promoted by the interferer occupying a lower region, and evoking a more salient pitch, than the target. From this point of view, it appears that, at least in the case of small $\Delta F0$'s (below about 3.5%), the measured amount of PDI may represent a lower bound on the effect, because performance could have been even poorer had subjects not been able to use spectral shifts as an alternative cue to the $\Delta F0$, in the presence of the interferer.

Another possible explanation for the large PDI effects observed at the 14%–20% $\Delta F0$'s, and the fact that at these $F0$'s performance in the presence of the interferer dropped to (or close to) chance, is that the internal representation of the target $F0$ in the central auditory system was almost completely “suppressed” by the more dominant interferer. Specifically, it is conceivable that in a situation where two groups of harmonics compete in determining the perceived pitch, a “harmonic sieve” tuned to the $F0$ of the dominant group is applied within the spectral region of the other group, and that harmonics that fall in between the teeth of the harmonic comb in that region are suppressed in order to avoid perceptual ambiguity. Under this “harmonic-cancellation” hypothesis, the suppression should be most effective in the situation where the target harmonics in the higher region fall almost exactly in between the comb teeth, as was the case when $\Delta F0$ was around 14%–20%. Whether the poor perfor-

mance in the presence of the interferer at $\Delta F0$'s of 14%–20% is due to listeners relying on misleading spectral cues or to the $F0$ of the dominant low-region harmonics suppressing the pitch of the harmonics in the higher region, the corresponding PDI effects should be eliminated when the spectral positions of the target and interferer are swapped, so that the target is in the lower spectral region because (a) the harmonics in that region are dominant, and (b) the spectral shifts in that region are coherent with the $F0$ shifts.

III. EXPERIMENT 2: PDI WITH THE TARGET IN A LOWER SPECTRAL REGION THAN THE INTERFERER

A. Methods

The procedure and listeners were the same as in the previous experiment. The only difference was in the stimuli: the spectral regions of the target and interferer were swapped, so that the target was now filtered into the lower spectral region (125–625 Hz) while the interferer was filtered into the higher region (1375–1875 Hz). Four listeners took part in this experiment, and only the 6-semitone rove range was tested. Because data from these four listeners were also collected under the conditions described in experiment 1, the data from both experiments could be compared using a within-listeners design. Importantly, the order of data collection was counterbalanced, such that two of the listeners completed experiment 2 before embarking on experiment 1, and the other two completed the two experiments in the opposite order. This, combined with the fact that in both experiments listeners' performance was stable across blocks of trials, makes it very unlikely that the differences in results between the two experiments were due to practice effects.

B. Results

The results of this experiment are shown in Fig. 3, as squares connected by solid lines. To facilitate a comparison with experiment 1, the data from Fig. 1 are replotted as dotted and dashed lines. Consider first the results obtained in this experiment without the interferer present (solid lines). The results are virtually indistinguishable from those obtained in the comparable conditions of experiment 1 (dashed lines); this was confirmed by the results of a two-way RMANOVA on the d' data, with spectral region (target in high or low region) and $\Delta F0$ as factors, which revealed no main effect of spectral region [$F(1, 4)=0.001$, $p=0.972$] and no interaction [$F(5, 20)=0.451$, $p=0.808$]. This outcome is consistent with earlier results showing similar $F0$ discrimination thresholds for complexes filtered into these spectral regions, as long as they contained resolved harmonics (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). As $F0$ difference limens are often taken as an indirect measure of pitch salience, it suggests that the salience produced by the resolved harmonics in the upper region was the same as that produced by the resolved harmonics in the lower region, at least when presented in isolation.

Consider next the effect of adding the interferer (compare empty and filled symbols connected by solid lines). Unlike in the previous experiment, where large PDI effects were

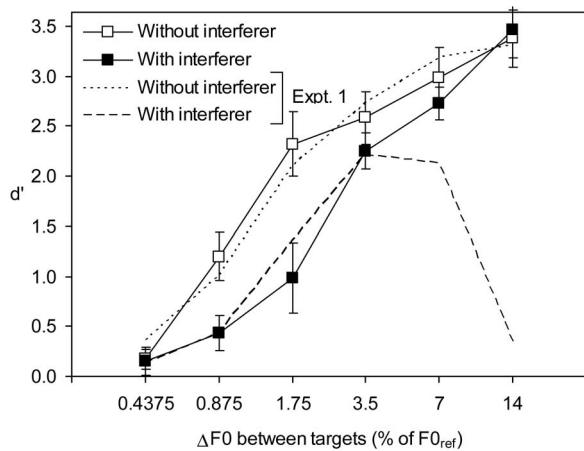


FIG. 3. $F0$ discrimination performance for a target complex filtered between 125 and 625 Hz as a function of the $F0$ difference between the two target tones, $\Delta F0$, in the absence and presence of a higher-region interferer (1375–1875 Hz). $\Delta F0$ is expressed in percent of the reference $F0$ around which the two target $F0$'s were symmetrically placed. The data from this experiment are shown as squares connected by solid lines. As in Fig. 1, empty symbols show performance for the target alone; solid symbols, performance in the presence of a simultaneous interferer complex. The error bars indicate standard errors of mean across listeners for the data of experiment 2. To facilitate comparison with experiment 1, the data from experiment 1 (Fig. 1) have been replotted here as dotted and dashed lines; to avoid clutter, the corresponding symbols and error bars are not shown here.

observed at $\Delta F0$'s larger than 7%, in this experiment, the interferer effect seems negligible at large $\Delta F0$'s. A two-way RMANOVA on these d' data showed no significant main effect of the interferer [$F(1, 4)=3.33$, $p=0.142$], but a significant interaction between interferer and $\Delta F0$ [$F(1, 4)=5.037$, $p=0.004$]. Planned comparisons (paired t-tests) indicated that the only $\Delta F0$'s at which the interferer significantly decreased performance were the 0.875% [$t=4.30$, $p < 0.05$] and 1.75% [$t=2.94$, $p < 0.05$] $\Delta F0$'s. In fact, at these small $\Delta F0$'s, as apparent in Fig. 3, the effect of the interferer was similar to that observed in experiment 1—a three-way RMANOVA on the data collected at these two $\Delta F0$'s in experiments 1 and 2 showed no significant main effect of experiment [$F(1, 4)=0.001$, $p=0.980$] and no significant interaction involving this factor.

C. Discussion

The results of experiments 1 and 2 indicate that for $\Delta F0$'s of 7% or more, the influence of the interferer was heavily dependent on the relative spectral positions of the target and interferer: although the interferer had a large detrimental effect on performance when the target was filtered into the higher spectral region and the interferer into the lower spectral region (experiment 1), it had essentially no effect in the converse situation (experiment 2). This outcome is consistent with the possible explanations described earlier, according to which the PDI effects at $\Delta F0$'s larger than 3.5% were due to the pitch evoked by the more dominant harmonics in the lower-region suppressing (partially or completely) that of the harmonics in the higher region, and/or to listeners basing their judgments on spectral shifts in the higher region.

In contrast, for $\Delta F0$'s less than 3.5%, the interferer was found to have a significant influence regardless of whether it

occupied the lower or the higher spectral region. This outcome indicates that, at these small $\Delta F0$'s, either PDI was mediated by mechanisms that depend little on harmonic dominance, or the harmonics in the higher region were not appreciably less dominant in determining pitch than those in the lower region. In an attempt to distinguish between these two possibilities, the final experiment used a pitch-matching procedure to provide a more direct test of what pitch(es) listeners were perceiving when listening to the combination of target and interferer.

IV. EXPERIMENT 3: PITCH MATCHES

A. Rationale

The various possible interpretations for the results of the previous two experiments provide some basis for predicting the pitch percepts produced by the experimental stimuli. This experiment used a pitch-matching procedure to gather more direct information regarding the pitches evoked by the simultaneous lower- and higher-region complexes as a function of their $F0$ separation. Although Shackleton and Carlyon (1994) collected pitch matches using harmonic complexes filtered into the same spectral regions as those used in the present study, their complexes were not presented simultaneously. It is possible that the value and salience of the pitch evoked by one complex is influenced by the simultaneous presence of the other complex, and that PDI is related not to the relative pitch saliences of the two complexes measured in isolation, but to their relative saliences when presented simultaneously. Accordingly, in this experiment, we sought to measure the pitch and relative pitch salience evoked by each group while the other group was present. Listeners adjusted the $F0$ of a single group of harmonics, which was filtered into the lower or the higher spectral region, to match the pitch of a group of harmonics, which was filtered into the same spectral region but was presented *simultaneously* with a group filtered into the other spectral region. This was done for three different $F0$ separations between the two simultaneous groups: a “small” separation (1.75%), a large separation (20%), and an intermediate separation (7%).

B. Method

This experiment used the same basic stimuli as experiments 1 and 2, with 400-ms harmonic complexes filtered into the lower and higher spectral regions, as described above. Each trial of the pitch-matching procedure involved the successive presentation of two sounds. The first, referred to as the “reference” stimulus, consisted of two simultaneous groups of harmonics filtered into the lower and the higher spectral regions, exactly like the target and the interferer in experiment 1. The $F0$ of the higher-region group was always set relative to that of the group in the lower region, which was roved over a 6-semitone range, as in experiments 1 and 2. Three $F0$ separations between the two groups were tested: 1.75%, 7%, and 20%; these three $F0$ separations correspond to those between the target and the interferer for $\Delta F0$'s (between the two targets) of 3.5%, 14%, and 40%, respectively, in experiment 1. Both positive and negative $F0$ separations were tested, meaning that in one set of conditions the $F0$ of

the group filtered into the higher spectral region was higher than the F_0 of the group filtered into the lower spectral region, while in the other condition, it was lower. Note that this is similar to the situation in experiment 1, where the F_0 of the higher-region target was higher than that of the lower-region interferer in one observation interval, and lower in the other one. The different F_0 separations were tested in random order.

The reference stimulus was followed, after a 500-ms gap, by a “comparison” tone, which was filtered into a single spectral region. In one condition, that region was the lower one; in the other condition, it was the higher one. These two conditions were tested in separate blocks. At the beginning of each matching procedure, the F_0 of the comparison tone was selected randomly from a ± 18 -semitone interval around the 250-Hz nominal F_0 , in 2-semitone steps.

Listeners were instructed to adjust the pitch of the comparison tone so that it matched as closely as possible the dominant pitch evoked by the stimulus in the first interval. They were informed that the reference stimulus might evoke more than one pitch and that when this happened, they should try to match the pitch that was the most salient to them. By presenting the comparison tone in just one region, we hoped to encourage listeners to attend more selectively to that spectral region in the reference tone. Listeners increased or decreased the F_0 of the comparison tone at will, using six virtual buttons on the computer screen: two large buttons for ± 4 -semitone (about 26%) steps, two medium-sized buttons for ± 1 -semitone (about 6%) steps, and two small buttons for $\pm 1/4$ -semitone (about 1.45%) steps. A “hear again” button was also available. After each button press, the reference and comparison stimuli were regenerated with the appropriate parameters. The stimulus F_0 was roved over a 6-semitone range from one stimulus presentation to the next, keeping the relative F_0 difference between the reference and the comparison tone as last adjusted by the listener. This across-trial roving of F_0 had two objectives: first, to be consistent with experiments 1 and 2, and second, to encourage listeners to compare the two sounds presented on each trial despite their different timbre, rather than rely on some longer-term memory trace.

Each listener performed at least 20 matches in each of the 12 test conditions (six F_0 separations \times two comparison-tone filtering-region conditions). At the end of each match, the final value assigned to the F_0 of the comparison tone relative to the F_0 of the nominal F_0 (250 Hz) was stored. Pitch-match histograms were produced by counting the number of matches at each $1/4$ -semitone step over a 2-octave range (i.e., ± 1 octave) around the nominal F_0 . In each listener and for each condition, the pitch-match histograms were normalized by dividing the counts in each $1/4$ -semitone bin by the total number of matches over the whole 2-octave range. In this way, the height of the bars in the resulting histograms directly indicates the proportion of matches made at the corresponding relative frequency, out of the total number of matches for that condition and listener.

In order to estimate the proportion of matches made to the F_0 of the lower- and higher-region groups, the pitch-match histograms obtained in each condition were modeled

as a mixture of two Gaussians, defined each by three free parameters: mean, amplitude, and standard deviation. This was particularly useful for the smallest F_0 separation tested (i.e., 1.75%), as this separation was similar in size to the measurement resolution ($1/4$ th of a semitone is approximately equal to 1.43%). The best-fitting parameters were determined using a maximum-likelihood procedure, based on the assumption that the proportions of pitch matches in each frequency bin were distributed binomially. The maximum of the likelihood hypersurface was found using the Nelder-Mead simplex algorithm in MATLAB.

Three of the four listeners who had taken part in experiments 1 and 2 were involved in the current experiment. This experiment was performed after experiments 1 and 2.

C. Results

Pitch-match histograms based on the pooled data of the three listeners who took part in this experiment are shown in Fig. 4. When the comparison tone was filtered into the lower spectral region (left-hand column), the pitch matches were concentrated near the F_0 of the lower-region group in the reference stimulus, which corresponds to the zero mark on the abscissa: 94% of the matches fell within a $\pm 1/4$ -semitone range around the lower-region reference F_0 at the 1.75% separation (top panel), 80% at the 7% F_0 separation (middle panel), and 92% at the 20% separation (bottom panel). Furthermore, at the 7% and 20% separations, the vast majority (i.e., over 90%) of the matches was closer to the F_0 of the lower-region reference group than to that of the higher-region group. For the 1.75% separation, determining what proportion of the matches was made to each group is less straightforward, due to the similar size of that F_0 separation and the measurement accuracy, as mentioned above. At that separation, the histogram bars on either side of the central bin may reflect inaccurate matches to the F_0 of the lower group and/or matches to the F_0 of the higher group. If it is assumed that the listeners sometimes matched the F_0 of the lower group and sometimes that of the higher group, the proportion of matches corresponding to each group may be estimated by fitting the distribution of matches with a mixture-of-Gaussians with means equal to the F_0 's of the lower- and higher-region groups, as described in the Methods section. Doing this, we found that the estimated proportion of matches to the F_0 of the lower group was about 66% although, because the two distributions are so close, the proportion of matches assigned to each F_0 remains highly uncertain. Nevertheless, Fig. 4 contains an indication that the higher-region complex had some influence on the listeners' responses at the 1.75% F_0 separation: comparing the heights of the empty and filled histogram bars closest to the two vertical dashed lines in the upper left-hand panel, it can be seen that for negative relative F_0 values (i.e., the dashed line on the left), the empty bar, which corresponds to the case where the F_0 in the higher region was lower than that in the lower region, has a higher amplitude than the filled bar, which corresponds to the case where the F_0 in the higher region was higher than that in the lower region, while for positive relative F_0 values (i.e., the dashed line on the right),

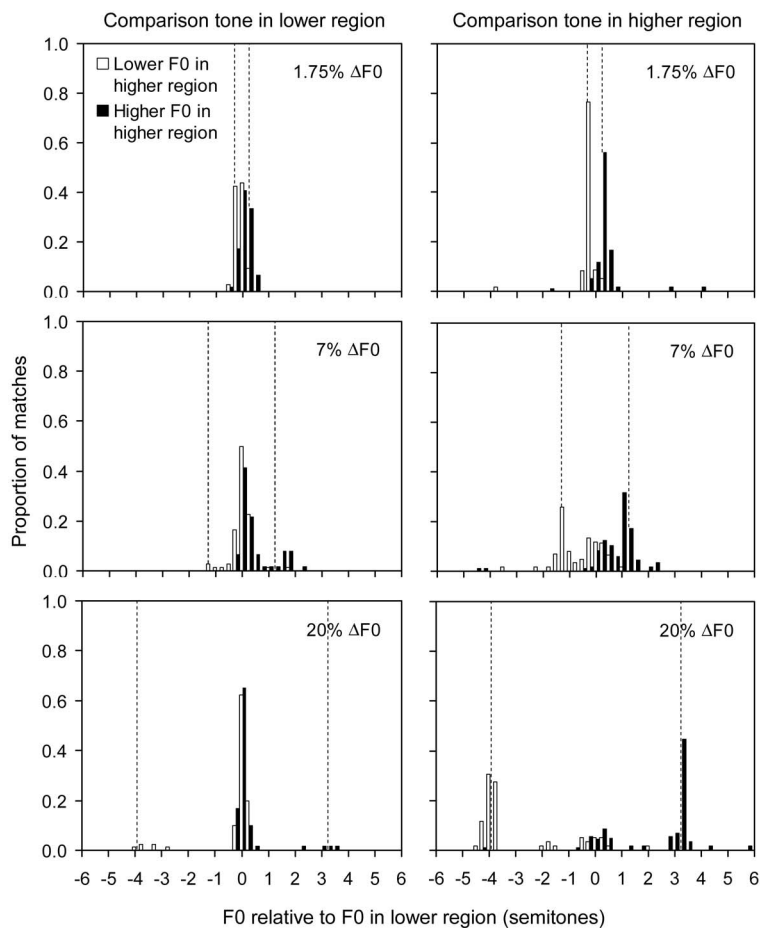


FIG. 4. Pitch-match histograms. The panels in the left-hand column show pitch matches obtained with the comparison tone filtered into the lower spectral region; those in the right-hand column, pitch matches obtained with the comparison tone in the higher spectral region. Each row corresponds to a given F_0 separation between the lower- and higher-region complexes in the reference stimulus: 1.75% (top), 7% (middle), and 20% (bottom); these correspond to the F_0 separations between the target and the interferer in the 3.5%, 14%, and 40% ΔF_0 conditions of experiment 1. The vertical dashed lines indicate the position of the F_0 of the higher-region complex in the reference stimulus, relative to that of the lower-region complex; the latter always corresponds to zero on the abscissa. The relative F_0 values on the abscissa are expressed in semitones; only matches falling between -6 and $+6$ semitones (corresponding, roughly, to $\pm 35\%$) around the reference (lower-region) F_0 are shown.

the converse is true. This suggests that, even if listeners always matched the F_0 of the complex in the lower spectral region, the perceived pitch of that complex was shifted towards the F_0 of the higher-region group. Although this shift of the perceived F_0 of the lower-region group in the direction of that of the higher-region group appears to be modest, it may have contributed to the drop in performance produced by the interferer at small ΔF_0 's in experiment 2.

When the comparison tone was filtered into the higher spectral region (right-hand column of Fig. 4), the majority of the pitch matches was close to the F_0 of the corresponding higher-region group in the reference stimulus, which is indicated by the vertical dashed lines. However, a substantial proportion of matches was also made to the F_0 of the lower-region group, as indicated by the central peaks (at 0 semitones) in the middle and lower right-hand panels. At the two corresponding F_0 separations of 7% and 20%, the proportions of matches to the F_0 of the lower region were estimated to be 48% and 25%, respectively, indicating that listeners tended to make more matches based on the F_0 of the lower-region group when the F_0 separation between the two groups was 7% than when it was 20%. In both cases, substantially more “across-region” matches were made in conditions where the comparison tone was filtered into the higher spectral region: less than 10% of the matches were across regions when the comparison tone was in the lower spectral region, and the maximum-likelihood mixture-of-Gaussian fits imputed only 3%–4% of matches to that F_0 . At

the 1.75% separation (top-right panel), no central peak can be distinguished. However, this could be due to peak overlap, and should not be taken as evidence that listeners never matched the F_0 of the lower-region group in this condition.

D. Discussion

The main finding from this experiment is the asymmetry in the pitch matches at the two larger F_0 separations, depending on which spectral region contained the comparison tone. When the comparison tone was in the lower spectral region, listeners usually matched its F_0 to that of the reference group filtered into the corresponding (lower) spectral region, and across-region matches were rare (less than 10% overall). However, when the comparison tone was in the higher region, listeners made a substantial proportion of matches (between about 20% and 44%) to the F_0 of the lower-region group. This asymmetry was observed at both the 7% and 20% F_0 separations, and is consistent with the hypothesis that when both groups were presented simultaneously, the pitch of the lower-region group was more salient or dominant than that of the higher-region group, even when the higher-region group was emphasized by the presence of the comparison tone.

For the 1.75% separation, things are less clear because, at this small separation, the relative inaccuracy of the pitch matches and the limited measurement resolution make it difficult to determine to which F_0 listeners were matching.

Therefore, the data do not permit a strong conclusion regarding the relative salience of the two groups of harmonics at small F_0 separations.

V. GENERAL DISCUSSION

A. Summary of results

The results from experiments 1 and 2 show that listeners' ability to discriminate changes in the F_0 of a group of resolved harmonics can be significantly impaired by a simultaneously presented group of resolved harmonics in a different (nonoverlapping) spectral region. The results extend the earlier findings of Gockel *et al.* (2004, 2005a) by showing that PDI is not restricted to situations in which the target contains only unresolved harmonics. The pitch matches of experiment 3 showed that the harmonics in the lower spectral region seemed to produce a somewhat more salient pitch than those in the higher spectral region. This difference in salience or dominance may explain why, at larger ΔF_0 's, PDI was only found when the target was in the higher spectral region (experiment 1). However, the differences in pitch salience are more difficult to reconcile with the fact that the PDI observed at smaller ΔF_0 's seemed to be independent of whether the spectral region of the target was higher or lower than that of the interferer.

B. Correspondence between performance and introspection: Are PDI and perceptual organization related?

When prompted, listeners typically reported that the simultaneous target and interferer were perceived as two separate sounds only when the F_0 separation between them was large, corresponding to about 20% (ΔF_0 of 40% in our experiments) or more. At F_0 separations between about 3.5% and 10% (i.e., ΔF_0 's between 7% and 20%), most listeners reported hearing only one sound, but did indicate that the stimuli sounded less "fused" than they did at smaller F_0 separations, below 3.5%.

These informal observations on how the stimuli sound bear some relationship to the pattern of results observed in our formal experiments. At very small F_0 separations, where listeners reported generally hearing a single fused sound, PDI was similar whether the spectral region of the target was above or below that of the interferer (Fig. 3). This may occur because the auditory system integrates the information from both regions and calculates the F_0 based on a sum of the information from all the components (e.g., Goldstein, 1973). However, in the present case, the weights assigned to the different harmonics appear not to be determined by the relative dominance of these harmonics (as estimated in other studies, e.g., Moore *et al.*, 1985; Dai, 2000), because PDI was found not to depend critically on the relative spectral positions of the target and interferer. Alternatively, at these small F_0 separations, PDI could be related to the unitary pitch evoked by the simultaneous target and interferer being less well defined than that evoked by the target alone. This could itself be due to the internal representations of the neighboring F_0 's merging into a broader peak—an effect that can be observed in the summary autocorrelograms pro-

duced by Meddis and O'Mard's (1997) model in response to stimuli consisting of two simultaneous sets of harmonics with different F_0 's: as the F_0 's become closer, the two corresponding peaks in the summary autocorrelogram start to overlap, forming a single, broader (less well-defined) peak.

At very large F_0 separations (e.g., 20%) between the target and the interferer, the F_0 difference between the two simultaneous groups is probably sufficient to overcome grouping based on common onset and offsets, allowing listeners to hear two separate pitches corresponding to the two F_0 's present. If so, any residual PDI observed at that large F_0 separation (when the target was in the upper spectral region) may have been due to the listeners confusing the pitch of the interferer with that of the target. The large roving range used in the experiments, and the fact that the interferer F_0 was always intermediate to the two target F_0 's, may have exacerbated this effect. This could explain why larger PDI effects were observed at these large F_0 separations between target and interferer here than in Gockel *et al.*'s (2004) first experiment. In that experiment, the F_0 separation between target and interferer was controlled by increasing the interferer F_0 above that of the target, so that at large separations the targets had a consistently lower pitch than the interferer; this perhaps made it easier for listeners to focus on the target pitch. Here, listeners could not adopt the strategy of listening consistently for the higher or the lower pitch, because the target had a higher pitch than the interferer in one interval, but a lower pitch in the other, and F_0 was roved across trials over a wide range.

At medium F_0 separations (ΔF_0 's of 14% and 20%), where the simultaneous target and interferer were generally heard as less fused than at smaller separations, yet not distinctly separated, strong PDI was found when the target was in the (less-dominant) higher spectral region (experiment 1), but no PDI remained when the target was in the (more-dominant) lower spectral region (experiment 2). As discussed above (cf. Sec. II C and Fig. 2), one possible explanation for the marked decrease in performance observed in the presence of the lower-region interferer at these ΔF_0 's is that the interferer, evoking a salient pitch, may have encouraged listeners to rely on shifts in the frequencies of individual harmonics (or in the spectral centroid) in the higher region, rather than on pitch changes; as illustrated in Fig. 2, this would have led to poorest performance at ΔF_0 's of 14% and 20%, as this is where spectral cues were most ambiguous. Another possible explanation for the large PDI effects at these ΔF_0 's, which was also suggested above, is that the lower-region interferer to some extent "suppressed" the pitch that would have otherwise been evoked by the target. The finding in experiment 1, that performance in the presence of the interferer was at (or very close to) chance at these ΔF_0 's, suggests that, in these conditions, the suppression was either complete or almost complete. This may appear to be inconsistent with the results of experiment 3, where it was found that even in the condition where the comparison group was in the lower region, listeners sometimes matched its F_0 to that of the higher-region group, indicating that they perceived the pitch of the latter. However, it is important to note that the stimuli in experiments 1 and 3 were not exactly the

same. Whereas in experiment 3 the first stimulus interval contained a single group of harmonics, in experiment 1, when the interferer was present, both intervals contained both groups. The latter design may have made it more difficult for listeners to perceptually segregate the two groups, which in turn may have made it harder to “hear out” the pitch of the less salient higher-region group, and at the same time, promoted suppressive interactions between the two groups. Thus, the fact that listeners did sometimes match to the pitch of the higher-region complex in experiment 3 does not imply that they always perceived the pitch of the higher-region complex in experiment 1.

C. Implications for models of pitch perception and source segregation

The results are consistent with models that integrate information across distant spectral regions in order to estimate pitch. However, they suggest that the channels on which pitch estimates are based cannot always be selected completely “at will,” based on the spectral region that they occupy. The results also suggest that $F0$ differences between the target and interferer of more than 10% (i.e., $\Delta F0$'s of more than 20%) are required for complete perceptual segregation based on $F0$. Although this difference appears rather large at first sight, particularly when compared with the mistuning of individual harmonics required to hear them as separate objects (Moore *et al.*, 1986), it is in fact compatible with experiments examining the effect of $F0$ differences on speech segregation (Darwin *et al.*, 2003), where differences in $F0$ between the target and interfering speech of at least 2 semitones (about 12%) were required for any benefit. This is different from the findings using two concurrent vowels (Scheffers, 1983; Assmann and Summerfield, 1990; Culling and Darwin, 1993); where a difference of 1 semitone (6%) already produces maximal segregation. However, it appears likely that within-channel cues, which were not available in our experiment, are important in mediating this effect (Assmann and Summerfield, 1994; Culling and Darwin, 1994).

D. Remaining questions

Some difficulties remain in reconciling our results with those of earlier studies. For instance, at small $F0$ separations, information from both spectral regions appears to be averaged, without regard to the pitch dominance region identified in earlier studies (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985; Dai, 2000; Gockel *et al.*, 2005b). Another issue is that the mistuning of one harmonic can lead to its perceptual separation from the rest of the complex at frequency differences of between 1% and 3% (Moore *et al.*, 1986), whereas our results suggest that much larger differences are needed to hear out two $F0$'s in different spectral regions. Also, whereas a single mistuned harmonic can be heard as a separate entity while still contributing to the overall pitch of the complex (Moore *et al.*, 1986), here one of the two possible explanations put forward earlier would imply that the $F0$'s in two separate spectral regions cease to be integrated at *smaller* $\Delta F0$'s than may be required to hear the two regions as separate tones. As suggested by Gockel

et al. (2004), part or all of these differences may be due to our use of two spectral regions, separated by noise, as opposed to earlier studies' use of contiguous harmonics within a single spectral region. The spectral separation may make the auditory system more tolerant of (or less sensitive to) mistuning across the two regions, and more willing to accept the two regions as emanating from a single source based on other cues, such as common onset and offset. In this respect, the background noise, which effectively masks the spectral gap between the two regions, may play an important role in fusing the two regions by making spectral continuity plausible in the same way that noise can make temporal continuity plausible (Grose *et al.*, 2002; Darwin and Hukin, 2004; Darwin, 2005). The influence of the background masking noise in pitch perception experiments clearly deserves further investigation. Finally, the present results raise the possibility that listeners modified their strategy in the $F0$ discrimination task, sometimes relying on $F0$ information, and sometimes on spectral-pitch or timbre cues, depending on the condition. This could be further investigated by using target complexes that are altered to reduce the utility of spectral cues by, for instance, roving the number of the lowest harmonic present (e.g., Houtsma and Smurzynski, 1990).

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant R01 DC 05216) and was completed while the authors were in the Research Laboratory of Electronics at the Massachusetts Institute of Technology. The authors are grateful to Hedwig Gockel for the time taken in helping to improve the clarity and presentation of this manuscript. Alain de Cheveigné, John Grose, and an anonymous reviewer also provided helpful comments on earlier versions of the manuscript.

¹Occasional cases where performance was 100% correct were dealt with by reducing the number of correct responses by one. With 300 trials, the highest possible proportion of correct responses after the application of this correction was approximately 0.997 (i.e., 299/300), which corresponds to a d' of approximately 3.8. As can be seen in Fig. 1, at the 14% $\Delta F0$, the average d' was around 3.5. This corresponds approximately to a proportion correct of 0.993, or a ratio of 298/300, indicating that at that separation, listeners made about two errors in 300 trials, on average. As this error rate was likely due to occasional attentional lapses, we consider that performance was essentially at ceiling in that condition.

²For this analysis, we assumed that the d' values corresponding to the no interferer condition at the 20% and 40% $\Delta F0$ were the same as those measured in the same condition at the 14% $\Delta F0$. This allowed all levels of the $\Delta F0$ factor to be included into the analysis. However, we also performed a similar analysis using only $\Delta F0$'s up to 14%, for which complete data were measured, both in the absence and in the presence of the interferer. This analysis also showed a significant main effect of interferer [$F(1, 4) = 248.69, p < 0.0005$]. For this and all other ANOVAs in this study, Greenhouse-Geisser correction was applied whenever appropriate, and the reported p values are the corrected values. The degrees of freedom indicated between parentheses with each F statistic are the original (uncorrected) degrees of freedom.

Assmann, P. F., and Summerfield, A. Q. (1990). “Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,” *J. Acoust. Soc. Am.* **88**, 680–697.

Assmann, P. F., and Summerfield, A. Q. (1994). “The contribution of waveform interactions to the perception of concurrent vowels,” *J. Acoust. Soc. Am.* **95**, 471–484.

- Beerends, J. G., and Houtsma, A. J. M. (1989). "Pitch identification of simultaneous diotic and dichotic two-tone complexes," *J. Acoust. Soc. Am.* **85**, 813–819.
- Bernstein, J. G., and Oxenham, A. J. (2003). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," *J. Acoust. Soc. Am.* **113**, 3323–3334.
- 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., van Wieringen, A., Long, C. J., Deks, J. M., and Wouters, J. (2002). "Temporal pitch mechanisms in acoustic and electric hearing," *J. Acoust. Soc. Am.* **112**, 621–633.
- Culling, J. F., and Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F_0 ," *J. Acoust. Soc. Am.* **93**, 3454–3467.
- Culling, J. F., and Darwin, C. J. (1994). "Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating," *J. Acoust. Soc. Am.* **95**, 1559–1569.
- Dai, H. (2000). "On the relative influence of individual harmonics on pitch judgment," *J. Acoust. Soc. Am.* **107**, 953–959.
- Darwin, C. J. (2005). "Simultaneous grouping and auditory continuity," *Percept. Psychophys.* **67**, 1384–1390.
- Darwin, C. J., and Hukin, R. W. (2004). "Limits to the role of a common fundamental frequency in the fusion of two sounds with different spatial cues," *J. Acoust. Soc. Am.* **116**, 502–506.
- 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.
- de Cheveigné, A., and Kawahara, H. (1999). "Multiple period estimation and pitch perception model," *Speech Commun.* **27**, 175–185.
- Gockel, H., Carlyon, R. P., and Micheyl, C. (1999). "Context dependence of fundamental-frequency discrimination: Lateralized temporal fringes," *J. Acoust. Soc. Am.* **106**, 3553–3563.
- Gockel, H., Carlyon, R. P., and Moore, B. C. J. (2005a). "Pitch discrimination interference: The role of pitch pulse asynchrony," *J. Acoust. Soc. Am.* **117**, 3860–3866.
- Gockel, H., Carlyon, R. P., and Plack, C. J. (2004). "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," *J. Acoust. Soc. Am.* **116**, 1092–1104.
- Gockel, H., Carlyon, R. P., and Plack, C. J. (2005b). "Dominance region for pitch: Effects of duration and dichotic presentation," *J. Acoust. Soc. Am.* **117**, 1326–1336.
- Goldstein, J. L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. Am.* **54**, 1496–1516.
- Grose, J. H., Hall, J. W., and Buss, E. (2002). "Virtual pitch integration for asynchronous harmonics," *J. Acoust. Soc. Am.* **112**, 2956–2961.
- Houtsma, A. J. M., and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," *J. Acoust. Soc. Am.* **87**, 304–310.
- Laguitton, V., Demany, L., Semal, C., and Liegeois-Chauvel, C. (1998). "Pitch perception: A difference between right- and left-handed listeners," *Neuropsychologia* **36**, 201–207.
- Licklider, J. C. R. (1951). "A duplex theory of pitch perception," *Experientia* **7**, 128–133.
- Meddis, R., and Hewitt, M. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **91**, 233–245.
- 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.
- Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1985). "Relative dominance of individual partials in determining the pitch of complex tones," *J. Acoust. Soc. Am.* **77**, 1853–1860.
- Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* **80**, 479–483.
- Plomp, R. (1964). "The ear as a frequency analyzer," *J. Acoust. Soc. Am.* **36**, 1628–1636.
- Plomp, R. (1967). "Pitch of complex tones," *J. Acoust. Soc. Am.* **41**, 1526–1533.
- 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.
- Ritsma, R. J. (1967). "Frequencies dominant in the perception of the pitch of complex sounds," *J. Acoust. Soc. Am.* **42**, 191–198.
- Scheffers, M. T. M. (1983). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleeck, S., Stippich, C., and Rupp, A. (2005). "Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference," *Nat. Neurosci.* **8**, 1241–1247.
- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* **95**, 3529–3540.
- Smoorenburg, G. F. (1970). "Pitch perception of two-frequency stimuli," *J. Acoust. Soc. Am.* **48**, 924–941.

Illusory percepts from auditory adaptation

Lucas C. Parra^{1,a)}

¹*Biomedical Engineering Department, City College of New York, New York, NY 10031*

Barak A. Pearlmutter^{2,b)}

²*Hamilton Institute, National University of Ireland Maynooth, Co. Kildare, Ireland*

(Received 25 April 2006; revised 6 December 2006; accepted 13 December 2006)

Phenomena resembling tinnitus and Zwicker phantom tone are seen to result from an auditory gain adaptation mechanism that attempts to make full use of a fixed-capacity channel. In the case of tinnitus, the gain adaptation enhances internal noise of a frequency band otherwise silent due to damage. This generates a percept of a phantom sound as a consequence of hearing loss. In the case of Zwicker tone, a frequency band is temporarily silent during the presentation of a notched broadband sound, resulting in a percept of a tone at the notched frequency. The model suggests a link between tinnitus and the Zwicker tone percept, in that it predicts different results for normal and tinnitus subjects due to a loss of instantaneous nonlinear compression. Listening experiments on 44 subjects show that tinnitus subjects (11 of 44) are significantly more likely to hear the Zwicker tone. This psychoacoustic experiment establishes the first empirical link between the Zwicker tone percept and tinnitus. Together with the modeling results, this supports the hypothesis that the phantom percept is a consequence of a central adaptation mechanism confronted with a degraded sensory apparatus. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431346]

PACS number(s): 43.66.Ki, 43.66.Mk, 43.66.Ba, 43.66.Dc, 43.66.Lj [RAL] Pages: 1632–1641

I. INTRODUCTION

A. Illusory auditory percepts

Tinnitus is the perception of a phantom sound often associated with hearing loss.¹ Mild tinnitus is rather common, reported by many subjects after a few minutes in a quiet environment.^{2,3} The subjective sound varies, often described as a “buzz,” “ring,” “hiss,” “hum,” or the like. Severe tinnitus is almost always indicative of hearing loss,⁴ with the pitch of the phantom sound generally corresponding to the frequencies of hearing loss⁵ and occurring predominantly at sharp edges of high-frequency loss.¹ To date, a variety of therapeutic approaches to alleviate tinnitus have given mixed results.^{3,6} It is therefore generally assumed that tinnitus may be the result of multiple physiological causes.⁷ It is believed that in most cases the tinnitus percept does not originate immediately at the cochlea. Instead it has often been associated with adaptive phenomena in the central nervous system.^{4,8} A variety of models for the physiological origin of this form of central tinnitus have been proposed.^{9,10}

The Zwicker tone, an auditory perceptual illusion named after the scientist who first characterized it, is a transient phantom sound that is perceived by most subjects after a notched broadband signal.^{11–13} The frequency of the illusory sound falls within the notched frequency band and is closer to the lower stop band edge for a wide-band notch. The strength and duration of the Zwicker tone percept depends on stimulus conditions and is quite variable across subjects.^{14,15} In fact, in the study presented below, some subjects with normal hearing did not hear a Zwicker tone.

Despite their apparent similarity, the exact relationship between the Zwicker tone and tinnitus is not well established. This paper (a) presents a conceptual model, fleshed out mathematically, that gives a common explanation for both phenomena; (b) shows that this model predicts a linkage between tinnitus and Zwicker tone; and (c) exhibits psychophysical data which matches the prediction of the model. The psychophysical data does support the proposed model, but also considered in isolation constitutes a novel empirical link between tinnitus and the Zwicker tone.

The remainder of the paper starts, in Sec. I B, with a review of the standard theoretical understanding of simple sorts of adaptation as mechanisms for matching sensory statistics. Section I C motivates a model of auditory adaptation based on these principles and Sec. II instantiates the model mathematically. Simulations of this model in Sec. III suggest an empirical link between tinnitus and the Zwicker tone. This link is confirmed by the psychosocial observations reported in Sec. IV.

Taken together, the experimental and modeling results support the hypothesis that tinnitus is a consequence of a gain adaptation mechanism that is confronted with hearing loss and an associated loss of nonlinear compression. Section V presents additional experimentally testable predictions that follow directly from this theory. Specifically, this predicts a relationship between the strength of nonlinear auditory phenomena such as combination tones^{16–18} with the sensitization observed following notched noise.^{19,20} Finally, it is noted that elevated hearing thresholds and loss of nonlinear compression are easily measured and may be partially restored with compressive hearing aids. This suggests a potential diagnosis and treatment option for those cases of tinnitus that can be linked to this particular form of hearing deficit.

^{a)}Electronic mail: parra@ccny.cuny.edu

^{b)}Electronic mail: barak@cs.nuim.ie

B. Adaptation and optimality

Gain and contrast adaptation is a common strategy of the perceptual system to match a large dynamic range of natural signals to the limited dynamic range of sensors and neurons.^{21,22} Perhaps the best known gain adaptation mechanism is the closing of the iris when stepping from a dark environment into a bright one. The analogous effect in hearing is the acoustic middle-ear reflex, which mechanically attenuates sound transmission to the cochlea in response to loud sounds.⁸

Adaptation to changes in stimulus statistics is a ubiquitous and long-studied phenomenon in the nervous system.²³ Visual neurons in the retina and visual cortex adjust the gain of their transfer functions to maintain a high sensitivity at varying luminance contrast levels.^{24–26} This allows the visual system to operate well under drastically varying external conditions. In the auditory system, adaptation is observed at various levels. Efferent feedback to outer hair cells are thought to control the gain of cochlear amplification,²⁷ while auditory nerve fibers are known to adapt their firing rate at various time scales.^{28,29} Finally, inferior colliculus neurons have been shown to adjust their response thresholds and gains to optimally encode variations in the auditory stimulus.^{30,31}

The main theoretical contribution of this paper is to demonstrate that some illusory auditory percepts can be explained as direct consequences of gain adaptation and internal noise in the presence of hearing loss. Gain adaptation and noise are basic features of the auditory processing stream.^{8,32,33} Since gain adaptation may operate at various levels of processing, a simple and generic model is constructed. It is then shown that, after gain adaptation, model frequency bands with reduced external input (due to permanent hearing loss or temporary deprivation) show enhanced steady-state activity resembling phantom sounds.

A generic argument is purposefully made, in order to be applicable at many stages of auditory processing, rather than suggesting which area or areas actually subserve this functionality. The auditory nerve is, however, discussed as one potential site where such a mechanism may play a role. The model is sufficiently generic and therefore similar phenomena are expected for any system in its broad general class: systems that perform local gain adaptation in the context of a global estimate of the stimulus energy.

C. Information bottleneck and gain adaptation

The cochlea transforms acoustic signals into neuronal activity by decomposing the signal into its various frequency components, which are then transmitted by the auditory nerve to the midbrain. The signal intensity in different frequency bands is encoded in the firing of different neurons which project into the auditory nerve. However, the dynamic range of the external stimuli is known to be much larger than the dynamic range of this neuronal activity.

Transmitting auditory information through this information bottleneck therefore requires adaptive mechanisms. The nervous system has developed various strategies to cope with this problem including, in particular, gain adaptation. It can

be argued that the main goal of the adaptive mechanism should be to transform the signal in different frequency bands into independent channels with optimally matched dynamic ranges. In this view, gain adaptation accomplishes two tasks. First, it adjusts signal variance to the effective dynamic range of each frequency channel, thus optimizing the information capacity in each frequency channel.²² Second, it removes redundancy across channels. Most acoustic signal have significant redundancy across frequency bands due to frequency comodulation—the simultaneous increase and decrease of amplitude in multiple bands.³⁴ In fact, humans can understand spoken language with as few as four distinct frequency bands, supporting the notion that information is redundantly encoded across different frequencies.³⁵ By normalizing signal power, channels become more independent.³⁶ A similar mechanism for reducing redundancy by divisive normalization has been proposed for visual processing^{37–39} and can be used for image compression.⁴⁰

A channel with fixed dynamic range will communicate maximum information if the transfer function matches the cumulative density function (CDF) of the input variable.⁴¹ In particular, the threshold and slope of the transfer function should match the mean and variance of the data. By adjusting the mean and/or variance of the input, one can optimize the information transmission for a given transfer function. The first-order correction is achieved by adjusting the variance (or power) of the signal to match the transfer function.⁴²

II. METHODS (MODELING)

A. Gain adaptation model

The model assumes (a) that instantaneous signal power is transduced by the cochlea into a neuronal signal and (b) that neuronal transmission has some inherent noise such as spontaneous background firing. $|S(\omega, t)|^2$ denotes the instantaneous power of the stimulus in frequency band ω at time t . If the random neuronal noise $N(\omega)$ is independent of the stimulus, the neuronal representation of signal power $P(\omega, t)$ in each frequency band is simply the sum of the signal and noise

$$P(\omega, t) = |S(\omega, t)|^2 + N(\omega). \quad (1)$$

It seems natural to suggest that gain adaptation should normalize this signal so that it lies within a limited dynamic range. The normalization should occur separately for each frequency band, using a long-term average of stimulus power. Computationally, the simplest estimate of signal power is a running average, which updates the long-term power estimate with the currently observed instantaneous power values. In a discrete-time formulation this is given by

$$\bar{P}(\omega, t) = (1 - \mu)\bar{P}(\omega, t - \Delta t) + \mu P(\omega, t), \quad (2)$$

where $P(\omega, t)$ is the neuronal representation of instantaneous stimulus power, $\bar{P}(\omega, t)$ is an estimate of the long-term average power, and $\mu < 1$ is a fraction that captures how much the current estimate contributes to the long-term average.⁴³ This is equivalent to a temporal integration of instantaneous powers with an integration window w_n :

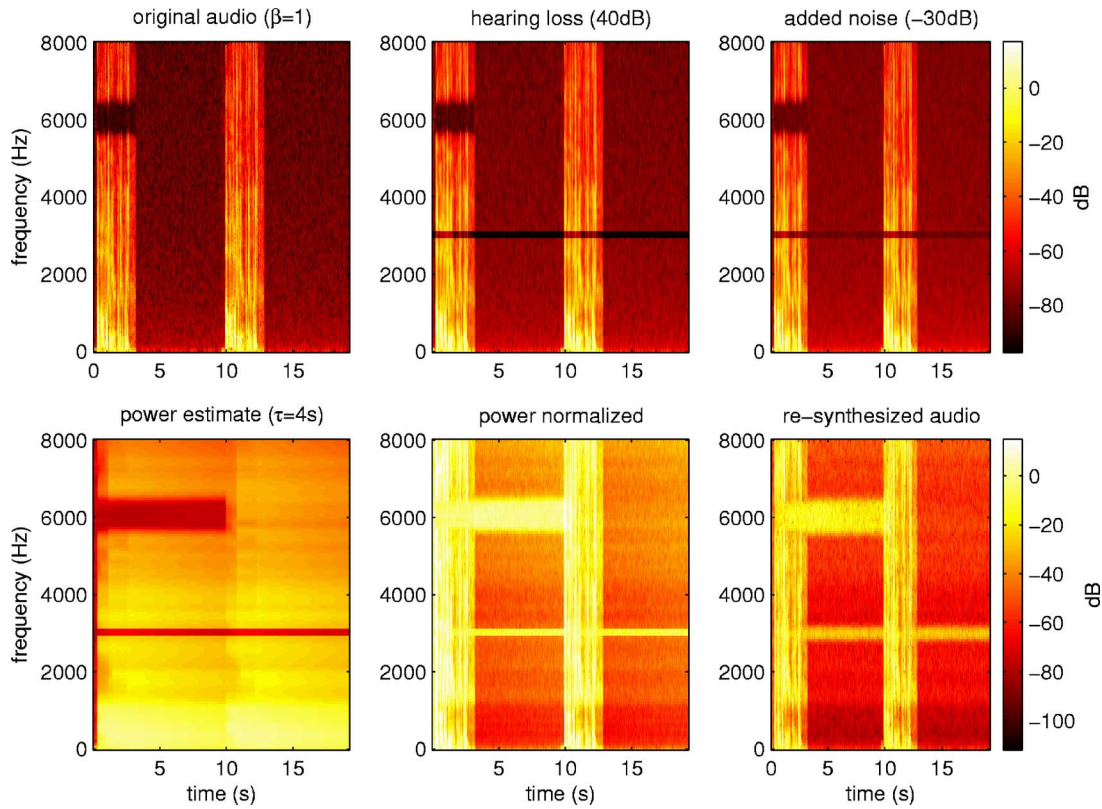


FIG. 1. (Color online) Stages of processing showing how phantom auditory percepts are generated by gain adaptation. A constant tone resembling tinnitus is generated at the frequency of hearing loss (3 kHz here). A transient tone resembling the Zwicker tone is generated following the notched sound (6 kHz here). The auditory signal is first decomposed into a time-frequency representation (top left). Frames of 16 ms (256 samples at a 16 kHz sampling rate) around time t are windowed with a Hanning window and Fourier transformed to obtain 128 frequency amplitudes $|S(\omega, t)|$ (top left) and phases $\arg[S(\omega, t)]$ (not shown). Image intensity represents power in decibels using the same color map for each row. Perceived powers (top center) are given by $H(\omega)|S(\omega, t)|^2$. Noise with a power profile $N(\omega) \propto 1/\omega$ is added to the perceived powers, giving the signal $P(\omega, t)$ (top right) according to Eq. (6). The gain and equalized signal powers (bottom left and center) are computed with Eqs. (4) and (5) using a time constant of $\tau=4000$ ms. The original signal powers are estimated from this activity using Eq. (8) and the signal is resynthesized. A spectrogram of this resynthesized signal is shown on the bottom right.

$$\bar{P}(\omega, t) = \sum_{n=0}^{\infty} w_n P(\omega, t - n\Delta t), \quad (3)$$

where the weights decay exponentially, $w_n \propto (1 - \mu)^n$. In order to ensure that the integration window is normalized, i.e., that $\sum_n w_n = 1$, a constant is introduced so $w_n = \mu(1 - \mu)^n$. The integration time can be characterized by a time constant $\tau = \Delta t / \mu$.

Now let the normalization gain for each band be

$$G(\omega, t) = \frac{1}{\bar{P}(\omega, t)} \quad (4)$$

so the normalized neuronal representation of stimulus power is

$$E(\omega, t) = G(\omega, t)P(\omega, t). \quad (5)$$

This normalization ensures that the neuronal signal is within some fixed dynamic range, and also removes the cross-frequency redundancy of comodulated power envelopes.

As seen below, these simple assumptions motivated by considerations of optimal signal processing give rise to illusory percepts resembling tinnitus and Zwicker tone in response to a reduced input in a given frequency band.

B. Sensitivity and hearing loss

The perceived signal intensity in each frequency band is affected by the sensitivity of the cochlea at that band. This is expressed by some gain function $H(\omega)$, with $H(\omega)|S(\omega, t)|^2$ replacing $|S(\omega, t)|^2$. Hearing loss is modeled by reducing $H(\omega)$ for the damaged bands. The broadening of the bandwidth associated with hearing loss has not been modeled here.

C. Compression of the cochlear amplifier

Note that the present model is linear in power and therefore does not include the nonlinear compression typically found for an intact cochlea.⁴⁴ Outer hair cells are thought to actively amplify faint sounds with high gain, whereas loud sounds are transduced with a lower gain.⁴⁵ This nonlinear amplification leads to a compression of dynamic range. Theoretical models⁴⁶ as well as psychophysical experiments⁴⁷ describe the compressive effect as power laws with powers between 0.3 and 0.5 for normal hearing. A power law is easily included in the present model with a power factor β . Together with the sensitivity discussed above, this gives

$$P(\omega, t) = H(\omega)|S(\omega, t)|^{2\beta} + N(\omega). \quad (6)$$

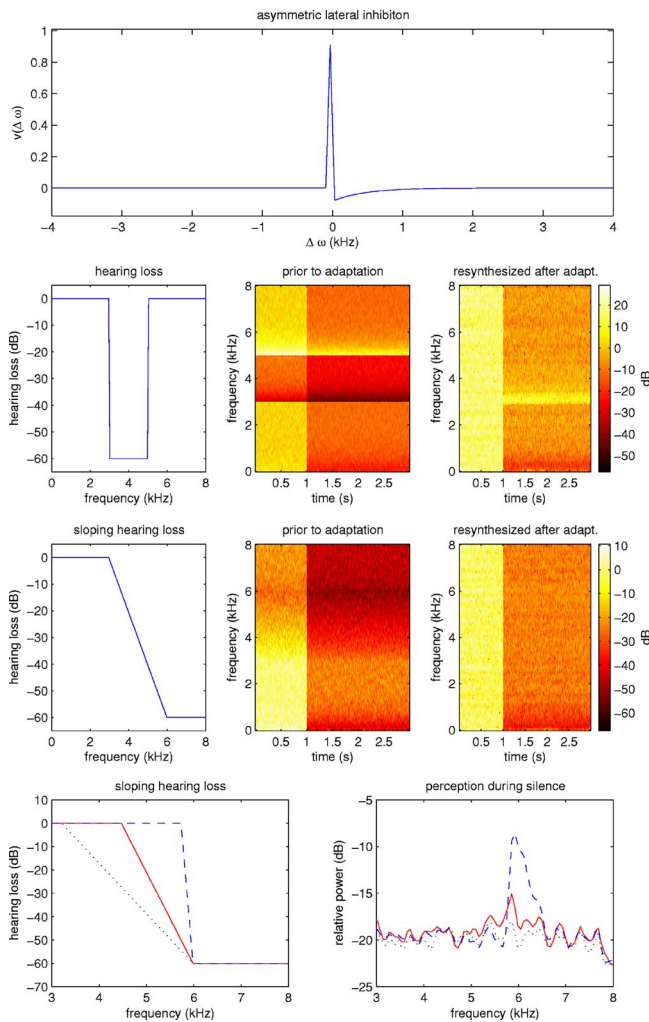


FIG. 2. (Color online) Effect of asymmetric lateral inhibition and the slope of hearing loss on the tinnitus percept. Top graph: Asymmetric lateral inhibition function $v(\Delta\omega)$. Second row: Loss function $H(\omega)$ with a sharp band gap (left panel) generates a phantom sound above the lower edge of the gap as evidenced by the spectrogram of the reconstructed signal (right panel). Center panel shows time-frequency representation of the response prior to adaptation. Third row: Panels as in second row, but with a sloping hearing loss. When simulating this with lateral inhibition (as in top graph) sloping loss does not generate a phantom sound (right panel). Fourth row: Comparison of the phantom percept during the silence period (1000–3000 ms) for different slopes of hearing loss. The right panel shows the average power during the 2000 ms of silence following the 1000 ms white noise stimulation, and is given in decibels relative to the total power of the preceding noise. An increasingly sharp loss accentuates the phantom percept. The phantom sound is at the lower edge of the loss band.

A power factor $\beta < 1.0$ reflects normal compression of the cochlear amplifier, whereas $\beta = 1.0$ reflects a lack of active amplification⁴⁸ resulting in a sharper increase in firing rate with increasing signal power observed for the damaged cochlea.⁴⁹

D. Lateral inhibition

The pitch of tinnitus is sometimes perceived at pronounced edges of hearing loss.^{10,50} Similarly, for wide-band notched noise the Zwicker tone is perceived preferentially at the lower edge of the notched band.¹³ To explain this phenomenon most models of tinnitus and Zwicker tone include some form of lateral inhibition.^{10,11,51,52} In fact, they require

asymmetric inhibition whereby high frequencies tones inhibit lower frequency units. Such asymmetric lateral inhibition has been reported in the auditory nerve, cochlear nucleus,⁵³ inferior colliculus,⁵⁴ and auditory cortex.⁵⁵ To demonstrate the effect of asymmetric lateral inhibition in the present model, the excitation can be modified

$$\log|S(\omega, t)| = \sum_{\Delta\omega} v(\Delta\omega) \log|S(\omega - \Delta\omega, t)| \quad (7)$$

with $v(\Delta\omega)$ representing the interaction of frequency bands separated by $\Delta\omega$. A simulation incorporating such asymmetric lateral inhibition is shown in Fig. 2, and corresponds to lateral inhibition prior to adding noise and gain adaptation (see discussion in Sec. V C). Similar effects are obtained if the asymmetric lateral inhibition is introduced after gain adaptation.

E. Recovered signal

To interpret the neuronal representations after gain adaptation, the neuronal model signal can be used to construct an estimate of the original signal. This step may seem artificial, as the nervous system does not need to regenerate the original signal in order to perceive it. Rather, the neuronal representation either itself constitutes, or is the precursor of, perception. Regardless of the physical basis of auditory qualia, it can be argued that if the representation is altered such that the stimulus cannot be regenerated, even approximately, then the percept must be equivalently distorted, and that the reconstruction technique provides an intuitive way to measure and visualize the distortion of the neuronal representation. Here, this method is used to show that the regenerated signals after gain adaptation exhibit artifacts that would be perceived as phantom sounds.

To interpret (and hear) the adjusted signal, an attempt must be made to reproduce the original signal $S(\omega, t)$ from $E(\omega, t)$. If the activity in band ω is associated with $E(\omega, t)$ and the total power of the signal is given by $P(t) = \sum_{\omega} P(\omega, t)$, one can estimate the original signal with

$$|\hat{S}(\omega, t)|^2 = P(t)E(\omega, t). \quad (8)$$

The assumption here is that the system does not know the varying gain it has applied to the signal. However, it does know the overall loudness of the signal as represented by $P(t)$.

Note that gain normalization as proposed here removes the common power of the signal on the time scale τ , so the overall loudness of a signal is therefore no longer reflected in the individual perceptual channels. Stationary silence would therefore be indistinguishable from loud stationary noise. The postulated mechanism therefore implies that the common signal power $P(t)$ must be separately encoded. For a frequency comodulated signal, power is redundantly distributed across bands. Removing this comodulation removes the redundancy and makes more efficient use of the information capacity of the channel. Communicating overall power as a variable separate from the power fluctuation in each frequency band is therefore a more efficient use of channel information capacity.

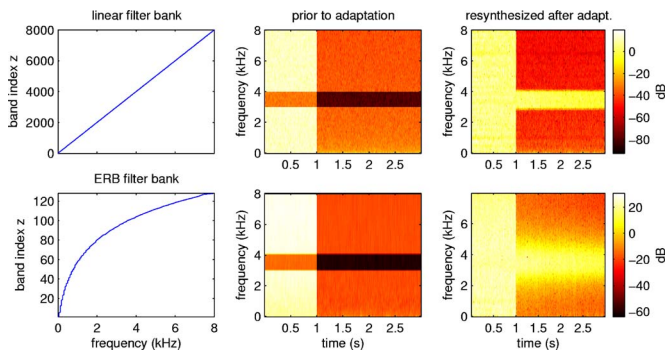


FIG. 3. (Color online) Effect of perceptual frequency scale. The center column shows a time-frequency representation of the response prior to adaptation, while the right column shows a spectrogram of the reconstructed signal following gain adaptation. Color represents signal intensity in decibels. First row: Effect of uniform hearing loss with a linear frequency scale as in Fig. 1. Second row: Perceptual frequency bands broaden the effect due to their broad overlapping frequency response. The best frequency of the filter bank is shown on the horizontal axis of the left panel. Compare this to the linear scale on the top-left panel.

Finally, in order to listen to the recovered signal, a time-domain signal must be generated from the time-frequency powers. The powers give amplitude but not phase information. This is a common problem in speech and sound synthesis. A standard engineering solution to this problem is to reuse the phase that was obtained when analyzing the original signal. This is done using a conventional overlap-add procedure: powers are combined with the original phase, $\arg[S(\omega, t)]$, inverse Fourier transformed, multiplied with a Hanning window, and added in half-overlapping frames. If the powers have not changed significantly, the resulting signal is perceptually similar to the original.

III. MODELING RESULTS

Hearing damage can be simulated by reducing the sensitivity $H(\omega)$ in a narrow frequency band. Figure 1 shows the result for a 40 dB hearing loss at 3 kHz and -30 dB internal noise. The lower right panel shows that gain adaptation generates steady-state power at the damaged frequency band. The reconstructed signal produces a sound similar to tinnitus.

The figure also shows the results obtained for a broadband sound with a notched response (power reduced by 40 dB at 6 kHz). Power normalization fills in the gap and generates an artificial tone following the notched noise. This is consistent with the Zwicker tone phenomenon. The increased gain following the noise is consistent with the sensitization²⁰ and release of masking¹⁹ observed after notched noise stimuli in psychophysical studies.

The effect of asymmetric lateral inhibition, and of the slope of the band edge, can now be analyzed. Lateral inhibition is used as shown in Fig. 2, and given by $v(\Delta\omega) = \delta(\Delta\omega) - c \exp(-\Delta\omega/B) \theta(\Delta\omega)$, where c and B determine the strength and bandwidth of the inhibitory effect, and $\theta(\Delta\omega)$ is a step function that implements the asymmetry. Figure 2 shows that for a sharp band gap the phantom sound is narrow band and appears at the lower edge of the band gap. In fact, the phantom sound is generated even in the absence of the high-pass edge. This is all in agreement with psychophysical

reports on Zwicker tone and tinnitus.^{13,50,51} Furthermore, under the assumption of lateral inhibition, it is found that a sloping edge will not cause a phantom percept, while an increasing slope will cause a stronger phantom percept. Also note that the pitch of the percept appears at the position of strongest loss and not the onset of the loss edge. These observations are consistent with empirical finding on tinnitus and nontinnitus subjects with varying hearing loss profiles.¹

The effects of a perceptual frequency scale, where bandwidth increases with center frequency,^{56,57} can also be analyzed (Fig. 3). The resynthesized signals show the same qualitative behavior, except that the phantom sound is broadened as a result of the broad bands on the perceptual scale.

The relevant parameters of the present model, most of which are summarized in Eq. (6), consist of: the level of signal loss $H(\omega)$, the amount of internal noise $N(\omega)$, the compression factor β , and the time constant τ which determines the integration time in Eq. (3). Figure 4 shows the effect of each of these parameters on the phantom sounds. The intensity of the phantom sounds increases with the level of internal noise and with the loss of signal intensity, but is fairly independent of τ . It is important to note that the Zwicker tone is decreased in magnitude and duration when considering the effect of nonlinear compression of the cochlear amplifier with typical compression factors of $\beta < 0.5$.

IV. MODEL PREDICTION AND EVALUATION

The modeling results shown in Fig. 4 (bottom panel) indicate that the Zwicker tone percept is strongest in the case of reduced compression. There is some evidence that tinnitus may be associated with a loss of compression,^{47,48} which according to the above model would also accentuate the perception of the Zwicker tone. When normal and tinnitus subjects were tested for differential perception of a Zwicker tone, it was found that in fact the two phenomena were empirically linked.

The perception of the Zwicker phantom tone following a two-second notched noise varied across subjects. Subjects reported different percepts, describing them variously as a “tone,” “hiss,” or “ringing” lasting a brief moment after the notched noise. More than half the subjects (54%) perceived a sound of varying strength for different notched bands, while others did not perceive a phantom tone following any of the notched noise sounds (47%) (see Sec. IV A).

All subjects were asked if, in their daily lives, they perceived spurious ringing on a regular basis. 25% of the subjects (11 of 44) responded positively.⁵⁸ There was a correlation between this self-reported tinnitus and the perception of the Zwicker tone ($r=0.42, p=0.004$). Table I and Fig. 5 show that tinnitus subjects are almost certain to hear a Zwicker tone, whereas fewer than half of normal subjects heard a phantom tone following a 2-s notched noise. These numbers show a significant association between self-reported tinnitus and perception of the Zwicker tone ($p=0.006$, two-tailed Fisher exact test).

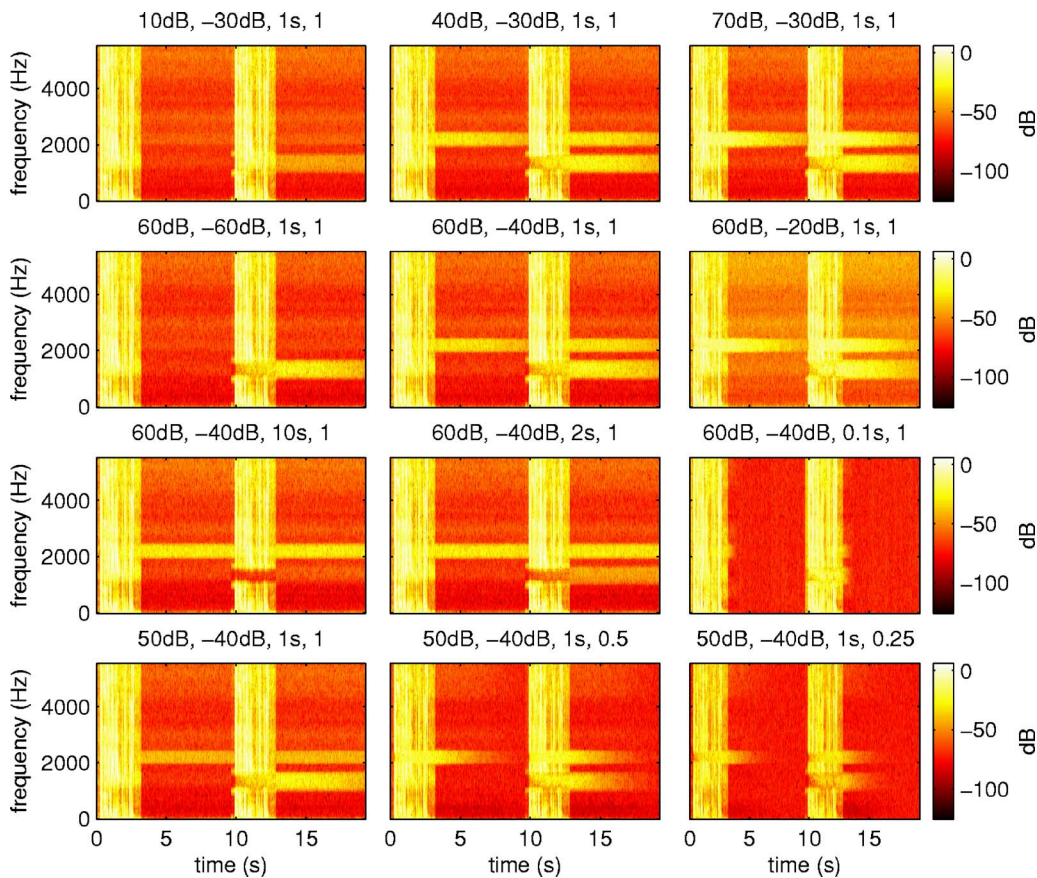


FIG. 4. (Color online) Dependence of reconstruction on signal loss, noise magnitude, power-averaging time constant τ , and nonlinear compression factor β (the heading of each panel specifies these parameters in that order). The two phantom percepts occur variably depending on the specific parameters. Importantly, the Zwicker tone is increased with abnormally high β corresponding to reduced cochlear compression.

A. Methods (psychophysical)

1. Subjects

Forty-four volunteers were recruited (22 male, 22 female, age 28 ± 8) among faculty and students at CCNY in accordance with the CCNY IRB guidelines. Subjects gave informed consent prior to experimentation. In this sample there was no significant correlation between the subject's age and self-reported tinnitus, nor between age and the Zwicker percept. The presence of the Zwicker percept was determined with the following procedure.

2. Zwicker tone perception test

Subjects were tested to determine if they reliably perceived a phantom tone by presenting four different noise sounds in random order. The control sound was white noise and the other three were notched noise with different notch

TABLE I. Number of normal and tinnitus subjects reporting Zwicker tone percept.

Tinnitus	Zwicker Perceived		Total
	Yes	No	
Yes	10	1	11
No	14	19	33
Total	24	20	44

bands, as described below. Subjects were instructed to report which of the four noises was followed by a perception of some form of ringing, however, faint it might be. The percept was considered factual if the subject consistently reported a percept for the same notched sounds (despite the random ordering) but not the white noise. Subjects that did not report any phantom percept, or that gave inconsistent answers to this test, were considered to not perceive the Zwicker tone.

3. Stimuli

The amplitude of the notched noise rises linearly for 1000 ms, holds for 1000 ms, and decays within 40 ms. The

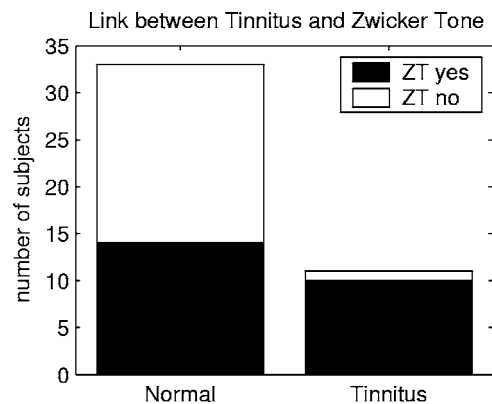


FIG. 5. Number of normal and tinnitus subjects reporting Zwicker tone percept. The Zwicker tone percept was significantly associated with tinnitus self-report ($p < 0.01$).

band gap of the notched noise is 4 kHz wide starting at 500, 1000, or 2000 Hz. The noise sequences were presented with a silence gap of three seconds. The signals were generated on a personal computer (PC) using MATLAB by zeroing the corresponding frequencies in the Fourier domain. They were reproduced using an Audiobook MAYA44 USB external digital-to-analog converter, and delivered binaurally with audio-technica ATH-M40f headphones at approximately 50–60 dB sound pressure level (SPL) adjusted for comfort. (Note that standard soundcards in today's PCs can typically not reproduce sharp band gaps.)

V. DISCUSSION

Illusory visual percepts were once thought to constitute regimes where the visual system breaks down and fails to process the data appropriately. For a number of broad classes of stimuli, this is no longer the accepted explanation. For example, many motion illusions can be explained as a consequence of Bayesian inferences being made from noisy data.⁵⁹ The present work has extended this to the auditory system, where it is proposed that a simple adaptive mechanism, when driven outside its normal operating regime, may generate uncertainties which makes the optimal interpretation contain illusory percepts. Specifically, the psychophysical and modeling results reported above support the hypothesis that tinnitus and the Zwicker tone may be a consequence of gain adaptation, and that the loss of compressive nonlinearity may accentuate and modify these percepts even in the absence of elevated hearing thresholds.

Taken together, the modeling and psychophysics results suggest distinct regimes of operation for normal and tinnitus subjects. It might be reasonably speculated that tinnitus subjects have lost the instantaneous amplification mechanism of outer hair cells in selective bands; that this disrupts the dynamic range compression inherent in the nonlinear amplification mechanism; and that as a result, a slower neuronal gain adaptation mechanism becomes the dominant factor.

A. Evidence related to tinnitus and hearing loss

In the auditory periphery there are at least two mechanisms that are thought to address the problem of dynamic range mismatch between the auditory nerve fibers, which lies between 20 and 40 dB, and the dynamic range in the auditory input of about 120 dB. First, outer hair cells are thought to actively amplify faint sounds with large gains, while at high signal intensities the gain is reduced. This nonlinear amplification leads to a compression of dynamic range. Second, inner hair cells are contacted by multiple auditory fibers with different response thresholds and gains. Therefore, as intensity increases an increasing number of fibers are recruited, which effectively increases the available dynamic range of neuronal firing for a group of fibers with a common characteristic frequency.

Peripheral hearing loss is associated with elevated thresholds. This results in a reduced diversity of response thresholds required by the recruitment mechanism. This is thought to be the origin of abnormally fast growth in

loudness.⁶⁰ In addition, outer hair cell damage, which is often associated with peripheral hearing loss, leads to a loss of active amplification, reducing the compressive effect of the nonlinear cochlear amplifier.⁶⁰ Here it is postulated that, when faced with these challenges, downstream mechanisms compensate by taking a more active role in coping with the dynamic range of the input. These mechanisms, when confronted with silence in selected frequency bands, increase internal gains, which then amplify neuronal noise to the point that it is perceived as phantom sounds. Tinnitus and the Zwicker tone, in this view, are not associated with increased activity in the periphery, yet are nonetheless ultimately caused by alterations in the peripheral apparatus.

Note that elevated thresholds are a common correlate of tinnitus,⁴ and abnormal growth of loudness is observed for frequencies matching the tinnitus percept.⁶¹ In addition, distortion products, which are thought to reflect the operation of the nonlinear cochlear amplifier, are selectively altered for frequency bands that have been matched to the tinnitus percept.^{62–64} Finally, release from masking by a secondary masking tone does not occur in tinnitus subjects, indicating once again that the nonlinear effect of this two-tone suppression ascribed to the cochlear amplifier is not operating in tinnitus subjects.^{65,66} All this supports the hypothesis that tinnitus is a result of hearing loss and degraded nonlinear compression.

A common strategy to alleviate tinnitus consists of masking the tinnitus percept with acoustic noise in the corresponding frequency band. While this method is effective in eliminating the tinnitus percept for the duration of the noise, it is seldom adopted by patients, as it accomplishes little more than replacing one auditory annoyance by another. Interestingly, a residual inhibition following the masking noise and lasting up to minutes is commonly observed.⁶⁷ It has also been reported that hearing aids properly fitted to the frequencies of hearing loss can sometimes alleviate tinnitus.^{3,68,69} Some reports indicate that tinnitus can be alleviated on a longer time scale by delivering *variable* signals in selected frequency bands,⁷⁰ in particular after noise-induced hearing loss preempting central adaptation.⁷¹ Perceptual training paradigms aiming at central adaptation mechanisms have also been used to alleviate tinnitus percepts.⁷² All this is in good agreement with the hypothesis proposed above, which maintains that the increased gains can be reduced by delivering signal *variance* to the damaged channel. This moreover suggests that a properly fitted compressive hearing aid may alleviate tinnitus for those subjects where tinnitus is caused by a loss of nonlinear amplification and/or a partial loss of sensitivity.

B. Neural substrate

The above model of gain adaptation explains the sensitization and release from masking that has been observed psychophysically following notched noise.^{19,20} The model makes minimal assumptions about the neural processing required in the gain adaptation mechanism. It assumes that intensity is encoded separately for each frequency band, presumably in neuronal firing rates of a group of neurons, and

that the overall loudness of the signal is encoded separately from the intensity of each individual band. Finally, it assumes that signal power can be accumulated over some time frame and that this estimate can be used to reduce or inhibit the activity in each band. Most of these assumptions are compatible with the present knowledge of neuronal function.

It has not been necessary above to specify at which level of neural processing the gain adaptation mechanism may be operating. In fact, several stages of adaptation may be possible. For example, the mechanism of gain control could be operating as part of the control of outer hair cell response through medial olivocochlear efferent feedback.^{27,73} In this context it is interesting to note that it has been reported that the efferent inhibition of outer hair-cell function as evidenced by distortion products is impaired in most tinnitus subjects.⁴ A central mechanism is also consistent with the finding that unilateral cochlear implants generally reduce contralateral tinnitus.⁷⁴ Gain could also be adjusted through inhibition and/or excitation of primary afferent nerve fibers through lateral olivocochlear efferents.^{75,76} A recent review suggests that most current electrophysiology on tinnitus implicates stages upstream of the auditory nerve.¹⁰ Similarly, the only current reports on physiological correlates of the Zwicker tone come from the auditory cortex.^{77,78} Most notably, gain adaptation has been demonstrated for inferior colliculus neurons³⁰ and is mediated by inhibitory input.⁷⁹

Section II E points out that after gain normalization, the response in separate frequency bands does not distinguish long-term silence from persistent uniform noise. It is argued that for efficient information transmission, overall loudness is better transmitted as a separate variable which can then be used to disambiguate silence from uniform noise. The questions that begs answering is: *what is the neuronal substrate for such a representation?* There are many cells in the auditory cortex with high spontaneous activity which respond only transiently with an increase in firing rate to the onset of sound. Few cortical cells respond tonically to a steady stimulus. A distinct representation between loudness and modulation may therefore not seem unreasonable for the auditory cortex. The situation for the auditory nerve is less straightforward. On the surface it would seem that loudness is encoded in the overall firing rate. Yet in fact, an increase in firing rate does not necessarily reflect an increase in loudness⁸⁰ and firing rate is not a sufficient model to explain level discrimination.⁸¹ Instead, other mechanisms such as synchrony and phase relations across fibers may be required to explain psychophysical performance. It is also conceivable that outer hair cell afferent fibers, which are just now being characterized,⁸² serve a role in this regard.⁸³ Despite much effort, the details of how overall level is encoded in the auditory nerve remains an open question.

Note that the basic mechanism proposed here works separately for each frequency band, and is therefore uniform across frequencies. In contrast, some reports on the Zwicker tone suggest that the phenomenon is asymmetric, in that subjects tend to match the perception with a tone that is somewhat above the lower edge of the notched band.^{11,13} In fact, a high-pass band edge may not elicit a Zwicker tone. In addition, a sharp edge is required to elicit the Zwicker tone. Simi-

larly, tinnitus is associated with heightened sensitivity at the edge of high-frequency hearing loss¹⁰ and requires that the slope of hearing loss is high.¹ These phenomena have been explained by asymmetric lateral inhibition for the Zwicker tone^{51,84} or a loss of lateral inhibition from a damaged band for tinnitus.¹⁰ Similarly, when asymmetric lateral suppression is included in the present model, the percept is more pronounced at a lower frequency edge and nonexistent for a sloping edge. However, one might argue that these effects are not a prerequisite to explain the phantom precept. Instead, this manuscript focuses on a simple explanation for a basic mechanism so that robust experimental predictions could be derived.

C. Logarithmic scale of intensity level

With the exception of Eq. (7), the model has been formulated in terms of the powers of the signal, i.e., directly in terms of SPL. Neuronal signals, however, typically scale with the logarithm of the SPL, which is therefore typically given in decibels. Within a narrow dynamic range, the conversion from linear SPL to a logarithmic scale occurs at the first stage of signal transduction, i.e., the hair cell.⁸⁵ Membrane potentials of hair cells⁸⁶ as well as the resulting firing rates in the auditory nerve⁸⁷ scale with the logarithm of signal power, provided the SPL is within the narrow dynamic range of the specific cell. A logarithmic response over a large dynamic range—as required to explain observed perceptual sensitivity (Weber's law)—is likely to involve a combination of cells with different thresholds and slopes.⁸⁸ To avoid making specific assumptions as to how this is accomplished by the auditory system, the gain adaptation model is formulated in the original domain of linear SPL. Only lateral inhibition was formulated in the log domain, reflecting the fact that it operates at the level of the auditory nerve and further upstream.

In a logarithmic perceptual scale a multiplicative factor H becomes an additive contribution, and the power factor β becomes a multiplicative factor. Gain adaptation, as formulated here, corresponds to an adaptation to the mean power of the signal by a horizontal shift in the conventional rate-response function. A loss in compression (increasing slope β) necessitates a correction of the slope, which has not been modeled above, or a fast adaptation of the threshold of the rate-response function. This is the hypothesized mechanism leading to a larger susceptibility to the Zwicker tone for tinnitus subjects.

The present work has analogs to a model for tinnitus developed concurrently,⁸⁹ which separately considers damage to inner hair cells (IHC), outer hair cells (OHC), and stereocilia. That model operates on the logarithm of the signal power and assumes that IHC damage is multiplicative on this logarithmic scale, that stereocilia damage is additive thus reducing background noise, and that OHC damage is additive therefore not affecting background noise. In the model described here, these three forms of loss correspond respectively to a loss in compression (power factor β), a multiplicative loss (the factor H) with a simultaneous reduction of spontaneous rate (the noise N), and a pure multiplicative loss

(the factor H). The edge effects discussed above are modeled differently in their work, however, the most significant difference between the models is the criterion that leads to the phantom percept: the alternative model⁸⁹ require two levels of adaptation, namely optimal information transmission at the level of the AN followed by homeostatic adaptation to a desired firing rate. The present model, in contrast, only requires optimal information transmission at some stage of processing.

D. Prediction

The mechanism proposed here predicts that gain adaptation should vary across frequencies for a given subject depending on the strength of the nonlinear compression at each frequency band. This work therefore predicts a link between sensitization following a notched noise^{19,20} and the various correlates that are commonly associated with the nonlinear effects of the cochlear amplifier, such as distortion products or two-tone suppression—both of which can be measured psychophysically or audiometrically using otoacoustic emissions.^{17,47,91,90}

VI. CONCLUSIONS

Two main conclusions follow from this work. The first is that a rather simple model of optimal auditory adaptation can account for tinnitus as a consequence of a mismatch between the design parameters of the adaptive system and the actual performance of the sensory apparatus. This account does not speak directly as to *where* in the brain tinnitus arises, but rather as to *why*. The model makes a novel testable prediction concerning a linkage between Zwicker tone and tinnitus, which was experimentally confirmed. It also predicts a link between the Zwicker tone and cochlear amplification, which will be tested in future experiments. The second conclusion is that, regardless of the motivation for the psychophysical experiment conducted, the experimental result in and of itself constitute a novel empirical link between tinnitus and Zwicker tone.

¹O. König, R. Schaette, R. Kemper, and M. Gross, "Course of hearing loss and occurrence of tinnitus," *Hear. Res.*, **221**, 59–64 (2006).

²J. C. Cooper, Jr., "Health and nutrition examination survey of 1971–75: Part II. tinnitus, subjective hearing loss, and well-being," *J. Am. Acad. Audiol.* **5**, 37–43 (1994).

³V. Vesterager, "Tinnitus: Investigation and management," *Br. Med. J.* **314**, 728–731 (1997).

⁴G. Hesse, H. Schaaf, and A. Laubert, "Specific findings in distortion product otoacoustic emissions and growth functions with chronic tinnitus," *Audiol. Neuro-Otol.* **11**, 6–13 (2005).

⁵A. Norena, C. Micheyl, S. Chery-Croze, and L. Collet, "Psychoacoustic characterization of the tinnitus spectrum: Implications for the underlying mechanisms of tinnitus," *Audiol. Neuro-Otol.* **7**, 358–369 (2002).

⁶*Tinnitus Handbook*, edited by R. S. Tyler (Thomson, Delmar Learning, 2000).

⁷D. M. Baguley, "Mechanisms of tinnitus," *Br. Med. Bull.* **63**, 195–212 (2002).

⁸A. R. Møller, *Hearing: Its Physiology and Pathophysiology* (Academic, New York, 2000).

⁹J. J. Eggermont, "Psychological mechanisms and neural models," in *Tinnitus Handbook*, edited by R. S. Tyler (Thomson, Delmar Learning, 2000), pp. 85–122.

¹⁰J. J. Eggermont and L. E. Roberts, "The neuroscience of tinnitus," *Trends Neurosci.* **27**, 676–682 (2004).

¹¹A. Norena, C. Micheyl, and S. Chery-Croze, "An auditory negative afterimage as a human model of tinnitus," *Hear. Res.* **149**, 24–32 (2000).

¹²E. Zwicker, "'Negative afterimage' in Hearing," *J. Acoust. Soc. Am.* **36**, 2413–2415 (1964).

¹³E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models* (Springer, New York, 1999).

¹⁴H. Fastl, "On the Zwicker-tone of line spectra with a spectral gap," *Acustica* **67**, 177–186 (1989).

¹⁵R. C. Lummis and N. Guttman, "Exploratory studies of Zwicker's 'negative afterimage' in Hearing," *J. Acoust. Soc. Am.* **51**, 1930–1944 (1972).

¹⁶F. Julicher, D. Andor, and T. Duke, "Physical basis of two-tone interference in hearing," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9080–9085 (2001).

¹⁷R. V. Shannon and T. Houtgast, "Psychophysical measurements relating suppression and combination tones," *J. Acoust. Soc. Am.* **68**, 825–829 (1980).

¹⁸R. Stoop and A. Kern, "Two-tone suppression and combination tone generation as computations performed by the hopf cochlea," *Phys. Rev. Lett.* **93**, 268103 (2001).

¹⁹R. P. Carlyon, "Changes in the masked thresholds of brief tones produced by prior bursts of noise," *Hear. Res.* **41**, 223–235 (1989).

²⁰L. Wiegand, M. Koessl, and S. Schmidt, "Auditory sensitization during the perception of acoustical negative afterimages: Analogies to visual processing?," *Naturwiss.* **82**, 387–389 (1995).

²¹J. J. Atick and A. N. Redlich, "Quantitative tests of a theory of retinal processing: Contrast sensitivity curves," Technical Report NYU-NN-90/2, New York University, 1990, also Institute for Advanced Studies IASSNS-HEP-90/51.

²²F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1996).

²³H. B. Barlow, R. Fitzhugh, and S. W. Kuffler, "Change of organization in the receptive fields of the cats retina during dark adaptation," *J. Physiol. (London)* **137**, 338–354 (1957).

²⁴I. Ohzawa, G. Sclar, and R. D. Freeman, "Contrast gain control in the cat visual cortex," *Nature (London)* **289**, 266–268 (1982).

²⁵R. M. Shapley and J. D. Victor, "The contrast gain control of the cat retina," *Vision Res.* **19**, 431–434 (1979).

²⁶Y. Yu, B. Potetz, and T. S. Lee, "The role of spiking nonlinearity in contrast gain control and information transmission," *Vision Res.* **45**, 583–592 (2005).

²⁷J. J. Guinan, Jr., "Physiology of olivocochlear efferents," in *The Cochlea* (Springer, New York, 1996), pp. 435–502.

²⁸R. L. Smith and J. J. Zwillocki, "Short-term adaptation and incremental responses of single auditory-nerve fibers," *Biol. Cybern.* **17**, 169–182 (1975).

²⁹L. A. Westerman and R. L. Smith, "Rapid and short-term adaptation in auditory nerve responses," *Hear. Res.* **15**, 249–260 (1984).

³⁰I. Dean, N. S. Harper, and D. McAlpine, "Neural population coding of sound level adapts to stimulus statistics," *Nat. Neurosci.* **8**, 1684–1689 (2005).

³¹N. J. Ingham and D. McAlpine, "Spike-frequency adaptation in the inferior colliculus," *J. Neurophysiol.* **91**, 632–645 (2004).

³²*Physiology of the Ear*, 2nd ed., edited by A. F. Jahn and J. R. Santos-Sacchi (Singular, 2001).

³³B. J. C. Moore, *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, New York, 2003).

³⁴I. Nelken, Y. Rotman, and O. Bar Yosef, "Responses of auditory-cortex neurons to structural features of natural sounds," *Nature (London)* **397**, 154–157 (1999).

³⁵R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science* **270**, 303–304 (1995).

³⁶L. C. Parra, C. D. Spence, and P. Sajda, "Higher-order statistical properties arising from the non-stationarity of natural signals," *Advances in Neural Information Processing Systems* **13**, 786–792 (2001).

³⁷J. R. Cavanaugh, W. Bair, and J. A. Movshon, "Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons," *J. Neurophysiol.* **88**, 2547–2556 (2002).

³⁸J. Malo, I. Epifanio, R. Navarro, and E. P. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Trans. Image Process.* **15**, 68–80 (2006).

³⁹R. Valerio and R. Navarro, "Optimal coding through divisive normalization models of V1 neurons," *Network* **14**, 579–593 (2003).

⁴⁰R. Buccirossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*

- 8, 1688–1701 (1999).
- ⁴¹C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.* **27**, 379–423; **27**, 623–656 (1948).
- ⁴²Frequency decomposition generates approximately zero-mean normal-distributed data, in which case the optimal transfer function has the shape of a Gaussian CDF centered at zero, and optimal transmission is achieved by adjusting the variance of the input to the slope of the CDF. Alternatively, log-power may be communicated, which is also well approximated by a normal distribution. In this case, overall the signal gain adjusts the mean of the Gaussian so that the mean power matches the center of the CDF. In either case, adjusting the data to have unit variance is the first-order correction to maximize transmitted information.
- ⁴³This can be viewed as a trivial Kalman filter; of course, a more sophisticated estimator could be used instead.
- ⁴⁴M. LeMasurier and P. G. Gillespie, “Hair-cell mechanotransduction and cochlear amplification,” *Neuron* **48**, 403–415 (2005).
- ⁴⁵A. J. Hudspeth, Y. Choe, A. D. Mehta, and P. Martin, “Putting ion channels to work: Mechano-electrical transduction, adaptation, and amplification by hair cells,” *PNAS* **97**, 11765–11772 (2000).
- ⁴⁶V. M. Eguluz, M. Ospeck, Y. Choe, A. J. Hudspeth, and M. O. Magnasco, “Essential nonlinearities in hearing,” *Phys. Rev. Lett.* **84**, 5232–5235 (2000).
- ⁴⁷E. J. Williams and S. P. Bacon, “Compression estimates using behavioral and otoacoustic emission measures,” *Hear. Res.* **201**, 44–54 (2005).
- ⁴⁸T. Janssen, P. Kummer, and W. Arnold, “Growth behavior of the 2 f1-f2 distortion product otoacoustic emission in tinnitus,” *J. Acoust. Soc. Am.* **103**, 3418–3430 (1998).
- ⁴⁹R. V. Harrison, “Rate-versus-intensity functions and related AP responses in normal and pathological guinea pig and human cochleas,” *J. Acoust. Soc. Am.* **70**, 1036–1044 (1981).
- ⁵⁰R. S. Tyler and D. Conrad-Armes, “Masking of tinnitus compared to masking of pure tones,” *J. Speech Hear. Res.* **27**, 106–111 (1984).
- ⁵¹J. M. Franosch, R. Kempter, H. Fastl, and J. L. van Hemmen, “Zwicker tone illusion and noise reduction in the auditory system,” *Phys. Rev. Lett.* **90**, 178103 (2003).
- ⁵²G. M. Gerken, “Central tinnitus and lateral inhibition: An auditory brainstem model,” *Hear. Res.* **97**, 75–83 (1996).
- ⁵³W. S. Rhode and S. Greenberg, “Lateral suppression and inhibition in the cochlear nucleus of the cat,” *J. Neurophysiol.* **71**, 493–514 (1994).
- ⁵⁴U. W. Biebel and G. Langner, “Evidence for interactions across frequency channels in the inferior colliculus of awake chinchilla,” *Hear. Res.* **169**, 151–168 (2002).
- ⁵⁵M. L. Sutter, C. E. Schreiner, M. McLean, K. N. O’Connor, and W. C. Loftus, “Organization of inhibitory frequency receptive fields in cat primary auditory cortex,” *J. Neurophysiol.* **82**, 2358–2371 (1999).
- ⁵⁶R. 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), Chap. 3, pp. 123–177.
- ⁵⁷M. Slaney, “An efficient implementation of the Patterson–Holdsworth cochlear filter bank,” Technical Report 45, Apple, Cupertino, CA, 1993. Auditory Toolbox version 2.0.
- ⁵⁸This subjective criterion was used because currently there is no objective test available for tinnitus.
- ⁵⁹Y. Weiss, E. P. Simoncelli, and E. H. Adelson, “Motion illusions as optimal percepts,” *Nat. Neurosci.* **5**, 598–604 (2002).
- ⁶⁰M. G. Heinz and E. D. Young, “Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss,” *J. Neurophysiol.* **91**, 784–795 (2004).
- ⁶¹M. J. Penner, “Magnitude estimation and the ‘paradoxical’ loudness of tinnitus,” *J. Speech Hear. Res.* **29**, 407–412 (1986).
- ⁶²H. Gouveris, J. Maurer, and W. Mann, “DPOAE-grams in patients with acute tonal tinnitus,” *Otolaryngol.-Head Neck Surg.* **132**, 550–553 (2005).
- ⁶³E. T. Onishi, Y. Fukuda, and F. A. Suzuki, “Distortion product otoacoustic emissions in tinnitus patients,” *Int Tinnitus J* **10**, 13–16 (2004).
- ⁶⁴Y. Shiomi, J. Tsuji, Y. Naito, N. Fujiki, and N. Yamamoto, “Characteristics of DPOAE audiogram in tinnitus patients,” *Hear. Res.* **108**, 83–88 (1997).
- ⁶⁵C. R. Mitchell and T. A. Creedon, “Psychophysical tuning curves in subjects with tinnitus suggest outer hair cell lesions,” *Otolaryngol.-Head Neck Surg.* **113**, 223–233 (1995).
- ⁶⁶M. J. Penner, “Two-tone forward masking patterns and tinnitus,” *J. Speech Hear. Res.* **23**, 779–786 (1980).
- ⁶⁷A. M. Terry, D. M. Jones, B. R. Davis, and R. Slater, “Parametric studies of tinnitus masking and residual inhibition,” *Br. J. Audiol.* **17**, 245–256 (1983).
- ⁶⁸J. Vernon, “Attempts to relieve tinnitus,” *J. Am. Aud. Soc.* **2**, 124–131 (1977).
- ⁶⁹J. Vernon and M. B. Meikle, “Tinnitus masking,” in *Tinnitus Handbook*, edited by R. S. Tyler (Thomson, Delmar Learning, 2000), pp. 313–355.
- ⁷⁰B. A. Goldstein, A. Shulman, M. L. Lenhardt, D. G. Richards, A. G. Madsen, and R. Guinta, “Long-term inhibition of tinnitus by UltraQuiet therapy: Preliminary report,” *Int Tinnitus J* **7**, 122–127 (2001).
- ⁷¹A. J. Norena and J. J. Eggermont, “Enriched acoustic environment after noise trauma reduces hearing loss and prevents cortical map reorganization,” *J. Neurosci.* **25**, 699–705 (2005).
- ⁷²C. Formby, L. P. Sherlock, and S. L. Gold, “Adaptive plasticity of loudness induced by chronic attenuation and enhancement of the acoustic background,” *J. Acoust. Soc. Am.* **114**, 55–58 (2003).
- ⁷³J. J. Guinan, Jr. and M. L. Gifford, “Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers I: Rate-level functions,” *Hear. Res.* **33**, 97–113 (1988).
- ⁷⁴N. Quaranta, S. Wagstaff, and D. M. Baguley, “Tinnitus and cochlear implantation,” *Int J Audiol.* **43**, 245–251 (2004).
- ⁷⁵J. A. Groff and M. C. Liberman, “Modulation of cochlear afferent response by the lateral olivocochlear system: Activation via electrical stimulation of the inferior colliculus,” *J. Neurophysiol.* **90**, 3178–3200 (2003).
- ⁷⁶W. H. Mulders and D. Robertson, “Diverse responses of single auditory afferent fibres to electrical stimulation of the inferior colliculus in guinea-pig,” *Exp. Brain Res.* **160**, 235–244 (2005).
- ⁷⁷E. S. Hoke, B. Ross, and M. Hoke, “Auditory afterimage: Tonotopic representation in the auditory cortex,” *NeuroReport* **9**, 3065–3068 (1998).
- ⁷⁸A. J. Norena and J. J. Eggermont, “Neural correlates of an auditory afterimage in primary auditory cortex,” *J. Assoc. Res. Otolaryngol.* **4**, 312–328 (2003).
- ⁷⁹N. J. Ingham and D. McAlpine, “GABAergic inhibition controls neural gain in inferior colliculus neurons sensitive to interaural time differences,” *J. Neurosci.* **25**, 6187–6198 (2005).
- ⁸⁰E. M. Relkin and J. R. Doucet, “Is loudness simply proportional to the auditory nerve spike count?,” *J. Acoust. Soc. Am.* **101**, 2735–2740 (1997).
- ⁸¹H. S. Colburn, L. H. Carney, and M. G. Heinz, “Quantifying the information in auditory-nerve responses for level discrimination,” *J. Assoc. Res. Otolaryngol.* **4**, 294–311 (2003).
- ⁸²M. A. Reid, J. Flores-Otero, and R. L. Davis, “Firing patterns of type II spiral ganglion neurons *in vitro*,” *J. Neurosci.* **24**, 733–742 (2004).
- ⁸³Their sparseness, convergent connectivity, slower responses, reduced accommodation, and association with the cochlear amplifiers make the outer hair cell afferent fibers ideal candidates to encode loudness.
- ⁸⁴A. Kral and V. Majernik, “On lateral inhibition in the auditory system,” *Gen. Physiol. Biophys.* **15**, 109–127 (1996).
- ⁸⁵C. J. Sumner, E. A. Lopez-Poveda, L. P. O’Mard, and R. Meddis, “A revised model of the inner-hair cell and auditory-nerve complex,” *J. Acoust. Soc. Am.* **111**, 2178–2188 (2002).
- ⁸⁶P. Dallos, “Response characteristics of mammalian cochlear hair cells,” *J. Neurosci.* **5**, 1591–1608 (1985).
- ⁸⁷B. Sachs, Murray and P. J. Abbas, “Rate versus level function for auditory-nerve fibers in cat: Tone-burst stimuli,” *J. Acoust. Soc. Am.* **56**, 1835–1847 (1974).
- ⁸⁸M. Winter, Ian R. Palmer, and Alan, “Intensity coding in low-frequency auditory-nerve fibers of the guinea pig,” *J. Acoust. Soc. Am.* **90**, 1958–1967 (1991).
- ⁸⁹R. Schaette and R. Kempter, “Development of tinnitus-related neuronal hyperactivity through homeostatic plasticity after hearing loss: A computational model,” *Eur. J. Neurosci.* **23**, 3124–3138 (2006).
- ⁹⁰A. J. Oxenham and C. J. Plack, “A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,” *J. Acoust. Soc. Am.* **101**, 3666–3675 (1997).
- ⁹¹C. J. Plack and A. J. Oxenham, “Basilar-membrane nonlinearity estimated by pulsation threshold,” *J. Acoust. Soc. Am.* **107**, 501–507 (2000).

Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration^{a)}

Julie Arenberg Bierer^{b)}

Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Box 354875, Seattle, Washington 98105-6246

(Received 15 May 2006; revised 14 December 2006; accepted 4 January 2007)

The efficacy of cochlear implants is limited by spatial and temporal interactions among channels. This study explores the spatially restricted tripolar electrode configuration and compares it to bipolar and monopolar stimulation. Measures of threshold and channel interaction were obtained from nine subjects implanted with the Clarion HiFocus-I electrode array. Stimuli were biphasic pulses delivered at 1020 pulses/s. Threshold increased from monopolar to bipolar to tripolar stimulation and was most variable across channels with the tripolar configuration. Channel interaction, quantified by the shift in threshold between single- and two-channel stimulation, occurred for all three configurations but was largest for the monopolar and simultaneous conditions. The threshold shifts with simultaneous tripolar stimulation were slightly smaller than with bipolar and were not as strongly affected by the timing of the two channel stimulation as was monopolar. The subjects' performances on clinical speech tests were correlated with channel-to-channel variability in tripolar threshold, such that greater variability was related to poorer performance. The data suggest that tripolar channels with high thresholds may reveal cochlear regions of low neuron survival or poor electrode placement. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436712]

PACS number(s): 43.66.Ts, 43.66.Cb [AJO]

Pages: 1642–1653

I. INTRODUCTION

The electrode configuration of a cochlear implant determines the spatial extent of cochlear activation. Computer and physical models of electrical current flow (Spelman *et al.*, 1995; Jolly *et al.*, 1996; Kral *et al.*, 1998; Briare and Frjins, 2000) have demonstrated the broadest spatial extent of electrical fields with the monopolar (MP) configuration, and progressively more restricted fields with bipolar (BP) and then tripolar (TP) (also known as quadrupolar) configurations. Accordingly, stimulation with the MP configuration should activate a larger population of neurons, and that neural activation pattern should become more focused with BP and TP stimulation. Electrophysiology studies in animals examining the effects of electrode configuration are consistent with these predictions. In the auditory nerve, individual fibers were more sharply tuned with BP and TP compared to MP (Kral *et al.*, 1998). A similar trend was observed across populations of auditory nerve fibers (van den Honert and Stypulkowski, 1987), and, more centrally, along the tonotopic axis of the inferior colliculus and auditory cortex (Rebscher *et al.*, 2001; Bierer and Middlebrooks, 2002; Raggio

and Schreiner, 2003; Snyder *et al.*, 2004). Importantly, the cortical activation patterns elicited by a single TP channel were most similar to that of an acoustic pure tone, while the broader patterns elicited by a MP channel more closely resembled activation by broadband noise (Arenberg *et al.*, 2000; Bierer and Middlebrooks, 2002).

Finer cochleotopic selectivity with the TP configuration was also revealed in cortical responses to two-channel stimulation (Bierer and Middlebrooks, 2004). In that study, the decrease in threshold caused by the addition of a second subthreshold channel was defined as *channel interaction*, an indication that the two channels were not activating the auditory system independently. Channel interaction was smaller for the BP and TP modes than for the MP mode, suggesting that restricted electrode configurations were activating more restricted neural populations. In addition, the greatest interaction occurred with simultaneous stimulation and decreased as the timing between the two channels was increased, suggesting that a temporal offset of two pulses can significantly reduce interaction. CI speech processing strategies in use today implement such interleaved pulses. Overall, the lower thresholds observed with two-channel stimulation were consistent with a facilitative mechanism of current summation in the cochlea, possibly mediated by summation of depolarization of auditory neurons (Cartee *et al.*, 2000; Middlebrooks, 2004).

Psychophysical measures of channel interaction in human cochlear implant (CI) subjects have also been performed using near-threshold stimulation (Eddington *et al.*, 1978; Favre and Pelizzone, 1993; Boëx *et al.*, 2003; deBalthasar *et al.*, 2003; Stickney *et al.*, 2006). Consistent with the physiological study by Bierer and Middlebrooks (2004), a reduc-

^{a)}Portions of this work were presented in: "Effects of electrode configuration on psychophysical measures of channel interaction" at the 54th Annual Society for Neuroscience Meeting, New Orleans, LA, November, 2003; "Effects of electrode configuration on psychophysical measures of channel interaction in cochlear implant subjects" at the 2003 Conference on Implantable Auditory Prostheses, Pacific Grove, CA, August, 2003; "The tripolar electrode configuration and measures of channel interaction in cochlear implant subjects", at the 28th Annual Association for Research in Otolaryngology Meeting, in New Orleans, LA, February 2005; and "The tripolar configuration reduces channel interaction" at the 2005 Conference on Implantable Auditory Prostheses, Pacific Grove, CA, August, 2005.

^{b)}Electronic mail: jbiierer@u.washington.edu

tion in threshold occurred when stimuli were presented to two channels simultaneously and the magnitude of shift decreased as the timing between channels increased. Moreover, a smaller degree of channel interaction was often obtained with the narrower BP electrode configuration than the MP configuration (White *et al.*, 1984; Boëx *et al.*, 2003; Stickney *et al.*, 2006). Those previous studies, however, did not examine the TP configuration. Given the physiological evidence that this type of stimulation elicits an even more focused activation pattern than BP, the use of TP might yield a consistently lower degree of channel interaction.

Psychophysical studies have demonstrated that dynamic range is influenced by electrode configuration. Larger dynamic ranges, which can provide better intensity resolution (e.g., Zeng *et al.*, 1998; Chatterjee, 1999) and improved speech perception (Zeng and Galvin, 1999; Fu and Shannon, 2000), were measured in response to progressively more focused BP electrode configurations (Chatterjee, 1999). A recent study by Drennan and Pfungst (2005) demonstrated that in subjects with a shorter duration of deafness, better intensity discrimination was obtained for the BP compared to MP electrode configuration. Further enhancement in dynamic range might be possible with the TP configuration.

Despite the potentially better selectivity, lower degree of channel interaction, and larger dynamic range with TP stimulation, there may be several disadvantages to this configuration. Perhaps the most prohibitive is that the TP mode will likely require even greater current levels to reach threshold and comfortable listening levels than the BP mode (e.g., human: Drennan and Pfungst, 2005; Mens and Berenstein, 2005; guinea pig: Bierer and Middlebrooks, 2002; Snyder *et al.*, 2004; cat: Rebscher *et al.*, 2001). Indeed, in human listener's maximum comfort levels cannot always be reached with the BP configuration (Pfungst and Xu, 2004; Drennan and Pfungst, 2005). High threshold and comfort levels could significantly decrease the life of the batteries that power an implant and could limit the growth of loudness. Another disadvantage of focused electrode configurations is the much larger channel-to-channel variability in comfort and threshold levels as compared to monopolar configuration, which has been observed both psychophysically (Pfungst and Xu, 2004) and physiologically (Rebscher *et al.*, 2001; Bierer and Middlebrooks, 2002; Snyder *et al.*, 2004). Variability can make clinical programming of CI settings more time consuming and difficult. Yet, it is unclear if psychophysical channel-to-channel variability would be greater with the TP configuration than has been reported with the BP configuration.

The goal of the present study is to evaluate single- and two-channel responses using the tripolar electrode configuration. Measures of single-channel threshold, channel-to-channel variability, and dynamic range are compared across electrode configurations (TP, BP, and MP). Measures of two-channel thresholds while varying the relative timing of pulses on the two channels are compared across electrode configurations and timing conditions (simultaneous and interleaved). Three timing conditions are studied: simultaneous, temporally adjacent, and symmetrically interleaved. The simultaneous condition is studied because it appears to

provide the most direct evidence of current summation at or before the neural membrane. Two nonsimultaneous conditions are studied because the symmetrically interleaved condition provides the greatest temporal separation between successive pulses, while the temporally adjacent condition provides the least. The temporally adjacent condition is most similar to what occurs in current speech processing strategies for multichannel cochlear implants. A reduction in threshold for two-channel versus single-channel stimulation is defined as channel interaction. By varying the degree of spatial overlap with electrode configuration and temporal overlap with relative timing, the underlying mechanisms of channel interaction can be better understood.

General methods: (1) *Subjects.* Nine adult cochlear implant users participated in this study. All subjects were implanted with the Clarion HiFocus-I electrode arrays with CII electronics (Advanced Bionics Corp., Sylmar, CA). Center-to-center electrode distance is 1.1 mm for that array. Subjects ranged in age from 30 to 84 years (2 males and 7 females) and had a wide range of speech perception abilities. Individual subject details are shown in Table I. All participants were native speakers of American English who became deaf after the acquisition of language. These experiments were done in accordance with guidelines set by the Committee for Human Research at the University of California at San Francisco and the Human Subjects Division at the University of Washington.

With the exception of one subject, all speech perception data were taken from standard clinical visits occurring within three months of the onset of the experiment. Subject S11's speech perception testing was completed approximately one year prior to the experiment. Consonant-Nucleus Consonant (CNC) word tests (Minimum Speech Test Battery for Adult Cochlear Implant Users, House Ear Institute and Cochlear Corporation, 1996) were performed in a double-walled sound attenuating room. Words were presented at a level of 70 dB SPL on an A-weighted scale.

(2) *Stimuli.* Pulses were biphasic and charge-balanced (cathodic phase first on the active electrode) with 102 μ s per phase presented at 1020 pulses/s. Stimuli consisted of pulse trains that were 510 ms in duration. The pulse trains for two-channel stimuli are shown in Fig. 1(b) and will be discussed in Sec. III A. A "channel" refers to an active electrode and the return electrode(s) that complete the current loop. The electrode configuration was varied from monopolar (MP) to bipolar (BP) to tripolar (TP) [see Fig. 1(a)], but was fixed within each run. The MP configuration consisted of an intrascalar active electrode and a distant ground return electrode. The BP electrode configuration consisted of an intrascalar active electrode with an adjacent intracochlear return electrode. We refer to the more basal of the pair as the active electrode. Finally, the TP electrode configuration consisted of an intrascalar active electrode and two flanking return electrodes, each carrying half of the return current. The channel number from 1 through 16 was defined by the location of the active electrode, from apex to base.

Stimuli were delivered to the subjects using a clinical interface controlled by the Bionic Ear Data Collection System version 1.15.158 software (BEDCS, Advanced Bionics

TABLE I. Subject information.^a

Subject number	Sex	Age (years)	Age of severe hearing loss (years)	Age of implantation (years)	Etiology of hearing loss	HINT sentences in quiet	CNC words
S3	F	48	43	47	Meniere's disease and/or autoimmune	42 97 ^b	26
S4	F	85	40	82	Unknown	CNT ^c	CNT ^c
S5	F	75	45	73	Unknown	81	26
S6	F	32	13	31	Cochlear dysplasia	9 44 ^b	4
S7	F	71	56	69	Unknown	97	76
S8	M	67	56	64	Meniere's disease	14	2
S9	F	63	40	62	Familial	90	32
S10	F	51	7	50	Usher's syndrome	92	48
S11	M	74	51	73	Industrial injury	97	60

^aThis subject was reimplanted following a device failure. The previous implant use was 6 years.

^bThese results are with the CI and a hearing aid in the nonimplant ear.

^cCould not test (S4 does not have enough open-set speech to perform the tests).

Corp., Sylmar, CA). All data represent stimulus levels supported by the implant, i.e. levels that would generate voltages on the electrodes below the implant compliance limit of 8 V. The BEDCS software was run on a personal computer and interfaced through the clinical interface with a dedicated Platinum Series Processor. The same computer was used for

on-line analyses (MATLAB, Mathworks, Natick, MA). A second PC-based computer allowed the subjects to respond to the stimuli using a mouse. Trial-by-trial feedback was not provided.

II. EXPERIMENT 1: ELECTRODE CONFIGURATION AND SINGLE-CHANNEL STIMULATION

A. Methods

Stimuli and procedures. All thresholds were measured using a 3-down 1-up, two-interval, forced-choice paradigm, converging at the 79.4%-correct level (Levitt, 1971). The amplitude of the pulse train was adaptively adjusted until a minimum of eight reversals had been completed. The first three reversals were reached using 2 dB steps between levels, and the remaining five or more reversals using 0.5 dB steps. Two or more repetitions of each measurement were obtained for S9, S10, and S11, and for some measurements in S3, S5, and S8. If the values obtained on the two runs differed by more than 1 dB, a third run was collected and all three runs were averaged. If the third run differed from the first two runs by more than 1 dB, a fourth run was performed and all four runs were averaged. In all other cases a single repetition was obtained. Different numbers of repetitions were obtained because of a change in the protocol part way through the study to increase the statistical power of the measurements. Data were analyzed separately for those subjects for whom at least two measurements were made and those for whom only one measurement was made.

The maximum comfortable level (MCL) was determined by initially presenting the stimulus at a low, suprathreshold level and allowing the subject to adjust the level by using the mouse to click a box labeled either "up" or "down." Subjects were instructed to set the level to be "Loud but still comfortable." The level changed in 1 dB steps until the first reversal in stimulus direction and thereafter changed in 0.5 dB steps. For subjects S9, S10, and S11, MCL was measured at least twice and averaged. For the other subjects a single measure-

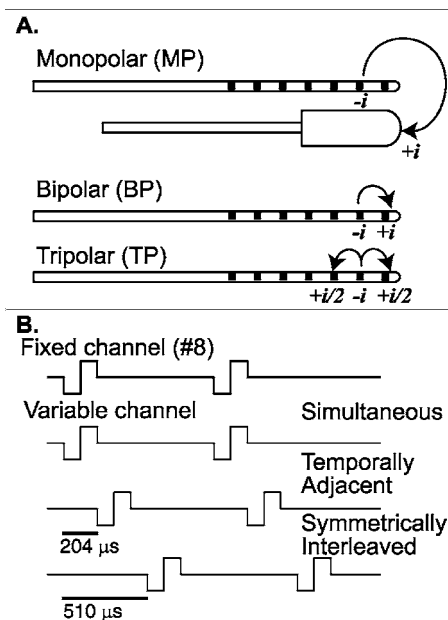


FIG. 1. Electrode configurations and pulse train stimuli for one and two channels. (a) Schematic of electrode configurations. The current path from active to return electrode(s) is represented by the arrows. The magnitude of current at the active electrode, $-i$, indicates that pulses were delivered with the cathodic phase first. For the monopolar configuration (top) the return electrode is outside the cochlea in the temporal bone. For the bipolar configuration (middle) the active and return electrodes are neighboring electrodes with the return electrode as the most apical of the pair. For the tripolar configuration (bottom) there are two return electrodes on either side of the active electrode, each carrying half of the return current. (b) Schematic of two-channel stimulation. The top line represents the fixed-channel pulse train, always set to channel 8. The subsequent lines illustrate the variable-channel pulse train for the three timing conditions: simultaneous ($0 \mu\text{s}$ offset between pulse onsets), temporally adjacent ($204 \mu\text{s}$ offset), and symmetrically interleaved ($510 \mu\text{s}$ offset).

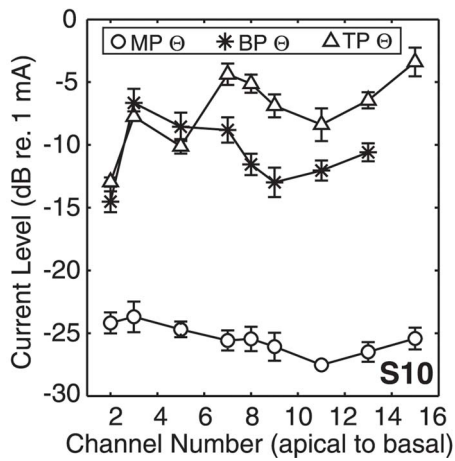


FIG. 2. Single-channel thresholds for the MP, BP, and TP configurations. Data are from one subject, S10. The abscissa represents CI channel from apical to basal while the ordinate represents current level in decibels relative to 1 mA. Thresholds were greatest for the TP configuration.

ment was obtained for MCL. A similar technique to that used with threshold measurements was used for determining the number of runs collected for MCL.

MCL could not be measured for some of the BP and most of the TP channels because of the limited compliance voltage of the equipment. The maximum allowable current for an active electrode, the compliance limit, was calculated from the clinically measured impedance adjusted to take into account phase duration (Advanced Bionics, personal communication). The compliance limit was defined as the maximum voltage allowed by the implant system (8 V) divided by the impedance.

B. Results

Threshold varied across electrode configuration (MP, BP, and TP) and active channel number (apical to basal position). Figure 2 displays these measures obtained for one subject, S10, for all three configurations as a function of channel number. In this subject, MP thresholds were much lower than those for TP and BP modes for all channels. For the TP configuration, the middle and basal channels had higher thresholds than the apical channels.

The threshold data for the six subjects for whom multiple runs were collected for each data point are displayed in Fig. 3, with each panel representing data for one subject. The panels are ordered based on the subjects' clinical performances on CNC word identification tasks, with the best performer in the top left panel and decreasing performance from left to right and top to bottom. Error bars represent plus and minus one standard deviation of the mean of the last eight reversals from all runs. Error bars are not shown when the bars were smaller than the symbols representing the means. Across all subjects and CI channels, threshold generally increased as the configuration changed from MP to BP ($p < 0.005$, paired sign test) and from MP to TP ($p < 0.005$), and from BP to TP ($p < 0.05$).

As noted earlier, thresholds in Fig. 3 for the MP configuration were relatively constant across CI channels. With BP and, especially, TP configurations, threshold levels were

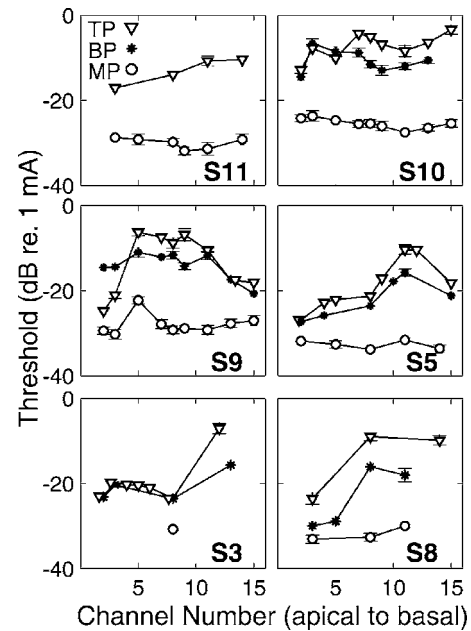


FIG. 3. Single-channel thresholds across configurations. Each panel represents the single-channel detection thresholds for one subject. The panels are sorted from highest to lowest scores (top left to bottom right) on CNC words in a quiet setting. The subject number is indicated in the bottom right corner. Conventions are as in Fig. 2. Thresholds increased from the MP to the BP and further to the TP mode. Data are only presented for subjects for whom at least two repetitions were obtained.

much more variable. A measure of variability was calculated for each subject as the standard deviation of the unsigned difference between thresholds of adjacent tested channels. This measure of variability is a variation on the calculation used by Pfungst and colleagues (Pfungst and Xu, 2004), which quantified the mean of the unsigned difference between thresholds of neighboring channels. Both measures are included in this study: The mean emphasizes the absolute magnitude of channel-to-channel threshold differences for each subject, while the standard deviation takes into account the expected channel-to-channel differences, emphasizing local rather than global variability. Only subjects with at least four channels of threshold measurements for each configuration were included in these calculations. Figures 4(a) and 4(b) display the distribution of channel-to-channel variability for five of these subjects who met the criteria and had multiple measures for each threshold data point. Data for the other subjects are plotted in Figs. 4(c) and 4(d). The left column data are based on the standard deviation variability measure and the right column data are based on the unsigned mean measure. In Fig. 4(a), four of the five subjects had greater channel-to-channel variability with either the BP or TP configuration compared to the MP configuration, and in three of those four subjects the TP variability was greater than the BP variability. In Fig. 4(b), either the BP or the TP configuration produced the largest variability for all of the subjects, and in four of the five subjects the TP variability was greater. The Wilcoxon signed-rank test, a nonparametric paired comparison, demonstrated that a significantly greater channel-to-channel variability occurred for TP than MP ($p < 0.1$) for both the standard deviation and mean variability measures. No significance was found for the BP and MP or TP and BP

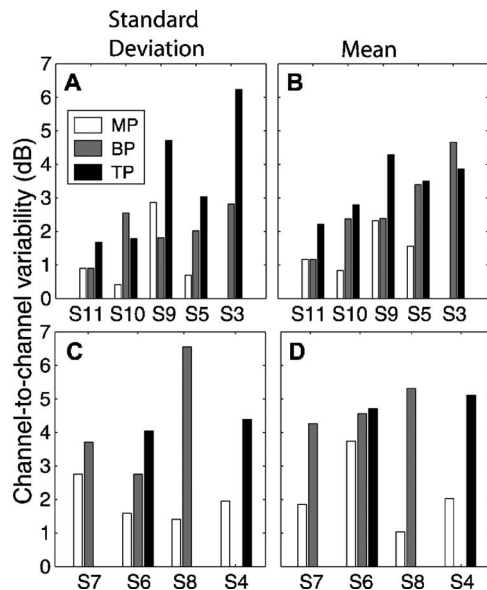


FIG. 4. Distribution of channel-to-channel variability. The histograms show the channel-to-channel variability in decibels for each subject (shaded bars) for the MP, BP, and TP configurations. The left and right columns represent channel-to-channel variability measured as the standard deviation and the mean of the threshold differences for neighboring tested channels, respectively. In the top row (a) and (b) data are presented for subjects for whom at least two repetitions were obtained for each threshold measurement. In the bottom row (c) and (d) data are presented for subjects for whom only one repetition was obtained for each threshold measurement. Channel-to-channel variability for all subjects was greater for the BP or TP mode.

comparisons. Similar trends are observed in the data presented in Figs. 4(c) and 4(d). When comparisons were made for all available subjects, greater channel-to-channel variability occurred for TP versus MP ($p < 0.05$, for six subjects) and TP versus BP ($p < 0.05$, for six subjects) for both measures of variability.

As exemplified by the trends for subject S10 (Fig. 2), threshold was dependent on electrode configuration and place of stimulation in the apical to basal dimension. Figure 5 represents a paired comparison of thresholds obtained for apical, middle, and basal channels across subjects. Thresholds from each region of the array were averaged within a subject. Apical channels were defined as channels between 1 and 5, middle channels between 6 and 10, and basal channels between 11 and 16. The top panel compares basal and apical thresholds, the middle panel compares middle and apical thresholds, and the lower panel compares basal and middle thresholds. Each data point is the average threshold for one subject and the symbols represent electrode configuration. Data points above the diagonal line indicate that thresholds were higher for the more basal channel of the pair. For the TP configuration, basal channels exhibited significantly higher thresholds than apical channels (TP apical versus basal, $p < 0.05$, Wilcoxon signed-rank test; TP apical versus middle, $p = 0.06$; all other comparisons showed no significant differences). As the BP and MP thresholds were somewhat constant regardless of channel number, no significant trends along the basal to apical dimension were observed with these configurations. This finding is consistent with physiological data that showed higher TP thresholds at progressively more basal channels (Bierer and Middlebrooks, 2002).

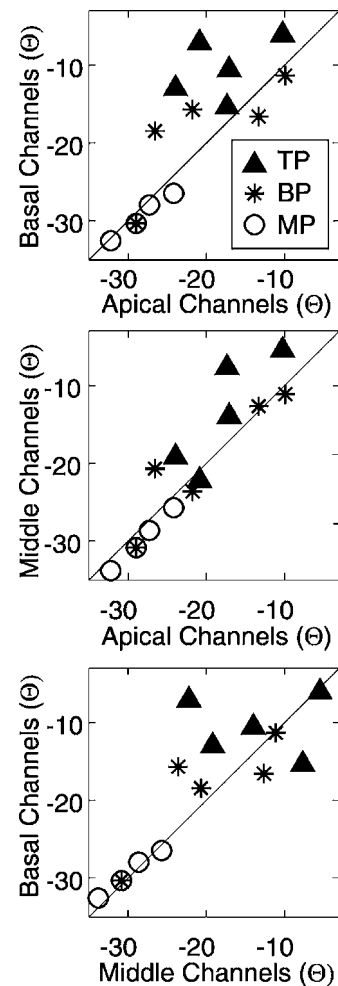


FIG. 5. Basal compared to apical thresholds. Each panel represents a scatter plot comparing thresholds for apical to basal (top), apical to middle (middle), and basal to middle channels (bottom). Channels 1–5=apical; channels 6–10=middle; channels 11–16=basal. The ordinate and abscissa are threshold in decibels relative to 1 mA. For each subject there is a data point for each tested electrode configuration, represented by symbols. Data are plotted only for subjects for whom at least two measurements for each threshold were obtained. All available data points are averaged within a subject for each segment of the electrode array (basal, middle, and apical). With TP, the thresholds were higher for basal than apical channels and higher for middle compared to apical channels. With MP and BP, the thresholds were independent of apical to basal location.

Figure 6 demonstrates the contributions of threshold and MCL to dynamic range. The data are from subjects for whom at least two measurements were made. Data for the MP configuration is shown in the top panel. Each solid vertical line represents the dynamic range for one subject and one channel. The lines extend from threshold to MCL (ordinate) as a function of dynamic range (abscissa). The regression lines represent the best fit lines to MCL (top of each line) and threshold (bottom of each line). This analysis shows that dynamic range increased primarily as a consequence of increasing MCL with a slope of 0.91 dB MCL/dB dynamic range ($r = 0.65$, $p < 0.001$, Pearson correlation coefficient, transformed t-statistic); threshold decreased with a slope of only -0.09 dB threshold/dB dynamic range ($r = -0.09$, $p = 0.66$).

For some BP and most TP channels, a MCL could not be obtained because the compliance limits were reached prior to

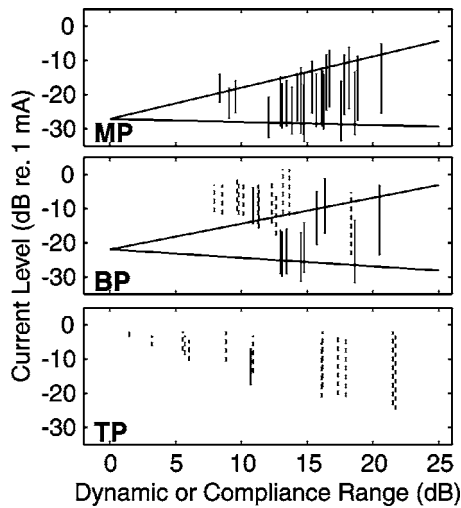


FIG. 6. Distribution of thresholds and comfort levels contributing to dynamic range. The top, middle, and bottom panels represent MP, BP, and TP data, respectively. The ordinate is the current level from threshold (the bottom of each vertical line) to either MCL or compliance limit (the top of each line) in decibels relative to 1 mA. Solid lines represent the range from threshold to MCL while dashed lines represent the range from threshold to the compliance limit for that electrode. The abscissa is dynamic or compliance range in decibels. With increasing dynamic range, for the MP mode the MCL increased significantly while the threshold was relatively constant.

the subjects' perception of comfortable loudness. In these cases, for which the true dynamic range could not be estimated, a *compliance range* is defined as the difference between threshold and the estimated compliance limit of each active electrode (see Sec. II A). These data appear as dashed lines. For the BP configuration (middle panel), both threshold and MCL were quite variable across dynamic ranges. Regression analyses performed only for true dynamic range measurements show that MCL increased with a slope of 0.75 dB MCL/dB dynamic range ($r=0.33$, $p=0.35$) and threshold decreased with a slope of -0.25 dB threshold/dB dynamic range ($r=-0.11$, $p=0.75$). Neither of these measurements was significant.

For the TP configuration (bottom panel), compliance limits were reached for most CI channels. Therefore the data are plotted as a function of compliance range, as described earlier, and no regression analysis was performed. Note that even TP channels with low thresholds could not reach MCL. Because the compliance range underestimates dynamic range, this finding suggests that these TP channels could have dynamic ranges larger than any of the MP channels tested. This panel also emphasizes the range and variability of TP thresholds. However, the variability of MCL, and thus, the true dynamic range could not be assessed.

In summary, single-channel stimulation using the TP configuration resulted in high perceptual thresholds compared to BP and MP. Progressively larger channel-to-channel variability in threshold was measured from MP to BP to TP. Additional variability for the TP configuration was apparent as higher basal channel thresholds compared to apical thresholds. Unfortunately, due to compliance limitations, conclusions about the relative contributions of threshold and MCL to BP and TP dynamic ranges could not be determined. Finally, dynamic range for the MP configuration was strongly

dependent on MCL. Together these results suggest that the TP configuration can potentially provide more focused spatial activation in the cochlea. In Experiment II, the comparison of TP, BP, and MP stimulation is extended to interactions between two channels.

III. EXPERIMENT II: ELECTRODE CONFIGURATION AND CHANNEL INTERACTION

A. Methods

1. Stimuli and procedures

Two-channel stimuli consisted of two different active electrodes and their respective return electrodes. Both channels were always stimulated using the same configuration, MP, BP, or TP. The two channels consisted of a fixed channel in the middle of the implant array, and a variable channel that could be any of the other available channels. In the simultaneous mode, allowable electrodes for the variable channel were those that did not contribute as an active or return electrode for the fixed channel. For example, with the TP configuration, channel 8 consisted of electrode 8 as active and electrodes 7 and 9 as returns; therefore, the nearest channels for simultaneous stimulation were 5 (active) and 4 and 6 (return), or 11 (active) and 10 and 12 (return). The variable channel was varied in both channel location and timing relative to the fixed channel. As shown in Fig. 1(b), two-channel stimuli were presented either simultaneously or with interleaved biphasic pulses. The temporal offset was defined by the timing between the onset of the first pulse on one channel and the onset of the first pulse on the second channel. Three timing conditions were used: simultaneous with a $0 \mu\text{s}$ offset, temporally adjacent with a $204 \mu\text{s}$ offset, and symmetrically interleaved with a $510 \mu\text{s}$ offset.

Detection thresholds were determined as for single-channel stimulation. The current levels on the two channels were always varied together, increasing or decreasing in equal step sizes (2 or 0.5 dB) relative to each channel's threshold when presented alone (as determined in Experiment I). Channel interaction was quantified by the threshold shift, which is the difference between the single-channel threshold and the two-channel threshold of the variable channel. If the threshold for the two-channel stimulus was lower, then the shift was a negative number indicating a decrease in current level. A larger negative threshold shift was interpreted to indicate greater channel interaction.

The measure of threshold shift to quantify channel interaction is a modification of methods used previously (e.g., Eddington *et al.*, 1978; Stickney *et al.*, 2006). In those studies, two-channel thresholds were compared for in-phase and out-of-phase pulses, such that the difference between thresholds serves as a measure of direct current summation. That measure allows for separation of channel interaction resulting from direct current summation, which should exhibit a phase effect, and channel interaction from overlapping neural excitation, which should not have a phase effect. The interpretation of the in-phase facilitative threshold shifts obtained in the present study assumes, based on previous findings (deBalthasar *et al.*, 2003), that most of the interaction for simultaneous near-threshold signals is a result of current summa-

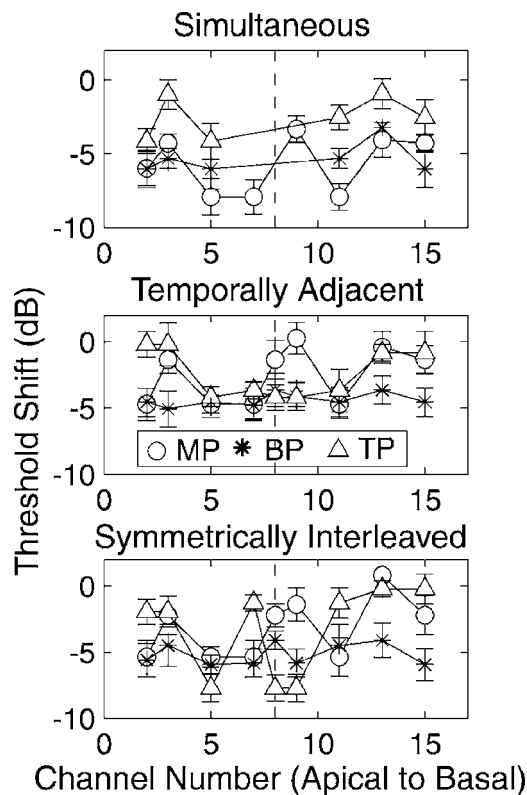


FIG. 7. Channel interaction for one subject. Threshold shift, measured as the difference between two-channel and single-channel thresholds of the variable channel, is given in decibels (ordinate). The abscissa is active channel number from apical to basal. The top, middle, and bottom panels represent simultaneous, temporally adjacent, and symmetrically interleaved conditions, respectively. The symbols represent electrode configuration. Error bars represent plus or minus one standard deviation from the mean. For the MP configuration, greater channel interaction was observed for the simultaneous condition compared with the two nonsimultaneous conditions.

tion, particularly with the MP configuration. The use of narrower electrode configurations and interleaved pulses allows for assessing channel interaction from other sources.

2. Subjects

For Experiment II, subjects S9, S10, and S11 were tested with at least two repetitions for each measurement. Subjects S3 through S8 were tested with only one repetition for each measurement. These two groups of subjects were analyzed separately. As with Experiment I, different numbers of repetitions were obtained in subjects as a result of a change in protocol.

B. Results

Channel interaction, as measured by the shift in threshold between single- and two-channel stimuli, is described across electrode configurations and the timing of two-channel stimulation. Figure 7 displays the threshold shifts for one subject, S9. The symbols represent the electrode configuration and the rows correspond to the three different two-channel timing conditions. Shifts in threshold at each channel tended to be largest (i.e., most negative) for the simultaneous MP condition. Threshold shifts for all other conditions were generally smaller and more variable across

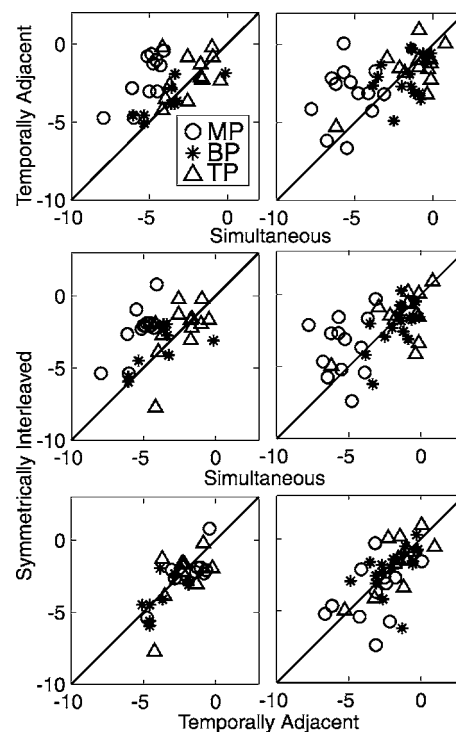


FIG. 8. Paired comparisons of channel interaction across two-channel timing conditions. The top row compares threshold shift in decibels for simultaneous (abscissa) vs temporally adjacent (ordinate) conditions. The middle row compares simultaneous (abscissa) vs symmetrically interleaved (ordinate) conditions. The bottom row compares temporally adjacent (abscissa) vs symmetrically interleaved (ordinate) conditions. Different symbols represent electrode configuration. The left column represents data from subjects S9, S10, and S11 and the right column represents data from subjects, S3, S4, S5, S6, S7, and S8. Larger negative threshold shifts were observed for the MP compared to the BP or TP modes, especially in the simultaneous condition.

channels and configurations. In the simultaneous condition (top), a marked difference between MP and TP threshold shifts occurred for most tested CI channels. A smaller difference occurred between the MP and BP configurations. For the two nonsimultaneous conditions (middle and lower panels, respectively), the BP and TP conditions sometimes showed a larger threshold shift than the MP condition. This result was not typical and the example likely reflects the variability in the data.

The threshold shift measurements pooled across S9, S10, and S11 are represented in the left column of Fig. 8 and measurements from subjects S3 through S8 (subjects with one repetition) are represented in the right column. Each data point corresponds to a single CI channel and each symbol represents a different electrode configuration. Only data from CI channels that were tested in all three configurations are plotted in Fig. 8. The top row compares simultaneous and temporally adjacent timing conditions. The middle row compares symmetrically interleaved and simultaneous conditions and the bottom row compares symmetrically interleaved and temporally adjacent conditions. Consistent with the example shown in Fig. 7, in the top two rows the majority of MP symbols (open circles) are above the diagonal line indicating a greater negative threshold shift for the simultaneous condition compared with either of the nonsimultaneous condi-

tions (paired t-test, $p < 0.001$, top two rows). For the BP and TP configurations, the magnitude of simultaneous threshold shifts was typically smaller than with MP, as seen by the progression from left to right for the MP, BP, and TP configurations, respectively. Also, threshold shifts were more similar across the two-channel timing conditions for BP and TP configurations (paired t-test, $p > 0.05$). The greater simultaneous threshold shift with the MP configuration was highly significant and suggests that by restricting the size of the current field (using TP or BP configurations), simultaneous interactions were greatly reduced.

Paired comparisons of simultaneous threshold shift between configurations demonstrate that for subjects S9, S10, and S11 (left column) MP threshold shifts were greater than the BP and TP configurations ($p < 0.05$, $p < 0.001$, t-test, respectively). Further, in the simultaneous condition, BP threshold shifts were larger than TP shifts ($p < 0.001$). In the temporally adjacent condition, the broader electrode configuration yielded a larger threshold shift than the narrow configurations for each comparison, such that MP was larger than BP ($p < 0.01$), MP was larger than TP ($p < 0.01$) and BP was larger than TP ($p < 0.01$). In the symmetrically interleaved condition, there was no significant difference in the threshold shift magnitude between any two electrode configurations ($p > 0.5$, for all comparisons). For subjects S3 through S8, similar results were obtained for the simultaneous condition (right column). However, no difference was observed between MP and BP or BP and TP configurations in the temporally adjacent condition and in the symmetrically interleaved condition, MP elicited a significantly greater threshold shift than BP and TP configurations ($p < 0.05$). Overall the three tested electrode configurations exhibited smaller threshold shifts for the nonsimultaneous conditions, consistent with previous studies (Shannon 1983; deBalthasar *et al.*, 2003). This finding suggests that summation of depolarization of auditory neurons is the main mechanism of the channel interactions measured (Cartee *et al.*, 2000; Middlebrooks, 2004).

In summary, two-channel simultaneous stimulation resulted in the largest threshold shifts for the MP configuration and became progressively smaller as the configuration was changed to BP and TP. For all configurations threshold shift was reduced for the two interleaved conditions compared to simultaneous stimulation.

C. Comparisons to speech perception scores

The results from Experiment I demonstrate that single-channel thresholds are generally higher with the BP and, even more so, with the TP configuration. Further, results from Experiment II demonstrate that channel interaction is smaller for the TP configuration. Thus, the potential benefits of reduced channel interaction with the TP configuration could be offset by its higher threshold and greater channel-to-channel variability. An important question, then, is how these measures relate to subject performance on speech perception tests. The relationship between the single-channel variability in threshold and each subject's clinical performance on CNC word identification tests are presented in Fig.

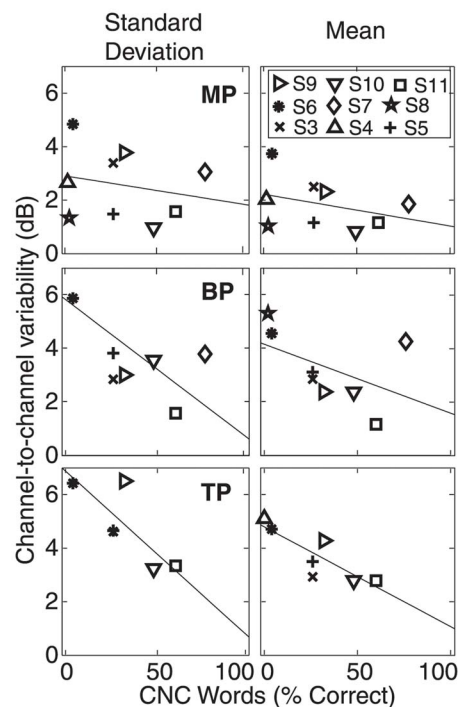


FIG. 9. Channel-to-channel variability in single-channel threshold as a function of speech perception performance on CNC word identification. The ordinate is channel-to-channel variability in decibels and the abscissa is subject performance on CNC words in percent correct. The left and right columns depict the standard deviation and the mean channel-to-channel variability measures, respectively. The symbols represent individual subjects. Data are presented for MP (top), BP (middle), and TP (bottom) configurations. TP threshold variability is negatively correlated with CNC word identification ($R = -0.82$; $p < 0.05$) while the BP and MP data are not correlated with word scores ($R = -0.69$; $p = 0.058$, $R = -0.14$; $p > 0.05$, respectively).

9. The data in the left column are based on the standard deviation of threshold differences across channels, while the data in the right column are based on the mean of threshold differences. With the MP configuration (top row), channel-to-channel variability was uniformly small with no correlation with CNC word scores (left panel, $r = -0.22$, $p = 0.6$, right panel, $r = -0.35$, $p = 0.36$). But the two narrower electrode configurations, BP (middle) and TP (bottom), exhibit trends such that subjects with smaller channel-to-channel variability performed better on CNC word tests. However, only the TP configuration exhibited a statistically significant negative correlation at the 95% confidence interval (TP: left panel, $r = -0.85$, $p = 0.01$, right panel, $r = -0.85$, $p = 0.01$; BP: left panel, $r = -0.69$, $p = 0.06$, right panel, $r = -0.5$, $p = 0.21$). Also, no correlations were found between speech perception scores and either single-channel thresholds or two-channel threshold shifts ($p > 0.05$). These preliminary results suggest that channel-to-channel variability in TP threshold, a broad measure of CI effectiveness in activating the auditory system, can predict speech perception ability.

IV. DISCUSSION

Previous studies have made comparisons between the MP and BP configurations (e.g., deBalthasar *et al.*, 2003; Pfungst *et al.*, 2004; Drennan and Pfungst, 2005). This study expands our present understanding of the perceptual influ-

ence of focused electrode configurations by including the TP configuration. The TP electrode configuration, even more so than the BP mode, was more variable in threshold and exhibited reduced channel interaction, compared with the broader MP mode.

A. Single-channel stimulation

Previous psychophysical studies have reported higher thresholds for BP compared to MP electrode configurations (Shannon, 1983; Chatterjee, 1999; Pfingst *et al.*, 2004; Pfingst and Xu, 2004). The present finding that TP thresholds were, in turn, higher than BP thresholds (Fig. 3) is supported by physiological studies at several levels of the auditory pathway showing higher thresholds for progressively narrower patterns of neural activation (Shepherd *et al.*, 1993; Bierer and Middlebrooks, 2002; Snyder *et al.*, 2004). These findings are in accordance with volume conduction models of the electrode/cochlea interface (Jolly *et al.*, 1996; Kral *et al.*, 1998; Briare and Frjins, 2000), which predict that TP stimulation yields the narrowest electrical fields and highest thresholds. Ultimately, at least two aspects of a restricted field are likely to affect the detection of a stimulus. First, volume conduction models have shown that reaching a criterion voltage level at a fixed distance from an electrode requires a higher level of current for TP than BP or MP configurations (Jolly *et al.*, 1996). Second, the activation of a minimum number of spiral ganglion neurons may be required for the implant listener to reach perceptual threshold (Frijns *et al.*, 1995; Bruce *et al.*, 1999), a greater hurdle for narrow electrical fields. Thus, the high thresholds obtained with the narrow TP electrode configuration likely reflect the activation of small populations of spiral ganglion neurons, in contrast to the broader swath of neurons activated by MP stimulation. This latter point is especially relevant if the neurons closest to a set of tripolar electrodes cannot be stimulated due to atrophy, ossification, or some other pathology.

Auditory neural survival is quite variable among subjects with profound hearing loss (e.g., Hinojosa and Marion, 1983). The focused activation elicited by narrow electrode configurations should be more susceptible to local neuron survival or electrode placement along the cochlea. Channels relatively far away from the spiral ganglion, or adjacent to a dead region with no surviving ganglion cells, would have higher thresholds for activation because of the longer current pathway to stimutable neurons. On the other hand, access to stimutable neurons is much less of an issue with broader stimulation. Indeed, previous studies comparing BP and MP stimulation have shown greater channel-to-channel variability in threshold and MCL for the BP configuration than for MP (Busby *et al.*, 1994; Pfingst *et al.*, 1997; Chatterjee, 1999; Pfingst and Xu, 2004). The current study extends those findings, demonstrating an even greater variability of threshold with TP stimulation.

The local channel-to-channel variations with TP stimulation were accompanied by a more global trend across the implant array, such that threshold decreased from base to apex (see Fig. 5). That this effect was not observed with the MP or BP configuration supports the hypothesis that the TP

mode is more sensitive to irregularities in nerve survival or electrode placement. The differences between basal and apical stimulation discussed earlier could have at least two explanations. First, nerve survival could have been better in the apex. For several pathologies, spiral ganglion cell degeneration tends to occur first in the base before progressing more apically (e.g., Schuknecht, 1974; Hinojosa and Marion, 1983; Zimmerman *et al.*, 1995). Second, electrodes may lie closer to the spiral ganglion at the implant's most apical extent. For instance, Cohen *et al.* (2003) compared psychophysical thresholds with x rays of the subjects' implant arrays and found that threshold varied with the distance of the electrode from the modiolus, where the spiral ganglion cells reside. All of the subjects who participated in the present study were implanted without the space-filling positioner designed to hold the HiFocus-I array near the modiolus, especially in the base. Thus, a larger distance from the spiral ganglion neurons would be expected at the base, resulting in higher required current levels for activation of those neurons. Further support for this notion comes from the work of Bierer and Middlebrooks (2002), who observed a gradient of neural thresholds in the auditory cortex of the guinea pig, with lower thresholds obtained for more apical stimulation. The authors attributed this gradient to the positioning of the cochlear implant (which was designed for the human cochlea): because the guinea pig cochlea tapers at the bend between the first and second turn, the implant likely had a slight buckle, pushing the apical electrodes more snugly against the modiolus wall.

In agreement with the present study, Mens and Berenstein (2005) recently found higher thresholds with a variant of the TP configuration (see Sec. IV C) compared to the MP configuration. The modified TP mode allowed them to measure MCL and dynamic range. In the present study, dynamic range could not be assessed with the true TP configuration because of the relatively long phase durations and the compliance limitations of the electrodes. However, the large compliance ranges (from threshold to compliance limits), particularly for TP channels with relatively low thresholds, suggests the potential for large dynamic ranges if the voltage limits were higher. This concept is consistent with the findings of Mens and Berenstein (2005), who demonstrated larger dynamic ranges with the modified TP configuration than with MP.

The potential for larger perceptual dynamic ranges with the TP configuration was suggested in a previous study of cortical activation patterns in the guinea pig (Bierer and Middlebrooks, 2002). In that study, dynamic range was defined as the difference between neural response threshold and saturation. The dynamic ranges were roughly equivalent across MP, BP, and TP configurations at the most sensitive cortical recording site. However, when cortical sites distant from the most sensitive site were examined, larger dynamic ranges were obtained with TP and progressively smaller dynamic ranges were obtained with BP and MP configurations. Further, the discrimination of stimulus channel based on cortical activation patterns persisted with increasing stimulus level with the TP configuration, more so than with MP or BP (Middlebrooks and Bierer, 2002). Overall, these physiologi-

cal results are consistent with a larger perceptual dynamic range for BP and TP stimulation, possibly represented as a distributed code in the auditory system (i.e., intensity coding by multiple neurons).

A large dynamic range is important for speech perception, both in normal and cochlear implant listening and especially in noisy listening environments (e.g., CI subjects—Zeng and Galvin, 1999; Fu and Shannon, 2000; normal hearing—Loizou *et al.*, 2000). Nevertheless, a measure of dynamic range is incomplete without a more thorough assessment of the growth of loudness. Determination of the number of perceptible loudness steps between threshold and MCL (e.g., Zeng *et al.*, 1998) or the study of loudness growth functions (Chatterjee, 1999) would provide a more complete understanding of cochlear implant function and the effects of electrode configuration. A recent study demonstrated better discrimination of current level with the BP compared to the MP electrode configuration for an apical channel in some subjects (Drennan and Pfingst, 2005). The authors speculated that this subgroup of subjects might have had better neural survival or electrode placement near the apical channel, allowing for better level discrimination with the BP configuration. Their finding is consistent with the present results in that both studies suggest a better neural survival or electrode placement for apical compared to basal channels and that such patterns of neural survival/placement are better detected with a focused electrode configuration such as BP or TP.

B. Two-channel stimulation

The large threshold shifts measured with the MP configuration support the hypothesis that the main contribution to simultaneous interaction is the summation of electrical fields in the cochlea. According to this hypothesis, if a population of auditory neurons receives overlapping current from two implant channels, then the amount of current from each channel necessary to produce a threshold perception should be lower than the amount required from either channel alone. Such a shift in current level was observed for nearly every pair of MP channels tested with the simultaneous condition (Fig. 7, top), similar to the results of previous studies (Shannon, 1983; deBalthasar *et al.*, 2003; Stickney *et al.*, 2001). The progressively smaller threshold shifts measured with BP and TP stimulation (Fig. 7, bottom) are also consistent with reduced field overlap. Finally, the reduction of MP threshold shift in the nonsimultaneous conditions is expected if the shifts result from the direct summation of current in the cochlea (Cartee *et al.*, 2000; deBalthasar *et al.*, 2003).

Two important results from studies of the guinea pig auditory cortex lend further support to the cochlear summation hypothesis. First, neural threshold shifts were much smaller with the BP and TP configurations than the MP configuration when two channels were stimulated simultaneously (Bierer and Middlebrooks, 2004). As in the present study, the MP threshold shifts decreased dramatically with nonsimultaneous stimulation. Second, with the MP configuration cortical recordings demonstrated substantial overlap of activated neural populations, and neural threshold shifts

could be modeled by simple addition of currents (Middlebrooks, 2004). In the present study, the degree of threshold shift observed in the simultaneous MP condition similarly is consistent with extensive overlap of excitation.

The existence of an interaction mechanism other than simple current summation is suggested by the threshold shifts that occurred for nonsimultaneous timing conditions. In normal hearing subjects, two equal-intensity clicks presented nonsimultaneously are detected at a lower threshold than a single click (Zwislocki, 1960; Viemeister and Wakefield, 1991). Similarly, longer duration stimuli are more easily detected than short stimuli (Gerken *et al.*, 1990). Both types of phenomena indicate that some type of temporal integration of stimulus energy occurs during auditory processing. Temporal integration was also demonstrated in CI listeners by Donaldson *et al.* (1997). In that study, the addition of pulses to a pulse train reduced the threshold for detection compared to that of a single pulse. In the present study, the addition of stimulus energy took the form of pulses from one channel interleaved with pulses from another channel. Temporal integration of each pair of pulses may thus have accounted for the lower thresholds, as in Donaldson *et al.*, assuming that such integration can occur across channels. Evidence from normal hearing listeners in which stimulus energy was combined across spectral bands to improve detection (Green, 1958) suggests that cross-channel integration is indeed possible. Further evidence in CI subjects is available from the loudness summation studies of McKay and colleagues (2001, 2003), in which it was demonstrated that the interleaved pulses from two channels were typically louder than the pulses on one channel, even when the two channels were far apart in the cochlea (and presumably did not have overlapping current fields). The actual mechanism of cross-channel integration is unknown, but it may be mediated by a "multiple looks" strategy (Viemeister and Wakefield, 1991), in which the auditory system makes independent evaluations of a stimulus over time and across tonotopic bands.

For the BP and TP configurations, it was observed that interpulse timing had little effect on threshold shift. This was also true for the nonsimultaneous MP timing conditions. One explanation, consistent with a temporal integration mechanism, is that the interpulse intervals used (ranging from 0 in the simultaneous condition to 510 μ s in the symmetrically interleaved condition) were much shorter than the integration time window of less than 10 ms for a single stimulus "look" inferred from acoustic studies (Viemeister and Wakefield, 1991). Nevertheless, a current summation mechanism for the nonsimultaneous condition cannot be ruled out. For instance, a small amount of residual depolarization, retained by the capacitive neural membrane of the cochlea, may linger following each leading pulse, adding to the charge of the subsequent pulse (Cartee *et al.*, 2000). On the other hand, the significantly larger MP threshold shifts in the simultaneous condition were presumably mainly due to current summation in the cochlea, though some temporal integration (and residual depolarization) may also have occurred.

For a given subject, channel-to-channel variability with the TP configuration was presumed to reflect the status of the

interface between the cochlear implant and surviving, stimu-
lable neurons. In fact, subjects with high TP channel-to-
channel variability in threshold exhibited relatively poorer
performance (Fig. 9). An underlying assumption is that the
high threshold channels are strongly contributing to the vari-
ability measure. These results suggest that the worst
channels—those with high thresholds—dominate the overall
performance of a subject. That implication is consistent with
previous studies relating single- and two-channel measures
to poor speech performance (Throckmorton and Collins,
1999; Pfungst *et al.*, 2004). In one of these studies, a similar
negative correlation between performance and variability in
threshold was found (Pfungst *et al.*, 2004). The correlation in
that study was greater for the MP compared to the BP mode,
whereas the present findings showed no correlation between
threshold variation and performance with the MP mode and a
slight but not statistically significant trend with the BP mode.
Several differences in methodology with the present study
may account for this discrepancy. One major difference is
that Pfungst and colleagues examined 22 channels in each
subject whereas between 4 and 14 channels were used in the
current study. It is possible that if all 14 available channels
were used, a greater variability in MP thresholds, and possi-
bly a greater correlation with performance, would have been
obtained. There were also differences in the CI devices [a
Nucleus Contour (Cochlear Corp., Sydney, Australia) versus
a Clarion HiFocus-I electrode] and the methods for obtaining
threshold (the method of adjustment versus two-interval
forced choice).

C. Clinical implications

There are several clinical implications of this study.
First, detection threshold was more variable with the TP than
with the BP and MP modes, and that variability was nega-
tively correlated with speech perception. Thus, TP thresholds
might be used as a diagnostic tool in the clinic to determine
suboptimal channels, whether the reason for the suboptimal-
ity is reduced nerve survival, poorly positioned electrode
channels, or some other factor. That information could be
obtained objectively and quickly using the electrically
evoked compound action potential or auditory brainstem re-
sponse. TP thresholds could also be used to program sub-
jects' speech processors to better stimulate individual chan-
nels. For instance, a region of poor interface between the
electrode and stimu-
lable neurons might be better stimulated
with a broad electrode configuration like MP or even by
inactivating channels in that region altogether. As this infor-
mation has not been readily available in the past, program-
ming possibilities have not yet been explored.

Second, the relationship between TP single-channel
threshold data and the proximity of individual electrodes to
the modiolus and surviving spiral ganglion neurons could be
explored. Three-dimensional reconstructions of computer to-
mography images of the CI *in vivo* could be performed in
conjunction with TP threshold measures. That information
could lead to speech processing strategies that are specific to
each patient's electrode placement and nerve survival pat-
terns.

Third, a new electrode configuration is under investiga-
tion referred to as partially tripolar or remote current frac-
tion. This configuration is similar to tripolar except that the
two return electrodes each carry less than half of the return
current while the excess current is sent to the monopolar
return electrode (Kral *et al.*, 1998). Recent studies with that
configuration in animal models (Bonhman *et al.* 2005) and
humans (Litvak *et al.*, 2005; Mens and Berenstein, 2005)
have demonstrated that it somewhat preserves restricted ac-
tivation patterns of true TP stimulation while lowering the
neural or perceptual thresholds. This new configuration may
provide the benefits of both TP and MP stimulation as a
recent study suggests for speech perception in fluctuating
background noise (Mens and Berenstein, 2005).

Finally, the minimal threshold shifts observed with the
TP configuration when two channels were presented simul-
taneously (at least for the 1020 pps trains used in this study)
have implications for clinical speech processing strategies. If
channel interaction remains small at suprathreshold stimulus
levels, one possibility is that multiple CI channels could be
stimulated simultaneously and potentially more temporal fine
structure could be presented to the CI listener.

ACKNOWLEDGMENTS

The author thanks Leonid Litvak and the Advanced
Bionics Corporation for technical support and for providing
the research and clinical interfaces. Great appreciation is ex-
tended to the implant subjects who participated in this study.
Thanks also to Amanda Seigfried for assistance in data col-
lection, Russell Snyder and Bob Shannon for consulting on
the study design and analyses, and John Middlebrooks,
Steven Bierer, and the reviewers for the Journal are thanked
for helpful comments. This work was supported by the Na-
tional Institutes of Health (NIDCD- F32 DC05883) and by
Hearing Research, Inc.

- Advanced Bionics Corp., Sylmar, CA, USA (personal communication).
Arenberg, J. G., Furukawa, S., and Middlebrooks, J. M. (2000). "Auditory
cortical images of tones and noisebands," *J. Assoc. Res. Otolaryngol.* **1**,
183–194.
Bierer, J. A., and Middlebrooks, J. M. (2002). "Auditory cortical images of
cochlear-implant stimuli: Dependence on electrode configuration," *J. Neu-
rophysiol.* **87**, 478–492.
Bierer, J. A., and Middlebrooks, J. M. (2004). "Auditory cortical images of
cochlear-prosthesis stimuli: Channel-interaction," *J. Assoc. Res. Otolaryn-
gol.* **5**, 32–48.
Boëx, C., de Balthasar, C., Kós, M., and Pelizzone, M. (2003). "Electrical
field interactions in different cochlear implant systems," *J. Acoust. Soc.
Am.* **114**, 2049–2057.
Bonham, B., Corbett, S., Johnson, T., and Rebscher, S. (2004). "Tripolar
focusing of auditory nerve activation: The effect of changing the remote
current fraction (RCF)," Neural Prosthesis Program Contract No. N01-
DC-02-1006, QPR#8.
Briere, J. J., and Frjins, J. H. M. (2000). "Field patterns in a 3D tapered
spiral model of the electrically stimulated cochlea," *Hear. Res.* **148**, 18–
30.
Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., and Clark, G. M.
(1999). "The effects of stochastic neural activity in a model predicting
intensity perception with cochlear implants: Low rate stimulation," *IEEE
Trans. Biomed. Eng.* **46**, 1393–1404.
Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark,
G. M. (1994). "Pitch perception for different modes of stimulation using
the cochlear multiple-electrode prosthesis," *J. Acoust. Soc. Am.* **95**, 2658–
2669.

- Cartee, L. A., van den Honert, C., Finley, C. C., and Miller, R. L. (2000). "Evaluation of a model of the cochlear neural membrane. I. Physiological measurement of membrane characteristics in response to intrameatal electrical stimulation," *Hear. Res.* **146**, 143–152.
- Chatterjee, M. (1999). "Effects of stimulation mode on threshold and loudness growth in multielectrode cochlear implants," *J. Acoust. Soc. Am.* **105**, 850–860.
- Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. C. (2003). "Spatial spread of neural excitation in cochlear implant recipients: Comparison of improved ECAP methods and psychophysical forward masking," *Hear. Res.* **179**, 72–87.
- de Balthasar, C., Boëx, C., Cosendai, G., Sigrist, V. A., and Pelizzone, M. (2003). "Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects," *Hear. Res.* **182**, 77–87.
- Donaldson, G. S., Viemeister, N. F., and Nelson, D. A. (1997). "Psychometric functions and temporal integration in electric hearing," *J. Acoust. Soc. Am.* **101**, 3706–3721.
- Drennan, W. R., and Pfungst, B. E. (2005). "Current-level discrimination using bipolar and monopolar electrode configurations in cochlear implants," *Hear. Res.* **202**, 170–179.
- Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (1978). "Auditory prostheses research with multiple channel intracochlear stimulation in man," *Ann. Otol. Rhinol. Laryngol.* **87**, 1–39.
- Favre, E., and Pelizzone, M. (1993). "Channel interactions in patients using the Ineraid multichannel cochlear implant," *Hear. Res.* **66**, 150–156.
- Frijns, J. H., de Snoo, S. L., and Schoonhoven, R. (1995). "Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea," *Hear. Res.* **87**, 170–186.
- Fu, Q. J., and Shannon, R. V. (2000). "Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users," *Ear Hear.* **21**, 227–235.
- Gerken, G. M., Bhat, V. K., and Hutchison-Clutter, M. (1990). "Auditory temporal integration and the power function model," *J. Acoust. Soc. Am.* **88**, 767–778.
- Green, D. M. (1958). "Detection of multiple component signals in noise," *J. Acoust. Soc. Am.* **30**, 904–911.
- Hinojosa, R., and Marion, M. (1983). "Histopathology of profound sensorineural deafness," *Forum Math.* **405**, 459–484.
- Jolly, C. N., Spelman, F. A., and Clopton, B. M. (1996). "Quadrupolar stimulation for cochlear prostheses: Modeling and experimental data," *IEEE Trans. Biomed. Eng.* **43**, 857–865.
- Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (1998). "Spatial resolution of cochlear implants: The electrical field and excitation of auditory afferents," *Hear. Res.* **121**, 11–28.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Litvak, L., Emadi, G., and Spahr, A. J. (2005). "Loudness growth with focused electrical stimulation: Model and cochlear implant data," *Assoc. Res. Otolaryngol. Abs.* 112, 28th Midwinter Meeting, New Orleans, LA, 19–24 February.
- Loizou, P. C., Dorman, M., and Fitzke, J. (2000). "The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants," *Ear Hear.* **21**, 25–31.
- McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (2003). "A practical method of predicting the loudness of complex electrical stimuli," *J. Acoust. Soc. Am.* **113**, 2054–2063.
- McKay, C. M., Remine, M. D., and McDermott, H. J. (2001). "Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level and mode of stimulation," *J. Acoust. Soc. Am.* **110**, 1514–1524.
- Mens, L. H., and Berenstein, C. K. (2005). "Speech perception with mono- and quadrupolar electrode configurations: A crossover study," *Otol. Neurotol.* **26**, 957–964.
- Middlebrooks, J. C. (2004). "Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds," *J. Acoust. Soc. Am.* **116**, 452–468.
- Middlebrooks, J. C., and Bierer, J. A. (2002). "Auditory cortical images of cochlear-implant stimuli: Coding of stimulus channel and current level," *J. Neurophysiol.* **87**, 493–507.
- Pfungst, B. E., and Xu, L. (2004). "Across-site variation in detection thresholds and maximum comfortable loudness levels for cochlear implants," *J. Assoc. Res. Otolaryngol.* **5**, 11–24.
- Pfungst, B. E., Xu, L., and Thompson, C. S. (2004). "Across-site threshold variation in cochlear implants: Relation to speech recognition," *Audiol. Neuro-Otol.* **9**, 341–352.
- Pfungst, B. E., Zwolan, T. A., and Holloway, L. A. (1997). "Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants," *Hear. Res.* **112**, 247–260.
- Raggio, M. W., and Schreiner, C. E. (2003). "Neuronal responses in cat primary auditory cortex to electrical cochlear stimulation. IV. Activation pattern for sinusoidal stimulation," *J. Neurophysiol.* **89**, 3190–3204.
- Rebscher, S. J., Snyder, R. L., and Leake, P. A. (2001). "The effect of electrode configuration and duration of deafness on threshold and selectivity of responses to intracochlear electrical stimulation," *J. Acoust. Soc. Am.* **109**, 2035–2048.
- Schuknecht, H. R. (1974). *Pathology of the Ear* (Harvard University Press, Cambridge).
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," *Hear. Res.* **12**, 1–16.
- Shepherd, R. K., Hatsushika, S., and Clark, G. M. (1993). "Electrical stimulation of the auditory nerve: The effect of electrode position on neural excitation," *Hear. Res.* **66**, 108–120.
- Snyder, R. L., Bierer, J. A., and Middlebrooks, J. C. (2004). "Patterns of inferior colliculus activity in response to acoustic and electric intracochlear electrical stimulation: Activation selectivity," *J. Assoc. Res. Otolaryngol.* **5**, 305–322.
- Spelman, F. A., Pfungst, B. E., Clopton, B. M., Jolly, C. N., and Rodenhiser, K. L. (1995). "Effects of electrical current configuration on potential fields in the electrically stimulated cochlea: Field models and measurements," *Ann. Otol. Rhinol. Laryngol. Suppl.* **166**, 131–136.
- Stickney, G. S., Louizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., and Opie, J. M. (2006). "Effects of electrode design and configuration on channel interactions," *Hear. Res.* **211**, 33–45.
- Throckmorton, C. S., and Collins, L. M. (1999). "Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects," *J. Acoust. Soc. Am.* **105**, 861–873.
- Van den Honert, C., and Stypulkowski, P. H. (1987). "Single fiber mapping of spatial excitation patterns in the electrically stimulated auditory nerve," *Hear. Res.* **29**, 195–206.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.
- White, M. W., Merzenich, M. M., and Gardi, J. N. (1984). "Multichannel cochlear implants: Channel interactions and processor design," *Arch. Otolaryngol.* **110**, 493–501.
- Zeng, F. G., Galvin, J. J., III, and Zhang, C. (1998). "Encoding loudness by electrical stimulation of the auditory nerve," *NeuroReport* **9**, 1845–1848.
- Zeng, F. G., and Galvin, J. J. III (1999). "Amplitude mapping and phoneme recognition in cochlear implant listeners," *Ear Hear.* **20**, 60–74.
- Zimmerman, C. E., Burgess, B. J., and Nadol, J. B. Jr. (1995). "Patterns of degeneration in the human cochlear nerve," *Hear. Res.* **90**, 192–201.
- Zwislocki, J. (1960). "Theory of temporal auditory summation," *J. Acoust. Soc. Am.* **32**, 1046–1060.

Quantifying the effects of fast-acting compression on the envelope of speech

Michael A. Stone^{a)} and Brian C. J. Moore

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 12 July 2006; revised 20 December 2006; accepted 20 December 2006)

Fast-acting compression has several effects on the envelope of speech signals, especially when a target and a background are mixed together. Three measures of the envelope are described, which can be used to quantify these changes: (1) Within-signal modulation correlation or coherence, the degree of correlation (or coherence) of the envelope (on a dB scale) of a single source across different frequency regions, which is reduced by fast-acting compression; (2) fidelity of envelope shape, the degree to which the envelope shape of the target speech in different frequency channels is preserved following compression; and (3) across-signal modulation correlation or coherence (ASMC), the extent to which the target and background acquire a common component of modulation when they are compressed together, which becomes greater in absolute value (more negative) when the target and background are compressed together. Values of these measures are presented and compared with intelligibility scores obtained using stimuli processed (with a noise-vocoder) so as to preserve mainly envelope cues in a limited number of frequency channels. The results suggest that the dominant factor affecting intelligibility is ASMC. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2434754]

PACS number(s): 43.66.Ts, 43.71.Gv, 43.66.Mk [DOS]

Pages: 1654–1664

I. INTRODUCTION

Fast-acting dynamic range compression is widely used in electroacoustic systems such as hearing aids and cochlear implants to reduce the range of signal levels at the output. The fast compression can increase the audibility of low-level signal components while preventing intense sounds from becoming uncomfortably loud (Villchur, 1973; Zeng, 2004). The increased audibility of low-level sounds sometimes leads to improved intelligibility of speech (Laurence *et al.*, 1983; Moore *et al.* 1992; Yund and Buckles, 1995), but does not always do so (De Gennaro *et al.*, 1986; Boothroyd *et al.*, 1988; Drullman and Smoorenburg, 1997). The lack of benefit found by some authors may have occurred because the benefits of improved audibility were offset by deleterious side effects of the fast-acting compression (Plomp, 1988). The existence of such deleterious effects is confirmed by the finding that, when speech is presented at a level at which it is clearly audible, and when the speech is processed so as to force listeners to rely mainly on envelope cues (Shannon *et al.*, 1995), then fast-acting compression tends to impair speech intelligibility relative to no compression or slow compression (Stone and Moore, 2003, 2004).

In this paper, we consider and quantify a variety of effects of fast-acting compression that may be considered as deleterious. We focus on the effects of compression on the envelope of speech, after the speech has been filtered into frequency channels; these filtered signals might be considered roughly to represent the outputs of the (broadened) auditory filters in an impaired auditory system, or the outputs of the analysis channels used in a hearing aid or cochlear

implant. The low-rate envelope fluctuations of speech in different frequency channels have been shown to be important for speech intelligibility (Plomp, 1983; Drullman *et al.*, 1994a, b), and this may be especially true when temporal fine structure information is not usable, as appears to be the case for people with moderate to severe cochlear hearing loss (Buss *et al.*, 2004; Moore *et al.*, 2006) and for users of cochlear implants (Moore and Carlyon, 2005). We do not explicitly consider the reduction of envelope modulation depth produced by fast-acting compression (although some of our measures are influenced by this), since the effects of this have been considered extensively elsewhere (Plomp, 1988; Stone and Moore, 1992, 2003, 2004). Nor do we consider the harmonic or intermodulation distortion in the audio-frequency domain produced by fast-acting compression, since that distortion can be low, and probably below the detection threshold, in well-designed compression systems (Stone *et al.*, 1999; Moore *et al.*, 2001; Tan *et al.*, 2004).

Perceptual segregation of two or more talkers may be especially difficult when envelope cues alone are available (Qin and Oxenham, 2003). Under these conditions, changes in the envelope of speech produced by fast-acting compression may markedly impair speech intelligibility (Stone and Moore, 2003, 2004). In what follows, we describe how compression produces three types of change in the envelope of a “target” talker in the presence of a “background” talker, and we quantify the magnitude of each type of change. We also compare the three types of envelope change with measures of the intelligibility of a target talker in a background talker, obtained under conditions where only envelope cues were available.

^{a)}Electronic mail: masl9@cam.ac.uk

II. CLASSIFICATION OF ENVELOPE CHANGES

In the following we describe three types of envelope change produced by fast-acting compression, and we pose a series of questions for each type of envelope change. We consider the effects of both single-channel and multi-channel compression.

A. Within-signal modulation correlation or coherence

If the speech of a single talker is filtered into frequency channels, and the envelope is extracted for each channel, then the envelope fluctuations are correlated to some extent in different frequency channels. The correlation tends to be higher for adjacent channels than for channels that are widely separated in frequency (Crouzet and Ainsworth, 2001; Healy and Bacon, 2002). The correlated fluctuations may help to bind the different frequency channels of the speech into a single perceptual stream, and when competing speech is present they may help to segregate the target speech from the background speech (Bregman *et al.*, 1985; Carrell and Opie, 1992; Carrell, 1993; Rappold *et al.*, 1993). We propose as a measure of this the within-signal modulation correlation or coherence (WSMC), which is defined as the degree of correlation (or coherence) of the envelope (on a decibel scale) of a single source across different frequency regions (i.e., after filtering into frequency channels).

We define a function $C(a_i, a_j)$ as an operator that extracts a scalar quantity from the envelopes of the i th channel of signal a , and the j th channel of signal a . This scalar quantity is either the Pearson correlation (r) or a summary statistic from the coherence function; more details are given in the following. We define the within-source modulation coherence or correlation (WSMC) as

$$\text{WSMC} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N C(a_i, a_j), \quad i \neq j, \quad (1)$$

where N is the number of channels. Effectively this quantity gives a measure of the similarity of the envelope of each channel with that of all the other channels, averaged across channels.

In this paper, we address two questions in connection with WSMC:

- (1) Does compression of a single source (target) affect its WSMC? To assess this, WSMC was compared for the original target ($\text{WSMC}_{\text{orig}}$) and the target after compression ($\text{WSMC}_{\text{comp}}$).
- (2) Does compression of a mixture of a target and a background affect the WSMC of the target? To assess this, the compression gain signal for the mixture was derived (separately for each channel in the case of multi-channel compression), and this gain signal was applied to the target alone. The WSMC of the resulting signal was compared to $\text{WSMC}_{\text{orig}}$ and $\text{WSMC}_{\text{comp}}$.

B. Fidelity of envelope shape

Fast-acting compression can markedly alter the envelope of speech from a single talker. The envelope may be dis-

torted in shape (Stone and Moore, 1992), and abrupt changes in envelope magnitude can lead to “overshoot” and “undershoot” effects (Verschuure *et al.*, 1996). The degree to which these effects occur depends on the design of the compressor, for example on whether the audio signal is delayed so as to time-align it with the gain-control signal (Robinson and Huntington, 1973). When compression is applied to a mixture of talkers, the envelope shape associated with the target talker will be altered, since the gain changes produced by the compressor are determined by the mixture rather than just by the target talker.

As a measure of these effects we propose the fidelity of envelope shape (FES), defined as the degree to which the envelope shape of the target in different frequency channels is preserved following compression. The FES can be measured by determining the correlation (or coherence) between the envelope of channel i for the original target (denoted a_i) and the envelope of channel i of the target after compression (denoted b_i), repeating this for all i , and averaging the results across i . This is expressed by

$$\text{FES} = \frac{1}{N} \sum_{i=1}^N C(a_i, b_i), \quad (2)$$

where $C(a_i, b_i)$ is an operator similar to that described in the previous section, except that now the correlation or coherence between two different signals (the target before and after compression) is determined.

In this paper, we address two questions in connection with FES:

- (1) To what extent does compression of a single source (the target) affect FES? This was assessed by calculating FES for the original target and for the target (alone) after compression.
- (2) To what extent does compression of a mixture of target and background affect the FES of the target? This was assessed by deriving the gain signal for the mixture (separately for each channel in the case of multi-channel compression), and applying this to the target alone. FES was then determined using the resulting signal as b and the original target as a .

C. Across-source modulation correlation or coherence (formerly called comodulation)

This form of envelope change was described by Stone and Moore (2003, 2004) as “comodulation.” Here, we use the term across-source modulation correlation or coherence (ASMC) to emphasize that this effect of compression is to introduce patterns of modulation that are partially correlated across previously independent sources, which may promote the perceptual fusion of the sources. ASMC arises when the variable gain produced by a compressor is applied to a mixture of speech signals with similar levels. Peaks in one signal cause a reduction in gain that is applied to all the signals in the mixture. Signals that were previously independently amplitude modulated, and therefore unlikely to fuse perceptually, acquire a common component of modulation from the gain control, and their independence is thereby reduced. The

reduced independence may lead to undesirable perceptual fusion of the target and background, and this may contribute to reduced intelligibility.

We propose as a definition of ASMC: the extent to which the target and background acquire a common component of modulation when they are compressed together. As a measure of this, we propose the following. For the mixture of target and background, the gain-control signal produced by the compressor is calculated (separately for each channel in the case of multi-channel compression). This gain-control signal is applied independently to the original target (giving $\text{Target}_{\text{postcomp}}$) and to the original background (giving $\text{Background}_{\text{postcomp}}$). The correlation (or coherence) is calculated between the envelope of channel i of $\text{Target}_{\text{postcomp}}$ and the envelope of channel i of $\text{Background}_{\text{postcomp}}$. This is repeated for all i and averaged across i to give the measure of ASMC. Mathematically, this is expressed as

$$\text{ASMC} = \frac{1}{N} \sum_{i=1}^N C(a_i, b_i), \quad (3)$$

where a_i and b_i are the envelopes of $\text{Target}_{\text{postcomp}}$ and $\text{Background}_{\text{postcomp}}$, respectively, in the i th channel.

Two questions are posed here:

- (1) What magnitude of ASMC is produced by fast-acting compression?
- (2) Does the magnitude of ASMC differ for single and multi-channel compression?

III. MEASURES OF ENVELOPE SIMILARITY

In the following we describe two measures of the similarity of envelope patterns across two channels, i.e., we describe the derivation of the operator $C(\cdot, \cdot)$ in Eqs. (1)–(3). We also justify calculating the measures based on the logarithms of the channel envelopes rather than on the linear amplitude values.

A. Correlation

The Pearson correlation coefficient is defined as

$$r_{ab} = \frac{\sum (a - \bar{a})(b - \bar{b})}{\sqrt{\sum (a - \bar{a})^2} \sqrt{\sum (b - \bar{b})^2}}, \quad (4)$$

where \bar{a} and \bar{b} denote the mean values of a and b , respectively. Here, the correlation was calculated for the entire duration of the stored envelopes, except for the first and last 0.5 s of the files. This was done to avoid possible effects associated with the start and finish of the files and of the processing. The resulting measure is denoted $C_{\text{cor}}(a, b)$.

B. Coherence

The magnitude squared coherence (MSC) function between two signals a and b is defined as

$$\text{MSC}(f) = \frac{|G_{ab}(f)|^2}{G_{aa}(f)G_{bb}(f)} \quad (5)$$

(Carter *et al.*, 1973), where $G_{xx}(f)$ is the average of the auto-spectrum of time signal x (related to the autocorrelation function) and $G_{ab}(f)$ is the average of the cross-spectra between signals a and b (related to the cross-correlation function). The short-term cross-spectrum, $S_{ab}(f)$, is defined as

$$S_{ab}(f) = a^*(f)b(f), \quad (6)$$

where $a^*(f)$ denotes the complex conjugate of the short-term spectrum of signal a and $b(f)$ denotes the short-term spectrum of signal b . The MSC is commonly calculated from the fast Fourier transform (FFT) of short segments of the time waveforms (the channel envelopes in our case) by averaging across many overlapping segments (Carter *et al.*, 1973). If a is the input to a system and b is the corresponding output, the MSC describes what fraction of the output power is linearly dependent on the input signal power. The single value that we require for our measure C can be obtained by averaging the MSC over the modulation frequency range of interest. Biases may be introduced into the estimate of the MSC unless certain conditions are observed when calculating the FFT (Carter *et al.*, 1973; Kates, 1992). We were careful to meet these conditions when performing our measurements.

The following sequence of steps was used to calculate the summary measure of coherence. The envelope was sampled at a rate of 1000 Hz (see later for justification of this sample rate). Segments of the envelope 4096 samples in length were windowed with a Kaiser window; the value of parameter β , which determines the relative level of “side-lobes” in the spectral domain, was set to 7. Compared to a Hanning window, the frequency resolution of a Kaiser window with $\beta=7$ has a slightly broader central lobe but smaller sidelobes. The MSC was calculated using 50% overlapped segments, starting at least 0.5 s after the start of the envelope signals and stopping at least 0.5 s before the end of the envelope signals, for the reason described earlier. The MSC was calculated by averaging in half-octave bands between 1 and 16 Hz. This effectively resulted in rejection of the DC component. This range of envelope modulation rates was chosen since the modulation spectrum of speech has its highest values in this range (Plomp, 1983) and because this range is important for speech intelligibility (Drullman *et al.*, 1994a, b; Stone and Moore, 2007). The MSC was collapsed from a function to a scalar by averaging the coherence values over modulation rates, giving the measure $C_{\text{coh}}(a, b)$.

C. Transformation to the logarithmic domain

Both the correlation and the coherence were calculated using the logarithms of the channel envelopes. One justification for this is that we would expect the measures obtained in this way to be more related to perception than if linear amplitude values were used. A second justification is as follows.

A compressor modifies the envelope of its input, $I(t)$, by use of a time-varying gain signal:

$$O(t) = g(t)I(t), \quad (7)$$

where $O(t)$ is the envelope of the compressor output, and $g(t)$ is the compressor gain. The function $g(t)$ is effectively a modulator. In order to prevent excessive spread of spectral sidebands, which may be perceived as distortion, the gain signal is usually bandlimited by the smoothing performed during the estimation of the envelope (for example, the envelope is often estimated by rectification and lowpass filtering), and by the smoothing used to produce the attack and recovery times. Stone and Moore (1992) showed that the compressor gain signal is given by

$$g(t) = k\tilde{I}(t)^{(1-cr)/cr}, \quad (8)$$

where k is a scaling constant, $\tilde{I}(t)$ is the smoothed envelope signal as extracted by the compressor, and cr is the compression ratio. Combining Eqs. (7) and (8), we get

$$O(t) = k \cdot \tilde{I}(t)^{(1-cr)/cr} I(t). \quad (9)$$

Since $O(t)$ and $I(t)$ are both always positive, we can take the logarithm of both sides. This gives:

$$\log(O(t)) = \log(k) + \frac{(1-cr)}{cr} \log(\tilde{I}(t)) + \log(I(t)). \quad (10)$$

Of the terms on the right-hand side of Eq. (10), $\log(k)$ is a constant that can be ignored for the purpose of calculating correlation or coherence. If the correlation or coherence between $\log(O(t))$ and $\log(I(t))$ is reduced below the maximum value of 1, then this must be caused by the middle term on the right-hand side of Eq. (10). Any reduction below 1 is directly related to the extent to which the estimate of the envelope of the input signal, $\tilde{I}(t)$, differs in shape from the actual envelope of the input signal. This is exactly the type of measure that we wish to achieve. This provides our second justification for basing the calculation of correlation or coherence on the logarithms of the envelopes.

IV. CHOICE OF TEST SIGNALS

The test signals were 90-s “master” files of continuous speech. The background talker was the same male as used in an earlier study using noise-vocoded speech (Stone and Moore, 2003). To produce the target speech, we used continuous speech from a talker with similar production characteristic to those of the male target talker used by Stone and Moore (2003, 2004); we avoided concatenation of target sentences, which would have produced a strong modulation component at the reciprocal of the sentence length. Pauses for breath were edited out of each file, although natural-sounding breaks of between 100 and 140 ms were left between sentences.

To evaluate the effects of single-channel compression, compressor gain control signals were derived for the target alone, background alone, and target plus background, for each compressor to be evaluated. The appropriate gain control signals were applied to the test signals. Then the compressed signals were bandpass filtered to extract channels, and the envelopes of the channel signals were determined for

calculation of WSMC, FES, and ASMC. To evaluate the effects of multi-channel compression, the signal files were processed by bandpass filtering to create channels, the envelopes were extracted, and gain-control signals were calculated for each compressor to be evaluated. The envelope signals for each channel, as well as the gain signals for each compressor, were saved for later analysis.

The starting alignment between target and background was the same for all conditions. All delays produced by the different processing stages were removed. The saved envelope signals were therefore all correctly time-aligned across all processing conditions and also still aligned with the master files. Envelope modulations for rates above 100 Hz were not preserved by the process of envelope extraction. To reduce computation time, the gain and envelope signals were downsampled to a rate of 1000 Hz. This rate was chosen to be higher than initially appears necessary because operations were to be performed on the logarithm of the envelope signals, thereby expanding the frequency range of the modulations and increasing the risk of aliasing (Rabiner and Schaffer, 1978).

The samples of continuous speech were recorded in a near-anechoic sound-isolated chamber. Consequently, in the wideband signal, the signal level dropped markedly (by about 60 dB) between sentences. In the channel signals, similar drops were observed over syllable-length time scales. The compressors evaluated here operated only over a limited range of levels (which is typical of the compressors used in hearing aids and cochlear implants), and since the effective compression ratios were small to moderate, only small gain changes, typically of the order of 6–10 dB, were applied to the signal envelopes. Use of the full dynamic range of the recordings to calculate the various measures would lead to an undue influence of the pauses in the speech.

To avoid undue effects of the low-level portions of the signals, the following strategy was adopted, as illustrated in Fig. 1:

- (1) For the i th bandpass filtered channel of the uncompressed signal, the root-mean-square level (RMS_i) was calculated.
- (2) A “key signal” for each channel was generated; the key signal for channel i was set to zero for all portions of the uncompressed signal with level, measured over a 10-ms interval, below ($RMS_i - 13$ dB) (which were assumed to be mainly noise) and was set to 1 for portions of the signal with levels equal to or above ($RMS_i - 13$ dB).
- (3) The envelope (of the unprocessed or compressed signal) in channel i was processed using the key signal for that channel, generated as described in (2), such that when the key signal was equal to 1 the envelope was left unchanged, and when the key signal was equal to 0 the envelope value was set to $RMS_i - 13$ dB. This is referred to as “thresholding” and is illustrated for channel 2 of an eight-channel vocoder in panel A of Fig. 1. The dashed line shows the original envelope for that channel and the solid line shows the thresholded envelope.
- (4) Following the processing described in (3), any portions of the channel envelope with value below $RMS_i - 13$ dB (which only occurred for compressed signals in parts of the

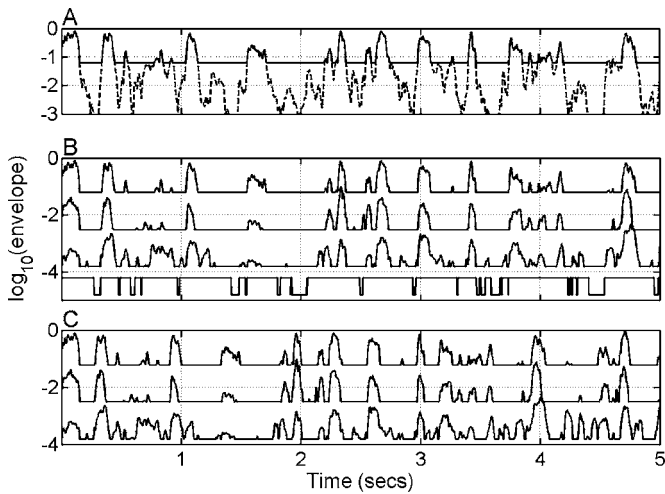


FIG. 1. Illustration of the processes of “thresholding” and generation of the “key signal,” as described in the text. For all panels the absolute positions of the traces are arbitrary. (A) The dashed trace shows the signal envelope for channel 2 of an eight-channel vocoder, and the solid trace shows the envelope after thresholding has been applied. (B) The lowest trace shows a key signal, generated as described in the text. The upper three traces show thresholded envelope signals for channels 2, 4, and 6 of an eight-channel vocoder. (C) The signals for channels 2, 4, and 6 of an eight-channel vocoder after thresholding and application of the key signal.

channel signal for which the key signal had a value of 1) were set to $\text{RMS}_i - 13$ dB; this is effectively a second stage of thresholding.

(5) The measures of ASMC were obtained using the thresholded channel envelope signals obtained as described earlier.

(6) When obtaining measures of WSMC and FES for single-channel compression, a single wideband key signal was generated by indexing only those portions of the wideband signal which were above the wideband compression threshold of whichever of the compressors had the lowest compression threshold (typically this threshold was close to $\text{RMS}_{\text{wide}} - 13$ dB, where RMS_{wide} refers to the RMS level of the wideband signal). To prevent excessive switching on and off of the key signal (for example, during voicing), gaps in the key of less than 10 ms were “ignored,” i.e., they were given a value of 1. When a gap in the key signal was longer than 10 ms, the first 10 ms of the gap was also included in the key signal, effectively extending the key signal by 10 ms. An example of the resulting key signal is shown as the bottom trace in panel B of Fig. 1. The other traces in Fig. 1 show thresholded envelope signals for channels 2, 4, and 6 of an eight-channel vocoder. The wideband key was applied to each thresholded channel envelope signal; portions of the channel signal where the value of the key signal was 0 were excised, and the remaining portions were concatenated. The resulting channel envelope signals are illustrated in panel C of Fig. 1, again for channels 2, 4, and 6 of an eight-channel vocoder. Note that the traces in panel C of Fig. 1 resemble the corresponding traces in panel B of Fig. 1 except that the traces in panel C of Fig. 1 appear progressively shortened in time because portions have been excised. The measures of WSMC and FES were calculated from the resulting channel signals.

(7) When obtaining measures of FES for multi-channel compression, the key signal for channel i (generated as described in step 2) was used to process the thresholded envelope signals for that channel (for the unprocessed and compressed signals). Again, portions of the channel signals where the value of the key signal was 0 were excised, and the remaining portions were concatenated. The measures of FES were calculated from the resulting channel signals.

(8) WSMC involves an across-channel measure, so, to determine the values of WSMC for multi-channel compression, the key signal derived from the i th channel of the unprocessed signal was applied to all channels; portions of the channel signals where the value of the key signal was 0 were excised, and the remaining portions were concatenated. This ensured that the temporal alignment of the channel signals was maintained. The measures of WSMC were calculated from the resulting channel signals.

Although the process of excluding portions of the channel envelope signals may have introduced some “spurious” high-frequency modulations into the envelopes, it ensured that the measures obtained were mainly influenced by the parts of the speech envelope of greatest perceptual relevance (Drullman, 1995). For the measure $C_{\text{coh}}(a, b)$, the effect of spurious envelope modulation at high frequencies was in any case reduced by restricting the range of modulation frequencies used in the analysis.

In what follows we present the above described measures for signals processed through a noise vocoder (Dudley, 1939; Van Tasell *et al.*, 1987 Shannon *et al.*, 1995), which removes temporal fine structure cues but preserves envelope cues in a limited number of frequency channels. This was used since we wished to compare the measures derived here with measures of speech intelligibility, under conditions where mainly envelope cues were available. The number of channels used to derive the measures WSMC, FES, and ASMC were the same as used in the vocoder processing. The envelope for each channel was extracted following all stages of processing, including single or multiple channel compression processing where appropriate, but prior to the recombination of the modulated noise band channel signals. “Frozen” noises were used, i.e., the same noise was used for unprocessed and compressed stimuli. To ensure that the measures were not influenced by the specific noise used, all measures were calculated as the average of measures from five different frozen noises.

V. RESULTS

A. Stimuli used by Stone and Moore (2003)

Stone and Moore (2003) measured the intelligibility of a target speech signal in a background of speech from a single talker. The target-to-background ratio was 5 dB, and the mixture of target and background was either uncompressed or subjected to fast-acting or slow-acting single-channel compression before being processed through a noise vocoder. Results are presented here for 8-channel and 11-channel processing. For the fast compressor, the attack time was 2 ms, the release time was 240 ms, and the static com-

TABLE I. Values of the various measures for the stimuli used by Stone and Moore (2003). The intelligibility data (bottom row) are taken from experiment 1 for the unprocessed stimuli, and are averaged across experiments 2 and 3 for the stimuli subjected to slow or fast compression.

	Unprocessed				Slow				Fast			
	8 channel		11 channel		8 channel		11 channel		8 channel		11 channel	
	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr
WSMC	0.277	0.383	0.271	0.350	0.267	0.347	0.263	0.317	0.215	0.213	0.216	0.209
FES	1.0	1.0	1.0	1.0	0.971	0.976	0.968	0.975	0.868	0.924	0.871	0.925
ASMC	0.024	0.011	0.025	0.012	0.023	-0.010	0.023	-0.006	0.033	-0.074	0.032	-0.061
% intell	80.6		90.6		81.4		90.0		77.4		85.1	

pression ratio was 7. The effective compression ratio (Stone and Moore, 1992) was very close to 7 for envelope modulation rates up to 1 Hz, but declined for higher rates, reaching about 2 for a modulation rate of 3 Hz. The slow compressor was similar to that described as “DUAL-HI” in Stone *et al.* (1999). This compressor incorporated two gain-control “loops” with different time constants (Moore and Glasberg, 1988; Moore *et al.*, 1991). The static compression ratio of the slow compressor was also 7, but the effective compression ratio was close to 1 (no compression) for envelope modulation rates above 0.5 Hz. Table I shows the values of the various measures described earlier for the unprocessed and compressed stimuli, and also shows the mean speech intelligibility scores for twelve normally hearing subjects obtained by Stone and Moore (2003), after correction for learning effects.

For each number of channels, the intelligibility scores did not differ significantly for the unprocessed stimuli and the stimuli subjected to slow-acting compression, but the scores were significantly lower for the stimuli subjected to fast-acting compression than for the other stimuli. The measures of WSMC are slightly lower for slow compression than for the unprocessed stimuli, and are markedly lower for fast compression. In other words, the fast compression reduced WSMC. The measures of FES are also slightly lower for slow compression than for the unprocessed stimuli, and are markedly lower for fast compression. The measures of ASMC are small for the unprocessed stimuli. In theory, one would expect values of zero. The non-zero values presumably are the result of chance fluctuations. However, none of the measures exceeds 0.025. For the slow compression, the measures of ASMC are also close to zero. However, for the fast compression the coherence measures of ASMC have small positive values, while the correlation measures are negative.

The negative values for the correlation measure are consistent with the arguments presented by Stone and Moore (2004). They pointed out that, when the speech of two talkers is mixed, the gain at any instant is determined mainly by the level of the talker with the higher level. Peaks in the speech of one talker result in a reduction in gain for the speech of the other talker. Hence, increases in level of one talker result in decreases in level of the other talker, and vice versa. This inverse relationship is not revealed by the coherence measure, since coherence is constrained to be between 0 and 1. This represents a limitation of the coherence measure.

For the data presented here, all three of the measures vary in an appropriate way for explaining the observed pattern of intelligibility scores. Intelligibility was lowest for the fast compressor, and for this compressor the WSMC and FES measures were lower than for the other conditions, while the ASMC measure of correlation was larger in absolute value (albeit negative). Thus, changes in any or all of the factors, WSMC, FES, and ASMC, might have contributed to the observed decrease in intelligibility for the fast compressor. In addition, the reduced envelope modulation depth produced by the fast compressor may have contributed to the reduced intelligibility. Thus the data do not allow us to tease out the relative importance of these different factors. However, the data presented next do allow us to do this.

B. Stimuli used by Stone and Moore (2004), experiment 1

Stone and Moore (2004) measured the intelligibility of a target speech signal in a background of speech from a single talker. The target-to-background ratio was 2 dB, and the mixture of target and background was subjected to single-channel compression before being processed through an eight-channel noise vocoder. Five types of fast-acting compression were used. One, referred to as the “envelope” compressor, had attack and release times that were similar to each other (10 and 13 ms, respectively). For this compressor the audio signal was delayed before the gain-control signal was applied to reduce “overshoot” and “undershoot” effects (Robinson and Huntington, 1973). The envelope compressor had a static compression ratio of 2.78, and it was effective in reducing the amplitude modulation depth (for sinusoidal modulation) over a wide range of modulation rates (Stone and Moore, 1992, 2003). However, it did this with minimal distortion of the shape of the envelope, so we anticipated that use of this compressor would lead to high values of the measured FES.

The other compression systems were more like those typically used in hearing aids and cochlear implants, where the attack time is much less than the release time, and there is little or no compensating delay of the audio signal. The release time was about 240 ms, and the attack times were 0.125, 0.5, 2 and 8 ms. The static compression ratio was 7, but, except for envelope modulation rates below 2 Hz, these compressors were less effective than the envelope compres-

TABLE II. Values of the various measures for the stimuli used in experiment 1 of Stone and Moore (2004). The bottom row shows the mean intelligibility scores for these stimuli. Intelligibility was not measured for the unprocessed (uncompressed) stimuli.

	Unprocessed		Env. Comp.		$t_{att}=0.125$ ms		$t_{att}=0.5$ ms		$t_{att}=2$ ms		$t_{att}=8$ ms	
	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr
WSMC	0.277	0.383	0.223	0.246	0.213	0.228	0.214	0.229	0.218	0.235	0.225	0.247
FES	1.0	1.0	0.914	0.953	0.870	0.926	0.872	0.927	0.878	0.930	0.877	0.934
ASMC	0.024	0.011	0.034	-0.084	0.033	-0.075	0.033	-0.074	0.033	-0.071	0.033	-0.071
% intell			32.6		37.4		38.1		39.7		44.2	

sor in reducing the amplitude modulation depth, but were all very similar to each other (see Fig. 1 of Stone and Moore, 2004).

Table II shows the values of the various measures described earlier, and also shows the mean speech intelligibility scores obtained by Stone and Moore (2004), after correction for learning effects. The two measures of WSMC (based on coherence and correlation) decreased as a result of the application of fast-acting compression. However, the measures did not vary markedly with the type of compression, and WSMC values were almost the same for the envelope compressor and for the compressor with 8-ms attack time, despite the fact that the latter gave significantly higher intelligibility than the former. This suggests that the WSMC is not the major factor affecting intelligibility for these stimuli. The measures of FES also decreased as a result of the application of fast-acting compression. However, in this case, the FES values were higher for the envelope compressor than for the other compressor, the opposite of the pattern found for intelligibility. Since high FES is presumably a “good thing,” other things being equal, this suggests that FES is also not critical in determining intelligibility for these stimuli.

The application of compression resulted in negative values for the correlation measure of ASMC, but in small positive values for the coherence measure. We focus here on the correlation measure of ASMC, for the reasons discussed earlier. For this measure, the correlation is most negative for the envelope compressor and for the other compressors the correlation becomes less negative as the attack time increases. This pattern of variation is similar to that observed for intelligibility. Note that the effect of attack time on intelligibility cannot be attributed to differences in the effectiveness of the compressors in reducing the envelope modulation depth, since (apart from the envelope compressor) the compressors were all very similar in this respect. Indeed, the relatively small effect of changing the attack time from 0.125 to 8 ms can be attributed to the fact that the ASMC depends on the range of gain values introduced by the compressor (Kates, 2000), and this range was similar across these compressors. Nevertheless, the pattern of ASMC values across all of the compressors (including the envelope compressor) does mirror the pattern of intelligibility changes observed in the results.

It seems reasonable to assume that the more the ASMC measure differs from zero (in either the positive or negative direction), the more the deleterious effect on speech intelligibility. Thus, it appears that, for the stimuli used in experi-

ment 1 of Stone and Moore (2004), the correlation measure of ASMC, which was described as comodulation by Stone and Moore, was the most important factor determining the changes in intelligibility across conditions.

C. Stimuli used by Stone and Moore (2004), experiment 2

In experiment 2 of Stone and Moore (2004), the task of subjects was again to identify speech from a target talker in the presence of speech from a background talker. Only the envelope compressor of their experiment 1 was used, and the single-channel compression was applied either after the target and background were mixed (called condition COMOD by Stone and Moore), or separately to the target and background before mixing (called condition INDEP by Stone and Moore). Vcoders with either six or eleven channels were used. The target-to-background ratio was 8 dB for the six-channel system and 3 dB for the eleven-channel system. These target-to-background ratios were chosen to give moderate values of intelligibility (avoiding floor and ceiling effects), so that the effects of changing conditions could readily be measured. The values of the various measures for these conditions are shown in Table III. Table III also shows the values for the uncompressed speech. Finally, Table III shows the mean intelligibility scores for eight subjects. For each number of channels, scores were significantly higher for condition INDEP than for condition COMOD.

The values of the measures of WSMC were decreased by the application of compression, and the WSMC measures were slightly lower for condition INDEP than for condition COMOD. The fact that WSMC decreased in condition INDEP indicates that the decrease does not depend on the target being mixed with a background signal. As for experiment 1 of Stone and Moore (2004), the pattern of results indicates that WSMC is not the main factor affecting intelligibility for these stimuli.

The values of the measures of FES also decreased following compression, and again the decrease was slightly greater for condition INDEP than for condition COMOD. Consistent with the outcome for experiment 1 of Stone and Moore (2004), this indicates that FES is not the main factor affecting intelligibility for these stimuli.

The correlation measure of ASMC became negative following application of compression in the COMOD condition. As found earlier, the coherence measures of ASMC were close to zero, reflecting the fact that coherence cannot

TABLE III. As in Table II, but for experiment 2 of Stone and Moore (2004).

	Unprocessed				COMOD				INDEP			
	6 channel		11 channel		6 channel		11 channel		6 channel		11 channel	
	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr
WSMC	0.282	0.366	0.271	0.350	0.220	0.194	0.228	0.232	0.216	0.178	0.223	0.203
FES	1.0	1.0	1.0	1.0	0.904	0.947	0.914	0.953	0.896	0.942	0.900	0.945
ASMC	0.025	0.012	0.025	0.012	0.038	-0.102	0.031	-0.071	0.025	0.016	0.025	0.017
% intell					65.6		68.7		89.0		79.7	

be negative. The ASMC measures for condition INDEP were, as expected, close to 0. Thus, the correlation measure of ASMC is the only measure that appears to be related to the obtained intelligibility scores.

D. Values of the measures for multi-channel compression

In this section we consider some stimuli used by Stone and Moore (2007) which were subjected to multi-channel compression. Speech from a target talker and a background talker was compressed by a fast multi-channel compressor either before (condition BEFORE) or after (condition AFTER) the speech from the two talkers was combined; in each channel the compressor was essentially the same as the envelope compressor described earlier. Six and eleven channels of compression were used with target-to-background ratios of +8 and +3 dB, respectively. Again, these ratios were chosen to give moderate intelligibility scores. The results are shown in Table IV, which also shows mean intelligibility scores for twelve subjects. For each number of channels, scores were significantly higher for condition BEFORE than for condition AFTER.

The values of the measures of WSMC were decreased by the application of compression, and for each number of channels the WSMC measures were similar for conditions AFTER and BEFORE. This indicates that WSMC is not the factor responsible for the difference in intelligibility between these two conditions.

The values of the measures of FES also decreased following compression, although they remained rather high (greater than 0.97). This confirms that the envelope compressor does not markedly change the shape of the envelope of the compressed signal. The FES values were similar for con-

ditions AFTER and BEFORE, indicating that FES is not the factor responsible for the difference in intelligibility between these two conditions.

The correlation measure of ASMC became negative following application of compression in condition AFTER. As found earlier, the coherence measures of ASMC were close to zero, reflecting the fact that coherence cannot be negative. The ASMC measures for condition BEFORE were, as expected, close to 0. Thus, the correlation measure of ASMC is the only measure that appears to be related to the obtained intelligibility scores.

VI. DISCUSSION

For three of the four sets of data analyzed here, the measure of ASMC based on correlation was the only one which showed a good correspondence with the obtained measures of speech intelligibility. Note that the absolute values of ASMC (or the other measures used here) cannot be meaningfully compared across different types of conditions (for example, different numbers of channels in the vocoder, different speech-to-background ratios, or different numbers of compression channels). What is important is the pattern of changes of ASMC with manipulation of a parameter such as attack time, or the application of compression before or after mixing of the target and background. Also, it should be noted that ASMC has been shown to be related to intelligibility only under difficult listening conditions when primarily envelope cues are available.

The measure ASMC is related to what was referred to by Stone and Moore (2003, 2004) as comodulation. When either single-channel or multi-channel fast-acting compression is applied to a mixture of a target talker and a background talker, the envelopes of the target and background become

TABLE IV. As in Table II, but for experiment 1 of Stone and Moore (2007), which used multi-channel compression.

	Unprocessed				AFTER				BEFORE			
	6 channel		11 channel		6 channel		11 channel		6 channel		11 channel	
	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr	Coh	Corr
WSMC	0.226	0.288	0.214	0.253	0.224	0.273	0.213	0.243	0.224	0.263	0.214	0.237
FES	1.0	1.0	1.0	1.0	0.971	0.977	0.971	0.980	0.980	0.978	0.984	0.983
ASMC	0.025	0.012	0.025	0.012	0.042	-0.111	0.033	-0.070	0.025	0.014	0.025	0.016
% intell					70.9		67.3		80.8		74.4	

partially negatively correlated. Thus the ASMC measure based on correlation becomes negative. The correlation of the envelopes may reduce the ability of the auditory system to perceptually segregate the target and the background, and this may account for the deleterious effect of applying fast-acting compression to a mixture of signals. The measure of ASMC based on coherence does not seem to be appropriate for revealing this effect, since coherence cannot be negative.

A negative correlation of the envelopes of the target and background implies that the target and background have envelope fluctuations that are partially out of phase. The implications of this were discussed by Stone and Moore (2004). They pointed out that, while the perceptual grouping of two signals appears to be affected by the similarity in the envelopes of the signals, the relative phase of modulation of the signals does not appear to be important. For example, the perceptual segregation of pairs of vowels is not influenced by the relative phase of amplitude modulation of the vowels (Summerfield and Culling, 1992). Also, modulation detection interference, which is believed to result at least partly from perceptual grouping processes (Hall and Grose, 1991; Moore and Shailer, 1992), is not greatly affected by the phase of modulation of the interferer relative to that of the target (Moore and Shailer, 1992; Moore, 1992; Hall *et al.*, 1995). Thus, the ASMC introduced by fast-acting compression could decrease perceptual segregation of the target and background even though the ASMC is negative.

The values of WSMC, based on either coherence or correlation, were reduced by the application of fast-acting compression (relative to those for no compression), even when the compression was applied to the target speech alone (conditions INDEP and BEFORE). This effect may appear surprising, especially for condition INDEP (when single-channel compression was used), since one might have expected that application of a time-varying gain to the broadband signal would result in an increase in the WSMC. The decrease probably occurs because, in response to a speech signal, the gain over a short time interval is determined by the spectral components that are most intense during that interval. When a given spectral region has a higher-than-average level, the gain for the whole signal is reduced, so components in other spectral regions become lower in level. This may account for the observed reduction in WSMC produced by fast compression.

For the unprocessed (uncompressed) speech, the value of the correlation measure of WSMC was higher than for the coherence measure. However, for the compressed speech, the two measures were more similar, and sometimes the correlation measure had a lower value than the coherence measure. Thus, while both measures appear to be suitable for characterizing WSMC, the correlation measure appears to be more sensitive to the effects of fast compression.

The values of the FES measure do not appear to be closely related to intelligibility. This may indicate that the exact shape of the envelope of speech in different frequency channels is not of critical importance for intelligibility. However, it may also be the case that the measure used here does not capture one or more salient aspects of the envelope shape. It has been proposed that the envelopes of auditory

signals in different frequency channels are subjected to a spectral decomposition in the modulation frequency domain, via a "modulation filter bank" (Kay, 1982; Dau *et al.*, 1997). It may be the case that alteration of the envelope shape does not have serious consequences if it results only in moderate changes in the relative outputs of different modulation filters, for example, reduction of the outputs of filters tuned to low modulation rates. However, if a compressor distorts the envelope shape in such a way as to introduce components in the modulation spectrum that were not originally present, this may have more serious consequences. This topic remains to be explored. It should be noted that FES, in common with the other measures introduced here, gives a single overall measure of the changes produced by compression, and does not give any insight into the nature of the changes or into the components of the modulation spectrum that have been most affected.

VII. SUMMARY AND CONCLUSIONS

Fast-acting compression can have a variety of deleterious effects on the envelopes of speech signals, especially when two speech signals, a target and a background, are mixed together. In this paper, we have described three types of envelope changes that may be produced by compression, and we have described measures which can be used to characterize those changes. The measures are:

- (1) Within-signal modulation correlation or coherence, which is defined as the degree of correlation (or coherence) of the envelope (on a decibel scale) of a single source across different frequency regions (i.e., after filtering into frequency channels). A high value of WSMC may help to perceptually bind together the different frequency components of the target speech, so a reduction in WSMC may lead to reduced intelligibility, especially when a background talker is present.
- (2) Fidelity of envelope shape, which is defined as the degree to which the envelope shape of the target speech in different frequency channels is preserved following compression. It is assumed that a reduction in FES may be associated with reduced speech intelligibility.
- (3) Across-signal modulation coherence, which is defined as the extent to which the target and background acquire a common component of modulation when they are compressed together. A value of ASMC different from zero, in either the positive or negative direction, may cause the target and background to be perceptually grouped, thereby reducing the intelligibility of the target.

For each of these quantities, measures based both on coherence and on correlation were determined. Both coherence and correlation appear appropriate for characterizing WSMC and FES. However, only correlation is appropriate for characterizing ASMC, since coherence cannot be negative, but fast-acting compression results in the envelopes of the target and background becoming negatively correlated.

The changes of the values across conditions, for a given type of processing, were compared with measured intelligibility scores, obtained using stimuli processed using a noise-vocoder, so as to preserve mainly envelope cues in a limited

number of frequency channels. The results suggest that the dominant factor affecting intelligibility, under difficult listening conditions when primarily envelope cues are available, is ASMC.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Tom Baer and three anonymous reviewers for helpful comments on an earlier version of this paper.

- Boothroyd, A., Springer, N., Smith, L., and Schulman, J. (1988). "Amplitude compression and profound hearing loss," *J. Speech Hear. Res.* **31**, 362–376.
- Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (1985). "Spectral integration based on common amplitude modulation," *Percept. Psychophys.* **37**, 483–493.
- Buss, E., Hall, J. W., III, and Grose, J. H. (2004). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," *Ear Hear.* **25**, 242–250.
- Carrell, T. (1993). "The effect of amplitude comodulation on extracting sentences from noise: Evidence from a variety of contexts," *J. Acoust. Soc. Am.* **93**, 2327.
- Carrell, T. D., and Opie, J. M. (1992). "The effect of amplitude comodulation on auditory object formation in sentence perception," *Percept. Psychophys.* **52**, 437–445.
- Carter, G. C., Knapp, C. H., and Nuttall, A. H. (1973). "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," *IEEE Trans. Audio Electroacoust.* **21**, 337–344.
- Crouzet, O., and Ainsworth, W. A. (2001). "On the various instances of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation," in *Workshop on Consistent and Reliable Cues for Sound Analysis* (Aalborg, Denmark).
- 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 Gennaro, S., Braidia, L. D., and Durlach, N. I. (1986). "Multichannel syllabic compression for severely impaired listeners," *J. Rehabil. Res. Dev.* **23**, 17–24.
- Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* **97**, 585–592.
- Drullman, R., and Smoorenburg, G. F. (1997). "Audio-visual perception of compressed speech by profoundly hearing-impaired subjects," *Audiology* **36**, 165–177.
- 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.
- Hall, J. W., III, and Grose, J. H. (1991). "Some effects of auditory grouping factors on modulation detection interference (MDI)," *J. Acoust. Soc. Am.* **90**, 3028–3035.
- Hall, J. W., III, Grose, J. H., and Mendoza, L. (1995). "Across-channel processes in masking," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
- Healy, E. W., and Bacon, S. P. (2002). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," *J. Speech Lang. Hear. Res.* **45**, 1262–1275.
- Kates, J. M. (1992). "On using coherence to measure distortion in hearing aids," *J. Acoust. Soc. Am.* **91**, 2236–2244.
- Kates, J. M. (2000). "Cross-correlation procedures for measuring noise and distortion in AGC hearing aids," *J. Acoust. Soc. Am.* **107**, 3407–3414.
- Kay, R. H. (1982). "Hearing of modulation in sounds," *Physiol. Rev.* **62**, 894–975.
- Laurence, R. F., Moore, B. C. J., and Glasberg, B. R. (1983). "A comparison of behind-the-ear high-fidelity linear aids and two-channel compression hearing aids in the laboratory and in everyday life," *Br. J. Audiol.* **17**, 31–48.
- Moore, B. C. J. (1992). "Across-channel processes in auditory masking," *J. Acoust. Soc. Jpn. (E)* **13**, 25–37.
- 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., and Glasberg, B. R. (1988). "A comparison of four methods of implementing automatic gain control (AGC) in hearing aids," *Br. J. Audiol.* **22**, 93–104.
- Moore, B. C. J., and Shailer, M. J. (1992). "Modulation discrimination interference and auditory grouping," *Philos. Trans. R. Soc. London, Ser. B* **336**, 339–346.
- Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (2006). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure information," *Hear. Res.* **222**, 16–27.
- Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (1991). "Optimization of a slow-acting automatic gain control system for use in hearing aids," *Br. J. Audiol.* **25**, 171–182.
- Moore, B. C. J., Stone, M. A., and Alcántara, J. I. (2001). "Comparison of the electroacoustic characteristics of five hearing aids," *Br. J. Audiol.* **35**, 307–325.
- Moore, B. C. J., Johnson, J. S., Clark, T. M., and Pluinage, V. (1992). "Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss," *Ear Hear.* **13**, 349–370.
- Plomp, R. (1983). "The role of modulation in hearing," in *Hearing - Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin).
- Plomp, R. (1988). "The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation-transfer function," *J. Acoust. Soc. Am.* **83**, 2322–2327.
- 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.
- Rabiner, L. R., and Schafer, R. W. (1978). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).
- Rappold, R. W., Mendoza, L., and Collins, M. J. (1993). "Measuring the strength of auditory fusion for synchronously and non-synchronously fluctuating narrow-band noise pairs," *J. Acoust. Soc. Am.* **93**, 1196–1199.
- Robinson, C. E., and Huntington, D. A. (1973). "The intelligibility of speech processed by delayed long-term averaged compression amplification," *J. Acoust. Soc. Am.* **54**, 314.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Stone, M. A., and Moore, B. C. J. (1992). "Syllabic compression: Effective compression ratios for signals modulated at different rates," *Br. J. Audiol.* **26**, 351–361.
- Stone, M. A., and Moore, B. C. J. (2003). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," *J. Acoust. Soc. Am.* **114**, 1023–1034.
- Stone, M. A., and Moore, B. C. J. (2004). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," *J. Acoust. Soc. Am.* **116**, 2311–2323.
- Stone, M. A., and Moore, B. C. J. (2007). "Effects on intelligibility of spectro-temporal modulation changes produced by multi-channel compression" (submitted to JASA).
- Stone, M. A., Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (1999). "Comparison of different forms of compression using wearable digital hearing aids," *J. Acoust. Soc. Am.* **106**, 3603–3619.
- Summerfield, Q., and Culling, J. F. (1992). "Auditory segregation of competing voices: Absence of effects of FM or AM coherence," *Philos. Trans. R. Soc. London, Ser. B* **336**, 357–366.
- Tan, C. T., Moore, B. C. J., Zacharov, N., and Matilla, V.-V. (2004). "Predicting the perceived quality of nonlinearly distorted music and speech signals," *J. Audio Eng. Soc.* **52**, 699–711.
- 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.
- Verschuure, J., Maas, A. J. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (1996). "Compression and its effect on the speech signal," *Ear Hear.* **17**, 162–175.
- Villchur, E. (1973). "Signal processing to improve speech intelligibility in perceptive deafness," *J. Acoust. Soc. Am.* **53**, 1646–1657.
- Yund, E. W., and Buckles, K. M. (1995). "Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise," *J. Acoust. Soc. Am.* **97**, 1206–1223.
- Zeng, F.-G. (2004). "Compression and cochlear implants," in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, A. N. Popper, and R. R. Fay (Springer, New York).

Word-internal versus word-peripheral consonantal duration patterns in three languages

Melissa A. Redford^{a)}

Department of Linguistics, University of Oregon, Eugene, Oregon 97403

(Received 12 April 2006; revised 8 December 2006; accepted 11 December 2006)

Segmental duration patterns have long been used to support the proposal that syllables are basic speech planning units, but production experiments almost always confound syllable and word boundaries. The current study tried to remedy this problem by comparing word-internal and word-peripheral consonantal duration patterns. Stress and sequencing were used to vary the nominal location of word-internal boundaries in American English productions of disyllabic nonsense words with medial consonant sequences. The word-internal patterns were compared to those that occurred at the edges of words, where boundary location was held constant and only stress and sequence order were varied. The English patterns were then compared to patterns from Russian and Finnish. All three languages showed similar effects of stress and sequencing on consonantal duration, but an independent effect of syllable position was observed only in English and only at a word boundary. English also showed stronger effects of stress and sequencing across a word boundary than within a word. Finnish showed the opposite pattern, whereas Russian showed little difference between word-internal and word-peripheral patterns. Overall, the results suggest that the suprasegmental units of motor planning are language-specific and that the word may be more a relevant planning unit in English. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431339]

PACS number(s): 43.70.Bk, 43.70.Fq, 43.70.Kv [AL]

Pages: 1665–1678

I. INTRODUCTION

Since Kozhevnikov and Chistovich (1965) and Lehiste (1970) researchers have used segmental duration patterns to gain insight into the representation of sequential speech action. Decades of work have provided a wealth of information on the intrinsic and sequential factors influencing articulation as well as on the relationship between linguistic units and speech planning. A widely accepted conclusion that has emerged from this work is that articulatory timing is determined in part by the dynamics of sequential segment articulation and in part by suprasegmental units. Research continues in order to ascertain the nature and representation of the suprasegmental units. The current study fits within this tradition.

Specifically, the current study focuses on consonantal duration patterns that are attributed to the syllable, which has been repeatedly proposed as the most basic suprasegmental unit of articulatory organization in the speech plan. A principal problem with this proposal is that the syllable is ill-defined in that its boundaries are often ambiguous. Such ambiguity makes it difficult to test whether or not the syllable is a relevant unit of speech motor control. The current study attempts such a test, but also explores an alternative hypothesis; namely, that syllable-related segmental duration patterns emerge from the articulatory dynamics of producing meaningful bits of speech—lexical items, for example—rather than from a phonological plan for executing these items in prespecified chunks.¹ This possibility suggests that some putative syllable effects on segmental duration may be

better explained by sequential and communicative factors rather than by the syllable per se. In this study, the effect of these other factors on consonantal duration patterns are explored and the results are discussed with reference to a model of speech production that is inspired by MacNeilage's (1998) frame/content theory.

A. Syllable-related segmental duration patterns

Production studies on the acoustic correlates of the syllable have found that vowel duration varies with syllable structure and consonant duration varies with syllable position. For instance, phrase-medial syllable-initial consonants are longer than syllable-final consonants (Boucher, 1988; Tuller and Kelso, 1991; Keating, Wright, and Zhang, 1999; Redford and Diehl, 1999) and the internal members of consonant clusters are shorter than the external member (Haggard, 1973; Klatt, 1976; Umeda, 1977). Perception studies indicate that these duration patterns are also used by listeners to infer syllable boundaries (Christie, 1977; Tuller and Kelso, 1991; Redford and Randall, 2005). For example, an intervocalic consonant-consonant sequence that is a possible onset cluster to the following vowel will be syllabified as an offset-onset sequence when produced with a short-long pattern, and as an onset cluster when produced with a long-short pattern (Redford and Randall, 2005). Although most of the work identifying consonantal duration patterns as correlates of syllable structure has been conducted in English, some work suggests that the relationship may be universal (e.g., Quené, 1992; Prieto, 2002; Maddieson, 1985, 2004).

The fact that segmental duration patterns covary with syllable structure and help cue syllable boundary perception has often been taken as evidence that syllables are supraseg-

^{a)}Electronic mail: redford@uoregon.edu

mental units of articulatory organization. Specifically, segments are often thought to be temporally adjusted (or timed relative to one another) within a syllable frame (Kozhevnikov and Chistovich, 1965; Lehiste, 1970; Haggard, 1973; Klatt, 1976; Fowler, 1979; Boucher, 1988; Browman and Goldstein, 1988; Campbell and Isard, 1991; Campbell, 1992; Munhall *et al.*, 1992; Turk, 1994; Krakow, 1999; de Jong, 2003). Campbell (1992), for instance, suggested that segments are elastic in their durations and are stretched or compressed uniformly across the syllable, which is independently lengthened or shortened according to prosodic factors or rate of speech. van Santen and Shih (2000) questioned this version of the syllable frame hypothesis, noting that syllable duration increases with the number of segments. In addition, van Santen and Shih showed that intrinsic segmental duration creates systematic and large duration differences between identically structured syllables in the same prosodic context. Their finding suggests that segmental factors may sometimes provide a better explanation than syllables for local consonantal duration patterns.

Another version of the syllable frame hypothesis emphasizes syllabic structure over the temporal constancy of the syllable, but assumes a temporal frame nonetheless. In this more abstract version of the hypothesis, syllable structure complexity results in compensatory lengthening or shortening of segments in particular positions. For instance, Munhall *et al.* (1992) explained that vowel duration decreases within a syllable as the number of postvocalic consonants increases because the additional consonants drive increasing overlap between vocalic and consonantal articulation. The increased articulatory overlap between adjacent vowels and consonants only makes sense if an organizational domain is assumed: some boundary forces earlier and earlier articulation of postvocalic consonants as consonant number increases. Munhall *et al.* (1992) suggested that the relevant domain/boundary was syllabic. However, their experiment examined articulatory organization within monosyllabic words, so it remains unclear whether word-internal syllable boundaries could also force the same pattern of increasing overlap. In the absence of evidence for syllabic effects on word-internal articulatory timing, it is unclear whether the relevant organizational domain is the syllable or the word.

In sum, segmental factors may often explain durational patterns attributed to the syllable, but higher-level units must be invoked to explain certain boundary-adjacent patterns. The identity of these higher-level units is still under dispute.

B. An alternative explanation for consonantal duration patterns

Munhall *et al.*'s (1992) study is not unique in confounding syllable and word boundaries. By far the most popular method of manipulating syllable structure in production experiments is to insert a word boundary where a syllable boundary is desired. This practice is justified by the fact that syllable boundaries are most easily defined with reference to a word boundary, but the problem is that the factors governing production at a word boundary are likely to be different than those governing production word-internally. For instance, we already know that articulatory strengthening and

lengthening characterize segmental production at word and phrase boundaries (Fougeron and Keating, 1997; Byrd and Saltzman, 2003; Keating *et al.*, 2003; Tabain, 2003). Similar effects have not been observed word-internally (Keating *et al.*, 1999). Although this may be because word-internal patterns have not been as thoroughly explored as word-peripheral patterns, the absence of word-internal strengthening and lengthening could also indicate that the relevant linguistic units for speech motor planning are those that are also important to message encoding. After all, the motor goal of speech is to convey meaning and the location of a word-internal syllable boundary is not itself meaningful. For example, a word such as *city* can be produced by different speakers with different syllable structures (or, at least, it can be analyzed differently by different linguists) without affecting meaning [see, e.g., Kahn (1976) versus Selkirk (1982)]. Although the different pronunciations of the same word can be described as yielding different syllable structure percepts (cf. [cɪt^hi] versus [cɪri]) this need not imply that the planning domain is different in the two cases. Instead, it may imply that different global production strategies have been employed to achieve different communicative goals (e.g., slow and careful versus fast and casual speech)—and these need not refer to syllable structure to produce effects that are attributed to syllable structure (e.g., aspirated /t/ versus the flapped allophone).

If sequential action is organized within word units rather than within syllable units, then the observed correlation between segmental duration patterns and word-edge structure must still be explained. One possibility is that the correlation evolved as a listener-oriented strategy to facilitate speech segmentation (Kohler, 1991; Quené, 1992; Fougeron and Keating, 1997; Keating *et al.*, 1999). If this is the case, then the research emphasis on English may have over-represented the strength of the correlation between segmental duration patterns and structure because English users may have a stronger need to cue boundary location than speakers of other languages: English has a large repertoire of sounds that can occur both word-finally and word-initially, and very little inflectional morphology to make boundaries more predictable. Further, the durational patterns at the edges of English words have been shown to help resolve word boundary ambiguity. For instance, the lengthening of /n/ relative to /s/ in the sentence *help us nail* distinguishes the *us nail* parse from an *a snail* parse (Christie, 1977). Intriguingly, these boundary-distinguishing patterns of relative consonantal duration emerge slowly in children's speech—at least a year after children are able to produce the relevant monosyllabic word shapes (i.e., syllable structures) in connected speech (Redford and Gildersleeve-Neumann, unpublished). This suggests that consonantal duration patterns around word boundaries may not be inherent to the production of different syllable structures, but learned as the child slowly gains control over timing one word relative to another. Again, the motivation for gaining such control may be to highlight meaningful boundaries for the listener. This possibility will be referred to here as the juncture hypothesis.

Of course, the juncture hypothesis begs the question of why particular durational patterns should be associated with

particular word-onset and -offset structures. Whereas initial lengthening and strengthening may increase the perceptual prominence of a consonantal onset, thereby signaling the beginning of a word, there is no equivalent explanation for why the second consonant in a word-onset or -offset sequence would be shorter than the first. However, if we assume that the word is the most basic unit in the speech plan, then these patterns could simply represent within-unit coarticulation or articulatory overlap. That is, an account like the one offered by Munhall *et al.* (1992) for vowel shortening in monosyllabic words might be invoked to explain why internal members of a word-onset or -offset cluster are truncated.

This example suggests an important corollary to the juncture hypothesis; namely, that word-internal durational patterns may not refer to syllable structure, but may instead reflect the articulatory dynamics of sequential segment articulation. These dynamics are influenced by segment identity and sequencing as well as by the word's rhythmic structure. Specifically, the different articulatory configurations associated with different segments are achieved at different rates, even when other factors are held constant, resulting in different intrinsic segmental durations (Klatt, 1976; Umeda, 1975, 1977). When segmental context is varied, the same configurations will be achieved in more or less time; that is, coarticulation affects segmental duration (Umeda, 1975, 1977). Rhythm also affects articulatory dynamics. This assertion is especially well documented for the rhythmic structure derived from lexical stress (Edwards and Beckman, 1988; de Jong, Beckman, and Edwards, 1993; de Jong, 1995; Harrington, Fletcher, and Roberts, 1995). For example, de Jong characterized lexical stress as localized hyperarticulation: "more extreme targeting of articulator movement" (de Jong, 1995, p. 497). More extreme articulator movement takes longer to execute, resulting in longer segmental durations (e.g., Moon and Lindblom, 1994).

In sum, the word is proposed as an alternative unit to the syllable to explain boundary-related consonantal duration patterns in English. By extension of the hypothesis, word-internal patterns are assumed to emerge from within-unit articulatory dynamics.

C. The current study

The syllable frame and juncture hypotheses make different predictions regarding word-internal segmental duration patterns. The syllable frame hypothesis predicts that word-internal patterns will reflect syllable structure. The juncture hypothesis predicts that such patterns emerge from within-unit articulatory dynamics, which are affected by lexical factors such as stress and the identity and sequencing of the sounds within a word. The goal of the current study was to test between these predictions to determine whether the syllable frame or juncture hypothesis provides a better explanation for segmental duration patterns. To do this, we used disyllabic nonsense word stimuli with intervocalic consonant sequences and manipulated syllabic affiliation of the consonants using stress and sequence order. The resulting consonantal duration patterns were compared to those that oc-

curred at the edges of words, where word-boundary location was held constant and stress and sequence order were varied.

The full manipulation was carried out in English. The results were then compared to those from Finnish and Russian, where the stress manipulation was across languages. The experiment was extended from English to Finnish and Russian to obtain more general conclusions. The extension was also motivated by the juncture hypothesis, which posits that boundary-adjacent consonantal patterns in English may be due to listener-oriented boundary highlighting. If it occurs, such highlighting could interfere with effects from stress and sequencing at the edges of words, making the word-internal results less interpretable as well (for lack of comparison). The assumption is that boundary highlighting is less crucial in Finnish and Russian. Both languages have extensive inflectional morphology, providing strong top-down cues to segmentation that do not exist in English. Also, word-edge phonotactics in Finnish are restrictive, providing an additional aid to segmentation.

II. COMPARISON OF WORD-INTERNAL AND WORD-PERIPHERAL CONSONANTAL DURATION PATTERNS IN ENGLISH

Syllable affiliation is most readily determined by position relative to the vowel at the edge of a word: a word-initial consonant is also syllable-initial, a word-final consonant is also syllable-final. Word-internal syllable affiliation is not as easy to determine for all sequence types. In particular, consonant sequences that can occur as word-onset clusters may be syllabified word-internally either as an offset-onset sequence or as an onset cluster to the following vowel. Psycholinguistic studies suggest that the particular syllabification will depend primarily on lexical stress since a stressed vowel attracts consonants (Treiman and Danis, 1988; Treiman and Zukowski, 1990; Derwing and Nearey, 1991; Redford and Randall, 2005). This means that one way to manipulate word-internal syllable boundaries is to manipulate stress, which is what was done in the present experiment. By itself, however, such a manipulation is inadequate for distinguishing between the syllable frame and juncture hypotheses: syllable structure changes with different stress patterns, but so do the local articulatory dynamics. One way to control for this problem is to vary stress while keeping the syllable boundary constant. This can be done at a word boundary. Accordingly, the experiment reported in this section compares consonantal duration patterns within a word (word-internal) and across a word boundary (word-peripheral) as a function of stress, which was manipulated both within the word and across the word boundary.

Another way to examine the independent effects of syllable structure and stress is to examine the effects of stress on word-internal consonant sequences that can only be syllabified as an offset-onset sequences. This can be done by inverting a sequence that forms a legal onset, thereby creating an illegal onset. Manipulating consonant sequences in this way has the added advantage of manipulating consonant position vis-à-vis a vowel and therefore syllable affiliation. For instance, when a stop-liquid sequence is inverted to be-

come a liquid-stop sequence, the liquid becomes syllable-final instead of syllable-initial, and vice-versa for the stop depending on word position and stress.

In addition to changing syllable affiliation, sequencing changes also affect articulatory dynamics. For instance, pre- and postvocalic liquids are articulated differently (Sproat and Fujimura, 1993; Gick *et al.*, 2005). In a prevocalic environment a liquid is formed first with an anterior gesture then with a posterior gesture. The reverse is true in a postvocalic environment. Such changes in articulatory timing could result in position-dependent differences in the acoustic duration of liquids. A change in sequence order will also affect consonant-consonant coarticulation, which could affect acoustic duration. For instance, stop coarticulation with a subsequent liquid will result in greater oral constriction during the stop release than when the stop is coarticulated with a subsequent vowel. Greater oral constriction during the release gesture in a stop-sonorant sequence may increase “the time required to re-establish the transglottal pressure drop necessary for the initiation of voicing” compared with the less constricted gesture associated with the stop release in a stop-vowel sequence (Docherty, 1992; p. 148). This, in turn, means that stop aspiration duration is likely to be longer in a stop-liquid sequence than in a liquid-stop sequence when the stop precedes a vowel in the latter instance (*viz.* Docherty, 1992).

Sequencing effects arising from vowel-consonant or consonant-vowel coarticulation may be inseparable from effects of syllable position, but it is nonetheless important to examine these effects if only to explain syllable-related patterns in terms of more basic production processes. On the other hand, sequencing effects arising from consonant-consonant coarticulation may be minimized at a boundary that separates consonants, since it is assumed that boundaries reduce coarticulatory pressure. Although an effect of syllable boundary on consonant-consonant coarticulation would provide strong support for the syllable frame hypothesis and would allow us to reject the juncture hypothesis, the lack of an effect will not allow us to reject the syllable frame hypothesis. Some will argue that any absence of an effect of syllable structure on articulation is consistent with the syllable’s position in the prosodic hierarchy: research on the articulatory dynamics at word versus higher-level phrase boundaries (accentual, intonational, and utterance) has shown that weaker prosodic boundaries have weaker effects on coarticulation than stronger prosodic boundaries [*see, e.g., Tabain (2003)*] and syllable boundaries are presumed to be weaker than word boundaries. If one accepts that the weakest prosodic boundaries may have no effect on coarticulation, then it would seem that only the juncture hypothesis can be falsified in the present experiment—a point that is relevant to interpreting the results that are presented after the experimental methods below.

A. Methods

1. Speakers and stimuli

Six native American-English speakers—three male and three female—produced a series of disyllabic nonsense

TABLE I. The experimental stimuli were designed so that the intervocalic stop-liquid and liquid-stop sequences occurred either within a word (internal) or across a word boundary (peripheral).

Stop-liquid		Liquid-stop	
Internal	Peripheral	Internal	Peripheral
jɪt4a	mɛjɪt ɪanou	ɟɪɪ4ta	mɛɟɪ tanou
kɛt4a	mɛkɛt ɪanou	kɛɪta	mɛkɛɪ tanou
lɪtɛ	mɛlɪt ɪɛnou	lɪɪtɛ	mɛlɪɪ tɛnou
mɛt4a	mɛmɛt 4anou	mɛ4ta	mɛmɛɪ tanou
nɛkla	mɛnɛk ɪanou	nɛlka	mɛnɛl kanou
sɛkla	mɛsɛk ɪanou	sɛlka	mɛsɛl kanou
tɪkla	mɛtɪk ɪanou	tɪlka	mɛtɪl kanou
vɪkle	mɛvɪk ɪɛnou	vɪlke	mɛvɪl kɛnou

words and nonsense word pairs (see Table I). The words contained stop-liquid or liquid-stop sequences that occurred word-internally or across a word boundary (*e.g., nɛkla* and *nɛlka* versus *mɛnɛk lano* and *mɛnɛl kano*). There were eight words and eight word pairs with each sequence order. Half of the set had intervocalic -kɪ- or -lɪ- sequences and half had intervocalic -tɪ- or -ɪt- sequences. These two consonant combinations were chosen with a cross-language comparison in mind: the sequences obey word-medial and word-edge phonotactics in English, Russian, and Finnish.

The stimuli were embedded in the frame sentence, *Say __ again*, and presented to speakers in writing as a randomized list of phrases. Each of these phrases was produced as a coherent utterance with one low-high-low intonational contour that focused the nonsense word stimuli and tended to peak on the first nonsense word when the stimulus was a nonsense word pair.

To ensure consistent vowel pronunciation across speakers, the stimuli were represented using the International Phonetic Alphabet (IPAs) *i.e.*, as a phonemic transcription, *e.g., nɛkla*). Speakers were reminded of or taught the sound correspondence of all non-English graphemes (*e.g., /ɛ/* and */ɪ/*) and were asked to practice producing all of the stimuli before the recording session began. Feedback during practice and recording was given only when a symbol was clearly misinterpreted (*e.g., /ɛ/* for */ɪ/*). With respect to pronunciation, it is important to emphasize that the experimenter and speakers treated the IPA representation as mere orthography: speakers produced whatever allophonic variation they found natural given the different contexts and conditions.

Each of the stimuli was repeated three times for a total of 96 phrases per speaker, that is, 24 tokens with word-internal stop-liquid sequences, 24 tokens with word-peripheral stop-liquid sequences, 24 tokens with word-internal liquid-stop sequences, and 24 tokens with word-peripheral liquid-stop sequences. All 96 phrases were spoken twice—once with the weak-strong stress pattern typical of English verbs, and once with the strong-weak stress pattern typical of English nouns. Stress was marked as vowel accent, and all stimuli were produced with the same weak-strong or strong-weak pattern in a given recording session (*e.g., session 1: nɛklá, mɛnɛk lano*; session 2: *nɛkla, mɛnɛk lano*). The two sessions were conducted on separate days, with the

different stress patterns completed in different random orders by different speakers. Speakers practiced the stress pattern of the day at the same time they practiced pronouncing the stimuli.

2. Measurement and analyses

The utterances were recorded in a sound-attenuated room using a Shure BG 5.1 microphone, and saved in digital format. They were later displayed in Praat (Boersma and Weenink, 2002) as oscillograms and spectrograms, and the acoustic durations of the intervocalic stop and liquid were measured. The measurement criteria were as follows. Stop boundaries were defined by a sudden drop/rise in the amplitude of a periodic wave form. Closure duration and burst plus aspiration duration (henceforth aspiration) were measured separately for clearer insight into the relationship between acoustic duration and articulatory changes in the vocal tract. Liquid boundaries were defined by amplitude and frequency changes in the periodic wave form on one side, and by the stop boundary on the other. In particular, the boundary between an /l/ and a vowel was located at the point where energy in the midfrequencies abruptly changed and wave form amplitude increased. The boundary between an /l/ and an adjacent vowel was also placed wherever there was an abrupt change in energy. If no such changes occurred, then the boundary was placed at the temporal midpoint of the transitioning third formant (Klatt, 1976). Ambiguity in defining segment boundaries was resolved by repeated listening to different sections of the wave form and by establishing the boundary where spectral and amplitude changes were observed that corresponded to an auditory segmentation of the wave form.

To test for measurement consistency, 10% of the tokens produced by all the speakers were randomly selected for re-measurement and then the intrarater reliability was assessed using Cronbach's coefficient alpha. In all cases, the coefficient alpha was high [stop closure, $\alpha=0.91$; aspiration, $\alpha=0.96$; liquid duration, $\alpha=0.93$], indicating excellent internal measurement consistency.

Since stress was expected to affect consonantal duration, the acoustic correlates of stress were also measured; namely, vowel pitch, amplitude, and duration. The goal of measuring the stress correlates was to establish that the stress manipulation had worked and that there were differences in the relative strength of the vowels on either side of the intervocalic consonants in question. Pitch (f0) and amplitude (dB) measures were made at the midpoint of each of the flanking vowels, whether these were in a single word (e.g., *nekla*) or whether these spanned a word boundary (e.g., *menεk lano). The duration of these vowels was measured from vowel onset to offset using the segmentation criteria discussed above for consonants. The pitch, amplitude, and duration measures were then converted to obtain a single dependent variable for the repeated measures analyses of variance (ANOVA) that would reflect the relative strength of the two flanking vowels. For the pitch and duration measures, this value was the ratio of the pre- and postconsonantal pitch or duration values. For the amplitude measure, it was a difference value.*

All measures were evaluated as a function of the within-subject factors in two- or three-way repeated measures ANOVAs. The relative strength of vowels was evaluated as a function of two within-subjects factors: stress (weak or strong) and word position (internal or across a word boundary). Consonantal durations were evaluated as a function of three within-subjects factors: stress, sequence order (stop-liquid and liquid-stop), and position. In both the vocalic and consonantal analyses, the levels in the factor stress represented the nominal stress pattern of the pre- and postconsonantal vowels, not the nominal stress pattern of the disyllabic nonsense word or word pairs. This means that items with a weak-strong pattern, for example, included the single nonsense words produced with weak-strong stress and nonsense word pairs produced with strong-weak stress (*SW#SW*). Stress was coded in this way for the analyses to facilitate interpretation of the results.

B. Results

1. Stress

The analyses showed that the relative strength of the pre- and postconsonantal vowels varied with stress, as expected [f0 ratio, $F(1,287)=467.21$, $p<0.01$, $\eta_p^2=0.62$; dB difference, $F(1,287)=786.71$, $p<0.01$, $\eta_p^2=0.73$; duration ratio, $F(1,287)=793.08$, $p<0.01$, $\eta_p^2=0.73$]. The second vowel was longer, louder, and higher pitched than the first vowel when these were produced with weak-strong stress. Conversely, the first vowel was longer, louder, and higher pitched than the second vowel when these were produced with strong-weak stress. This effect of stress on relative pitch and amplitude was greatly reduced when a word boundary intervened, as is evident from Fig. 1 where the significant interaction between position and stress pattern is shown for the three dependent variables [f0 ratio, $F(1,287)=393.04$, $p<0.01$, $\eta_p^2=0.58$; dB difference, $F(1,287)=350.87$, $p<0.01$, $\eta_p^2=0.55$; duration ratio, $F(1,287)=157.26$, $p<0.01$, $\eta_p^2=0.35$]. Post hoc tests indicate that the stress pattern did not affect the relative pitch of vowels on either side of a word boundary [$F(1,287)=3.09$, NS], though it did affect their relative amplitude [$F(1,287)=56.43$, $p<0.01$]. In contrast to relative pitch and amplitude, stress had a large effect on the relative duration of vowels on either side of a word boundary, as shown in the bottom panel of Fig. 1.

The different patterns of interaction between stress and position for the three dependent variables is likely due to an interaction with focal accent. The six speakers placed the focal accent of the phrase on the nonsense word. If the stimulus consisted of two disyllabic words, the focal accent was typically placed on the first word rather than on the second. Focal accent likely increased the pitch and amplitude of the stressed syllable on words where accent peaked, but not on those outside the accent peak (van Sluijter and van Heuven, 1996). Vowel duration, which appeared to be unaffected by peak focal accent, is therefore the more reliable indicator of the nominal stress pattern.

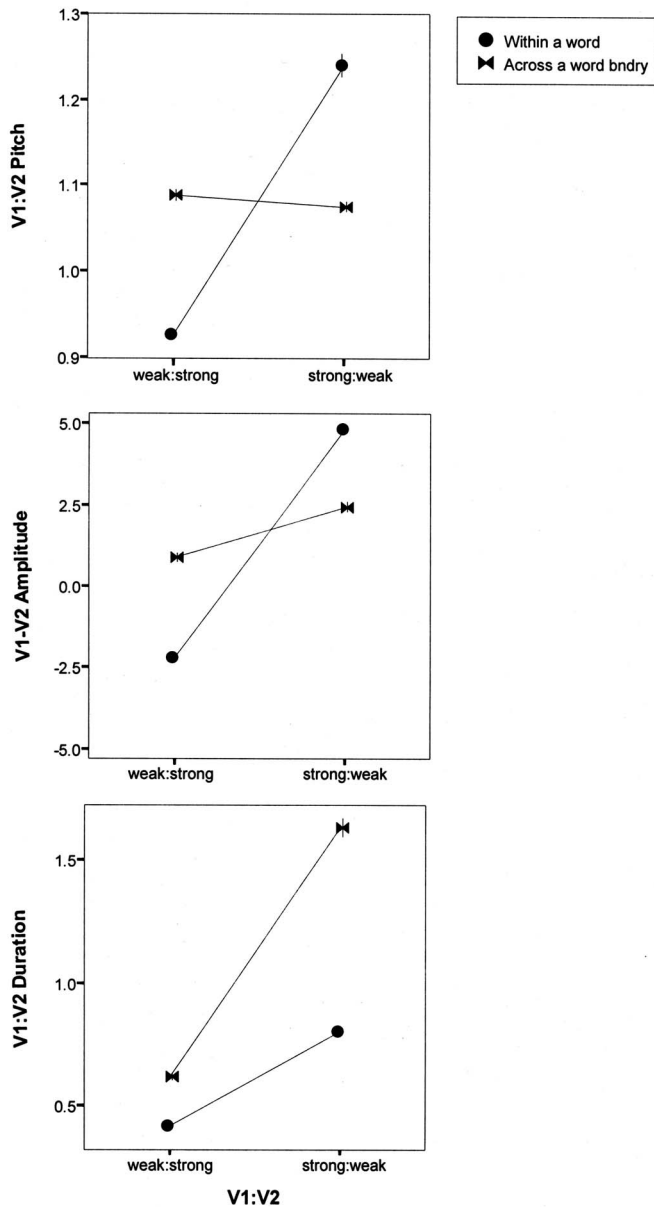


FIG. 1. The relative pitch, amplitude, and duration is shown for the vowels on either side of the consonant sequence that was the focus of the present experiment. The three panels of the figure show from top to bottom the f_0 ratio, the dB difference, and the duration ratio for the pre- (V1) and postconsonantal (V2) vowels as a function of their word position and stress specifications (V1:V2).

2. Consonantal durations

The analyses showed that the effects of stress and sequencing on consonantal duration varied with word position and was different for the different consonantal articulations under consideration. The results on stop closure, stop aspiration, and liquid duration are described in turn with respect to the hypotheses.

Closure duration varied systematically with stress and sequencing [$F(1,143)=83.99, p<0.01, \eta_p^2=0.37$], but the simple effects of stress and sequencing were not statistically significant. Figure 2 shows that postvocalic stop closure is longer when the stop follows a stressed vowel than when it follows an unstressed vowel. Similarly, prevocalic stop closure is longer when the stop precedes a stressed vowel than an unstressed vowel. Thus, the pattern of interaction between stress and sequencing suggests that stop closure duration is influenced only by adjacent vowel strength. Stress does not influence closure duration when a liquid consonant intervenes between the stop and the vowel.

Figure 2 also shows that the pattern of interaction between stress and sequencing varied with word position [$F(1,143)=8.24, p<0.01, \eta_p^2=0.06$]. In particular, stress had a greater influence on closure duration at the edges of words than within a word (i.e., the interaction between stress and sequencing was even stronger at the edges of words). This result is consistent with the juncture hypothesis, which predicts that consonantal duration patterns are amplified at a word boundary. The result also parallels the finding that stress-induced differences in vowel duration were larger across a word boundary than within a word.

Aspiration duration also varied systematically with stress and sequencing [$F(1,143)=26.01, p<0.01, \eta_p^2=0.15$], but the simple effects were also significant [stress, $F(1,143)=208.86, p<0.01, \eta_p^2=0.59$; sequencing, $F(1,143)=307.4, p<0.01, \eta_p^2=0.68$], indicating that the pattern of interaction between the factors was different than that observed for closure duration. In particular, Fig. 3 shows that aspiration duration was influenced by the strength of the following vowel, rather than by adjacent vowel strength. Aspiration duration was longer before a stressed vowel than before an unstressed vowel regardless of whether the stop was first or second in the consonant sequence.

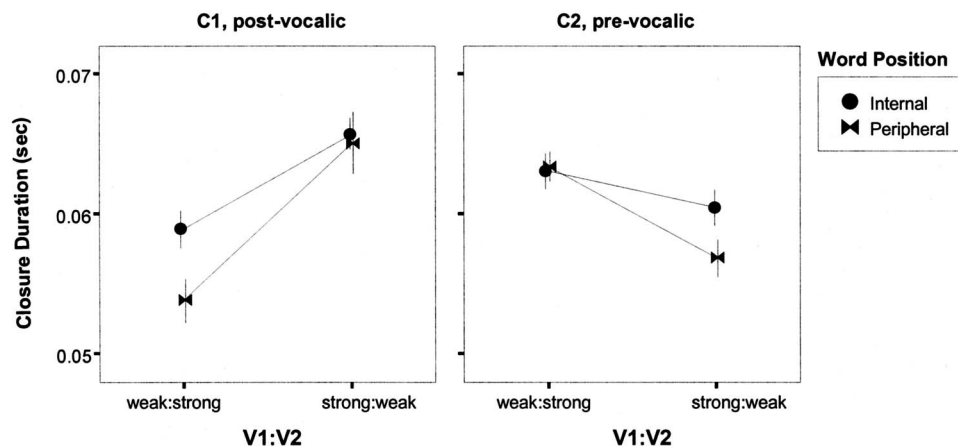


FIG. 2. Stop closure duration is shown as a function of sequence order (C1 or C2), word position (internal or peripheral), and stress specification of the adjacent pre- and postconsonantal vowels (weak:strong or strong:weak).

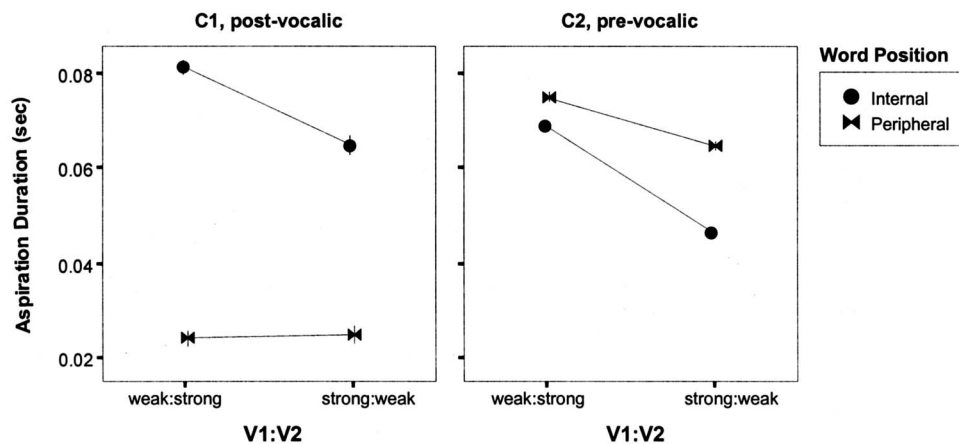


FIG. 3. Stop aspiration duration is shown as a function of sequence order (C1 or C2), word position (internal or peripheral), and stress specification of the adjacent pre- and postconsonantal vowels (weak:strong or strong:weak).

The effect of sequence cannot be discussed without describing the interaction between sequencing and position that is also evident in Fig. 3 [$F(1,143)=500.04$, $p < 0.01$, $\eta_p^2 = 0.78$]. Word-internally, the effect of sequencing was consistent with the prediction from coarticulation: aspiration duration was greater when it preceded a liquid (C1, postvocalic) than when it preceded a vowel (C2, prevocalic). At the word boundary, however, only prevocalic (C2, word-initial) stops were significantly aspirated. Postvocalic (C1, word-final) stops were produced with little to no aspiration. Prevocalic stops also had longer aspiration durations overall at a word boundary than they did word-internally.

The observed interaction between sequence and position is consistent with the juncture hypothesis, but not with the syllable frame hypothesis. In particular, the syllable frame hypothesis predicts aspiration in syllable-onset position and no aspiration in syllable-offset position, consistent with position-dependent allophonic variation of stop aspiration in English. This prediction was clearly upheld at the word boundary. However, word-internal stops were aspirated before a liquid regardless of stress even though the strong-weak stress manipulation was expected to result in offset-onset stop-liquid sequences. Nonetheless, word-internal syllable boundaries are difficult to establish according to independent criteria. So, the word-internal pattern could indicate that speakers syllabified word-internal stop-liquid sequences as onset clusters to the second syllable regardless of lexical stress. If this is the case, then the results can be interpreted to

support a syllable frame hypothesis. If it is not the case, then the results are only consistent with the juncture hypothesis.

The problem of distinguishing between the syllable frame and the juncture hypotheses is further exacerbated by the results on liquid duration. Liquid duration varied with stress [$F(1,143)=20.16$, $p < 0.01$, $\eta_p^2 = 0.12$], sequencing [$F(1,143)=27.39$, $p < 0.01$, $\eta_p^2 = 0.16$], and position [$F(1,143)=291.02$, $p < 0.01$, $\eta_p^2 = 0.67$]. All interactions between these factors were also significant [stress \times sequencing, $F(1,143)=222.12$, $p < 0.01$, $\eta_p^2 = 0.61$; stress \times position, $F(1,143)=29.91$, $p < 0.01$, $\eta_p^2 = 0.17$; sequencing \times position, $F(1,143)=69.21$, $p < 0.01$, $\eta_p^2 = 0.33$; stress \times sequencing \times position, $F(1,143)=11.97$, $p < 0.01$, $\eta_p^2 = 0.08$]. This complex pattern of results is best understood with reference to Fig. 4.

Figure 4 shows two principal results: (1) liquid duration is greater at a word boundary than it is word-internally and (2) adjacent vowel stress affects liquid duration except when liquids follow a stop word-internally. These results are consistent with the juncture hypothesis, which predicts boundary highlighting and unimpeded coarticulation of word-internal stop-liquid sequences. The word-internal result may also be consistent with the syllable frame hypothesis, but only if syllable boundary location is unaffected by lexical stress. In particular, short prevocalic liquids are expected if the word-internal stop-liquid sequence forms an onset cluster to the second vowel (Haggard, 1973).

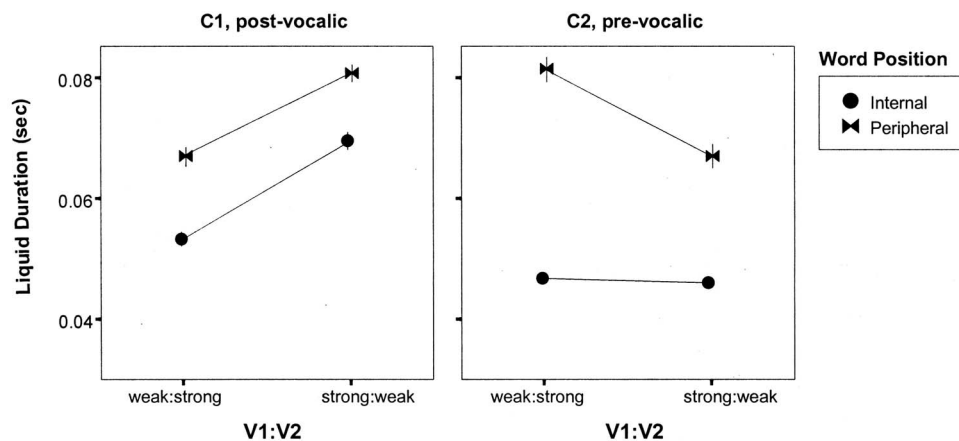


FIG. 4. Liquid duration is shown as a function of sequence order (C1 or C2), word position (internal or peripheral), and stress specification of the adjacent pre- and postconsonantal vowels (weak:strong or strong:weak).

C. Discussion of stress, sequencing, and juncture effects on segmental duration

Overall, the results show that segmental duration patterns within a word differ from the patterns that occur across words, and that stress and sequencing may be sufficient for explaining word-internal consonantal duration patterns. The duration ratio that captures the flanking weak-strong and strong-weak vowel patterns is greater across a word boundary than within a word. Closure duration was also affected by stress and boundary strength: adjacent vowel stress affected stop closure duration, but the relative increase or decrease was greater across a word boundary than within a word. These findings support the juncture hypothesis, which predicts durational patterns that highlight word boundaries. The hypothesis also receives some support from the results on stop aspiration and liquid duration. While postvocalic stop aspiration was nearly absent in word-final position, prevocalic stop aspiration was greater at a word boundary than within a word. Liquid duration was longer overall at a word boundary than word-internally, suggesting boundary-dependent lengthening and within unit coarticulation.

The latter two results on stop aspiration and liquid duration are also consistent with the syllable frame hypothesis, but only if syllable boundaries are defined in a post hoc fashion. In particular, the effect of sequence on word-internal stop aspiration and liquid duration suggests that stop-liquid sequences were always coarticulated with the following vowel. Coarticulation of the stop-liquid sequence with the following vowel is consistent with an onset cluster syllabification. However, in the strong-weak condition this syllabification is at odds with the psycholinguistic literature, which indicates that listeners would syllabify the sequence as an offset-onset when the preconsonantal vowel is stressed (Treiman and Danis, 1988; Treiman and Zukowski, 1990; Derwing and Nearey, 1991; Redford and Randall, 2005). Nonetheless, it is possible that listener boundary perception is irrelevant to production. If this is the case, then the results on word-internal patterns cannot be used to disconfirm the syllable frame hypothesis.

On the other hand, the possibility that syllable structure is better defined in speech according to output patterns rather than according to independent criteria (e.g., listener judgments) suggests that the syllable may not define articulatory dynamics so much as emerge from them. Such a view of syllables is consistent with the juncture hypothesis, which presumes that the word-internal consonantal duration patterns produced by American English speakers reflect basic coarticulatory processes rather than the effects of a linguistic boundary. This possibility is considered further below.

III. COMPARISONS OF ENGLISH CONSONANTAL DURATION PATTERNS WITH RUSSIAN AND FINNISH PATTERNS

Whether or not we assume that the stress manipulation resulted in two different syllabifications of the word-medial stop-liquid sequence, the acoustic results indicate that word-internal patterns in American English can be explained to result from sequencing and the implementation of stress. In

particular, stop aspiration may be greater before a liquid than before a vowel for aerodynamic reasons; namely, coarticulation with the following liquid will mean greater closure in the vocal tract during the stop release gesture, which may make it difficult to re-establish the transglottal pressure differential needed for voicing (Docherty, 1992). Liquids that occur after stop consonants have shorter acoustic duration for a similar reason; namely, because the onset of voicing is deferred after a stop consonant. In other words, the dynamics of articulatory sequencing create different consonantal duration patterns word-internally. Further differences are induced by adjacent vowel strength in the case of stop closure duration and liquid duration, and with subsequent vowel strength in the case of stop aspiration. Stressed vowels are hyperarticulated, that is, they are produced with greater jaw opening and more accurate target attainment (de Jong, Beckman, and Edwards, 1993; de Jong, 1995). The longer articulatory trajectories result in vowel lengthening and, it would seem from the present results, in the lengthening of consonantal gestures that are coarticulated with the vowel.

Although stress and sequencing are sufficient to explain the word-internal patterns in isolation, they are not sufficient to explain (1) why the effects of adjacent vowel strength on closure duration are stronger at a word boundary than word-internally or (2) why stops aspiration is nearly absent in postvocalic (word-final) position and especially robust in prevocalic (word-initial) position or (3) why liquid duration is longer overall at a word boundary than it is word-internally. The juncture hypothesis assumes that all of these differences exist to highlight word boundary location because English phonology and morphology interact to make such boundaries opaque: phonotactics are insufficient for establishing boundary location because different types of syllable onsets and offsets are possible (e.g., *an aim versus a name, odd rum versus a drum*); and word beginnings and endings are not consistently marked with a set of recurring sounds or sound strings (cf. *What is your last name? Kakaja tvoja familija?*). It follows from the English case that languages with simpler syllable structure or more synthetic-inflexional structure will have more transparent word boundaries. It is possible that in such languages stress and sequencing will suffice to describe the consonantal duration patterns both word-internally and at word boundaries. This prediction is tested here in Russian and Finnish: two languages that are more highly synthetic and inflected than English, and one of which—Finnish—also has a simpler syllable structure.

A. Methods

Six native Russian speakers—three male and three female—and six native Finnish speakers—one male and five female—participated in the experiment. Russian and Finnish speakers produced the same stimuli as the English speakers, namely, disyllabic stimuli with intervocalic stop-liquid and liquid-stop sequences that occurred either within a word (e.g., *nɛkla* and *nɛlka*) or across a word boundary (e.g.,

mēnek lano and *mēnek lano*). Again, there were eight of each type—half with -kl- or -lk- sequences and half with -tr- or -rt- sequences (see Table I).

The Russian and Finnish stimuli were embedded in frame sentences, *Skazhi___snova* and *Sano___taas*, which means *Say___again* in Russian and Finnish, respectively. As in the English experiment, each of the stimuli was repeated three times for a total of 96 phrases per speaker. Russian and Finnish speakers were presented with the phrase in randomized order on a sheet of paper in normal orthography—Cyrillic for Russian and Roman for Finnish. Normal orthography could be used in place of the IPA because, unlike English, Russian and Finnish orthography is fairly transparent. Russian speakers produced the stimuli with weak-strong lexical stress, a common pattern in Russian. Finnish speakers produced the stimuli with strong-weak lexical stress, the obligatory pattern in Finnish.

The segmentation criteria and measurements were made as before. The principal difference between the sound patterns of the English and the Russian and Finnish stimuli was in the rhotic, which was trilled and sometimes tapped in both Russian and Finnish. The boundary between this segment and an adjacent vowel was easily discovered by an abrupt decrease in the amplitude of the periodic waveform. As before, intrarater reliability was assessed using Cronbach's coefficient alpha computed over 10% of the data (randomly selected and remeasured). In all cases, the coefficient alpha was high [Russian: stop closure, $\alpha=0.90$; aspiration, $\alpha=0.96$; liquid duration, $\alpha=0.92$; Finnish: stop closure, $\alpha=0.98$; aspiration, $\alpha=0.93$; liquid duration, $\alpha=0.92$], indicating excellent internal measurement consistency.

Once again the design was within-subjects so the effects of the within subjects factors (sequence order and/or position) on the vowel measures and consonantal durations were assessed in one- and two-way repeated measure ANOVAs. Instead of treating Russian and Finnish as a between-subjects factor, we analyzed the data from the different languages separately. The separate language analyses were in keeping with the English-only analyses reported in the first experiment (II above). A more fundamental reason for splitting the data, though, was the fact that decades of research on language-specific phonetics give us every reason to expect language-specific differences in phonetic detail, such as segmental duration. Thus, a qualitative comparison of the durational patterns across languages was deemed more meaningful than a statistical demonstration of difference. The results from the analyses are reported below.

B. Results

1. Stress

As expected, the relative pitch, amplitude, and duration of vowels varied with lexical stress placement. This effect was manifest as pitch, amplitude, and duration differences in the vowels that flanked the consonantal sequence. In Russian, the f_0 ratios, amplitude differences, and duration ratios of the pre-consonantal to postconsonantal vowel were higher across a word boundary than they were within a word [f_0 ratio, $F(1,287)=34.71$, $p<0.01$, $\eta_p^2=0.11$; dB difference,

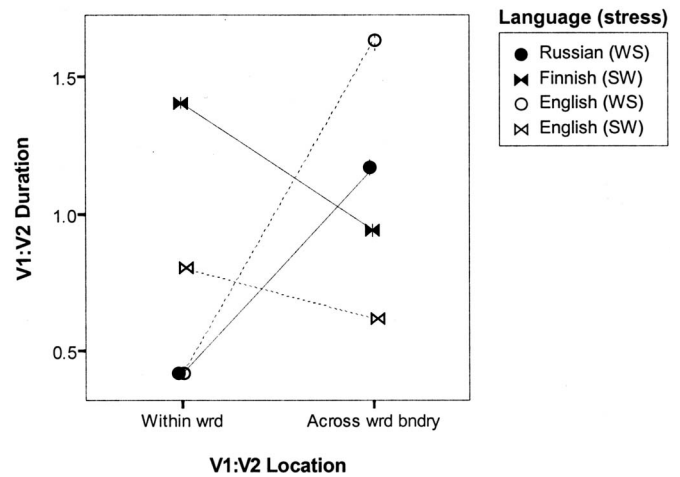


FIG. 5. The relative duration of the pre- and postconsonantal vowels is shown for the different languages and word positions. Note that the stress specification of these vowels changes across a word boundary within a language (e.g., #WS# is compared with #WS#WS#). The American English patterns are shown for comparison. These patterns are the same as those displayed in Fig. 1, but in Fig. 1 the duration ratios were displayed as a function of the stress specification for the flanking vowels (e.g., #WS# was compared with #SW#SW#).

$F(1,287)=65.59$, $p<0.01$, $\eta_p^2=0.19$; duration ratio, $F(1,287)=734.35$, $p<0.01$, $\eta_p^2=0.72$]. This is because the pre-consonantal to postconsonantal vowel pattern in Russian was weak-strong within a word and strong-weak across a word boundary. In Finnish, the vowels that flanked the consonantal sequence had the opposite pattern: strong-weak within a word and weak-strong across a word boundary. Accordingly, the f_0 ratios, amplitude differences, and duration ratios were higher within a word than across a word boundary in Finnish [f_0 ratio, $1,287F(1,287)=113.88$, $p<0.01$, $\eta_p^2=0.28$; dB difference, $F(1,287)=423.99$, $p<0.01$, $\eta_p^2=0.60$; $F(1,287)=221.02$, $p<0.01$, $\eta_p^2=0.44$]. The Russian and Finnish results on the vowel-to-vowel duration ratio are shown as a function of vowel location in Fig. 5. The American English results from the weak-strong and strong-weak stress conditions are included for comparison.

Again, it is important to note that the pitch, amplitude, and duration patterns are different within a word than they are between words. This is because stress was always placed on the second syllable in Russian and on the first syllable in Finnish. When these stress patterns are realized on sequential disyllabic words, the result is a shift in stress across a word boundary. For example, Russian *nēklá* with weak-strong stress has a weak V1 and a strong V2, but the same vowels in Russian *mēnēk lanó* are strong-weak (WS#WS). This stress shift is also apparent in the American English data, when these are graphed by stress condition rather than by the stress pattern associated with the vowels on either side of the focal consonant sequence (viz. Figs. 1–4).

2. Consonantal durations

The stress shift across a word boundary means that the effects of adjacent vowel stress on consonantal durations in Russian and Finnish will be shown by an interaction between sequencing and word position. This type of interaction was

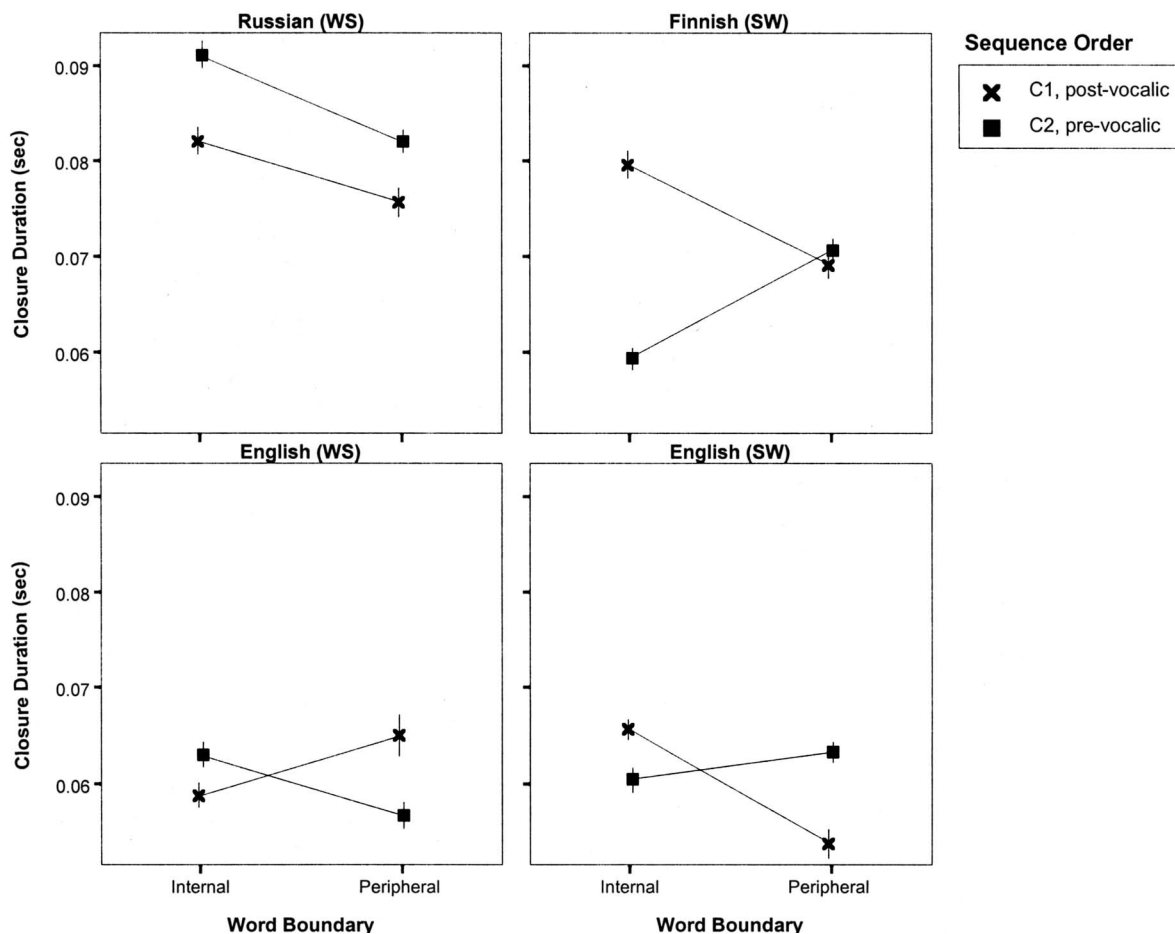


FIG. 6. Stop closure duration is shown as a function of language (stress pattern), sequence order (C1 or C2), and word position (internal or peripheral). The American English data from Fig. 2 are redrawn here to facilitate comparison with the Russian and Finnish data.

statistically significant for stop closure duration in Finnish [$F(1,143)=92.75$, $p < 0.01$, $\eta_p^2=0.39$] and for liquid duration in both Russian and Finnish [Russian, $F(1,143)=20.46$, $p < 0.01$, $\eta_p^2=0.13$; Finnish, $F(1,143)=85.22$, $p < 0.01$, $\eta_p^2=0.37$]. Note that the American English data show a similar effect of stress on stop closure duration in both stress conditions and on liquid duration in the strong-weak stress condition. The results on stop closure duration in Russian and Finnish are shown in Fig. 6 and those on liquid duration are shown in Fig. 7. The English data are included for comparison.

Figures 6 and 7 are also instructive because they show that sequencing explains stop closure duration in Russian, even when stress does not; and that sequencing is generally an important factor in explaining stop closure and liquid duration across both languages [Russian stop closure, $F(1,143)=21.46$, $p < 0.01$, $\eta_p^2=0.13$; Finnish stop closure, $F(1,143)=49.74$, $p < 0.01$, $\eta_p^2=0.26$; Russian liquid, $F(1,143)=46.50$, $p < 0.01$, $\eta_p^2=0.25$; Finnish liquid, $F(1,143)=131.34$, $p < 0.01$, $\eta_p^2=0.48$]. In particular, prevocalic stop closure (C2) is longer than postvocalic stop closure (C1) in Russian regardless of word position, and postvocalic liquids (C1) are longer than prevocalic liquids (C2) regardless of position.

Sequencing also influences liquid duration in American English, but the effect of word position is even more obvious

(see Figs. 4 and 7). Word-internal liquids are short, except when they occur after a stressed vowel. Liquids at a word boundary are long, though influenced by the strength of an adjacent vowel.

Finally, Figs. 6 and 7 show that the effect of word position in Russian and Finnish is different from their effect in English. In English, sequencing effects are amplified at a word boundary, that is, the difference between postvocalic and prevocalic stop closure is greater at a word boundary than word-internally. In contrast, position induced differences in closure duration are smaller at a word boundary in Russian and Finnish than they are within a word.

Finnish stops and liquids show the opposite pattern of Russian liquids, but this pattern is nonetheless quite different from English. In Russian, liquid duration is more influenced by adjacent vowel stress at a word boundary than word-internally, but the interaction with sequence order negates a simple effect of word position on duration [$F(1,143)=2.95$, NS]. This means that there is no net lengthening effect for liquids at a word boundary in Russian as there is in English. In Finnish, the effects of stress are stronger word-internally than they are at a word boundary. This negates a simple effect of word position on duration for stop closure [$F(1,143)=0.13$, NS], though the effect on liquid duration reaches significance—on average liquids are longer at a word boundary [$F(1,143)=4.04$, $p < 0.05$, $\eta_p^2=0.03$]. Although liq-

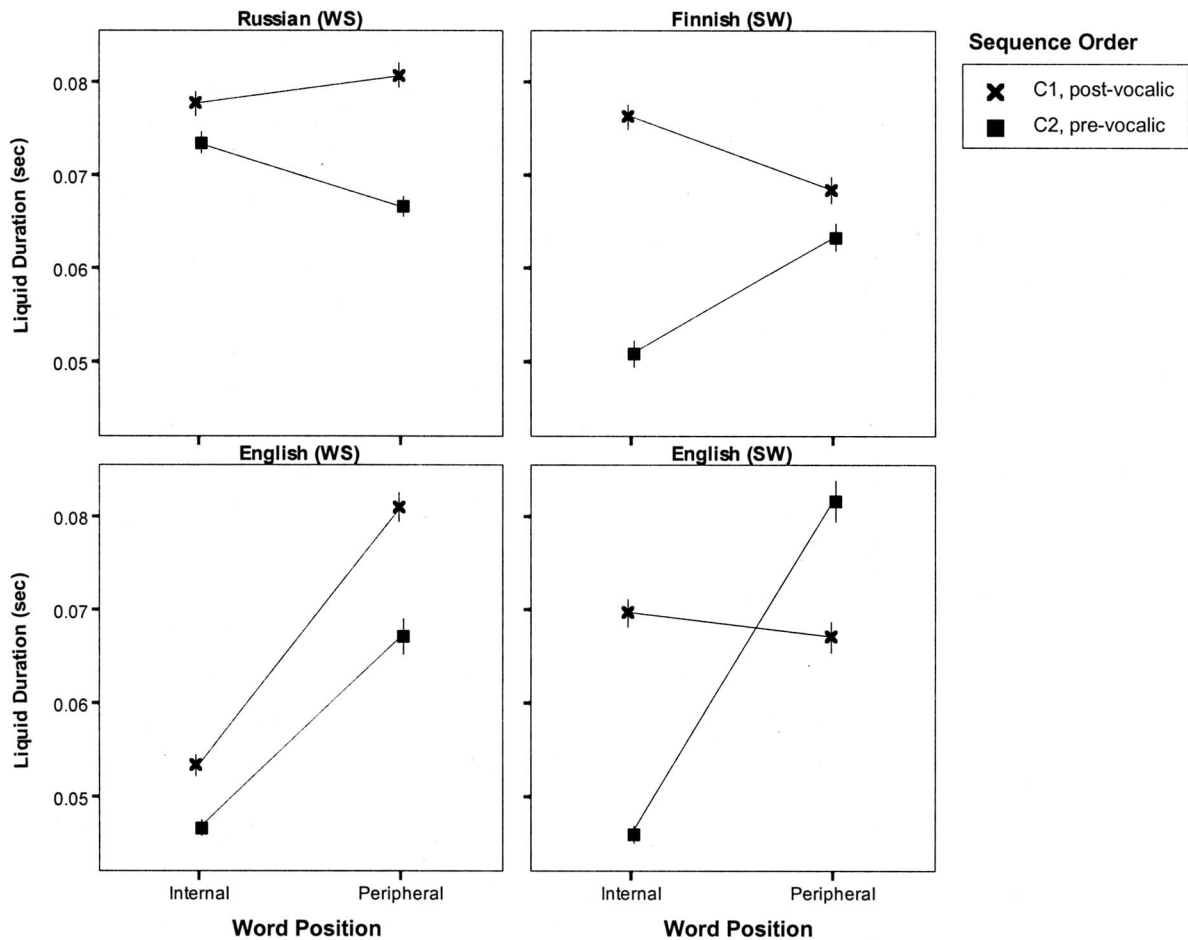


FIG. 7. Liquid duration is shown as a function of language (stress pattern), sequence order (C1 or C2), and word position (internal or peripheral). The American English data from Fig. 4 are redrawn here to facilitate comparison with the Russian and Finnish data.

uids may be longer on average at a word boundary in Finnish, examination of Figs. 6 and 7 leaves the impression that consonantal duration patterns are amplified word-internally in Finnish, which is opposite of the English results where such patterns are amplified at a word boundary. However, the Finnish pattern parallels the greater word-internal differences in vowel duration, and this is similar to the English pattern: stop closure and liquid duration differences are also greater in English where vowel duration differences are greater (albeit across a word boundary rather than word-internally as in Finnish).

The minimal effect of word boundaries in Russian and Finnish is most apparent in the results on stop aspiration duration. Figure 8 shows that Russian stop aspiration is unaffected by word position, and minimally affected by adjacent vowel stress (i.e., there is a nonsignificant trend towards an interaction between sequence and position [$F(1,143) = 3.65$, $p = 0.058$, $\eta_p^2 = 0.03$]). Finnish stop aspiration is affected by sequence [$F(1,143) = 198.62$, $p < 0.01$, $\eta_p^2 = 0.58$] and by word position [$F(1,143) = 13.18$, $p < 0.01$, $\eta_p^2 = 0.08$], which likely reflects the strength of the subsequent vowel as it does word-internally in English.

Figure 8 shows that the effect of sequencing in Finnish is the same within a word as it is across a word boundary. Stops that occur before liquids are produced with more aspiration than stops that occur before vowels. This result differs

from English, where only word-internal stops are longer before liquids. At the word's edge, English stop aspiration is determined by whether the stop occurs word-finally or word-initially.

C. Discussion of the comparison between Russian, Finnish, and English

A comparison of Russian, Finnish, and American English consonantal duration patterns shows that stress and sequencing affect duration in similar ways across languages, though there are language-specific differences in the extent to which any one factor explains the durational patterns. For instance, vowel stress has a lengthening effect on adjacent consonants in both Russian and Finnish, but stress affects consonantal duration less in Russian than in Finnish.

Whereas the effect of stress and sequencing is similar across languages, the effect of word position is different. In English, differences in stop closure duration and aspiration are especially strong at a word boundary, and liquids are especially long. The opposite pattern is found in Russian and Finnish. In Russian, stop closure duration is longer within a word than at a word boundary, and there is no effect of word position on aspiration. In Finnish, the differences in stop closure and liquid duration are especially strong within a word compared to at a word boundary.

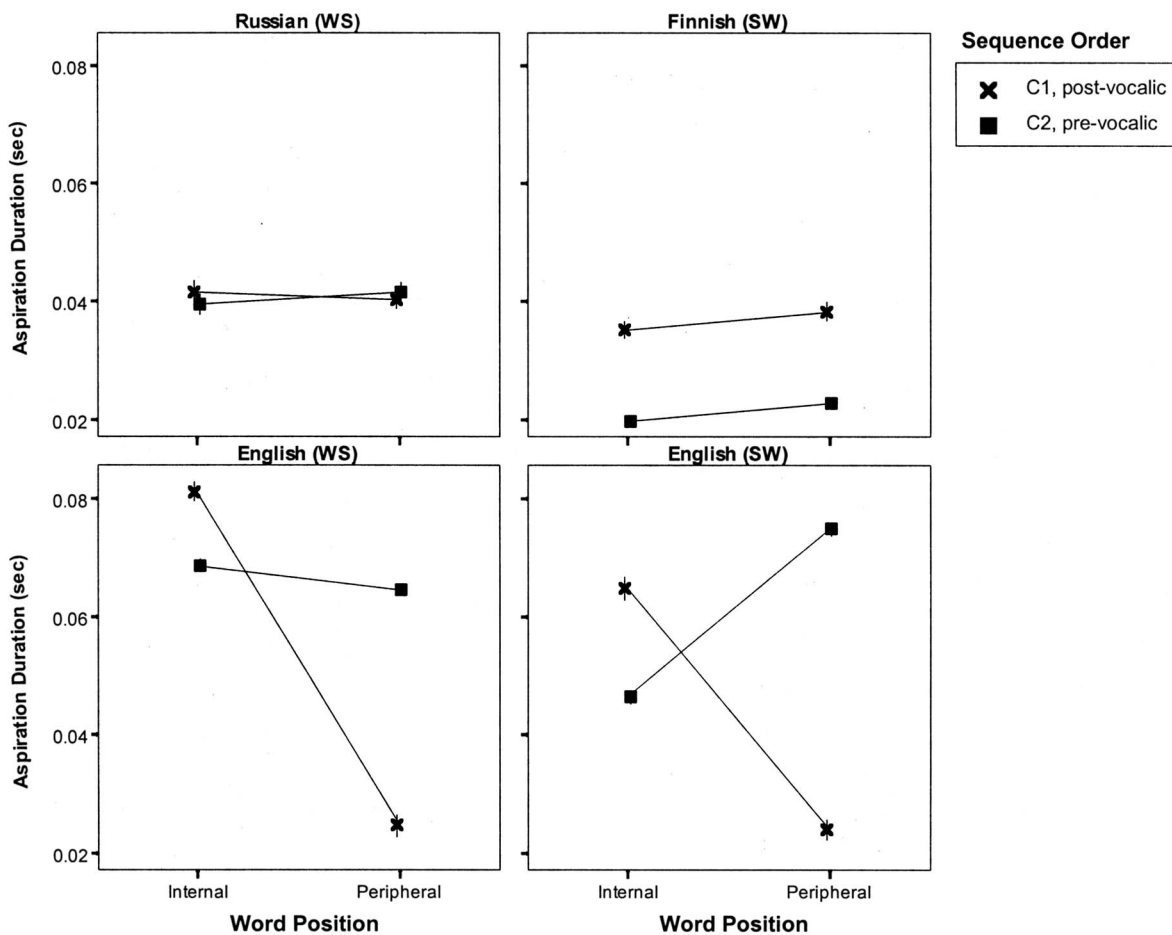


FIG. 8. Stop aspiration duration is shown as a function of language (stress pattern), sequence order (C1 or C2), and word position (internal or peripheral). The American English data from Fig. 3 are redrawn here to facilitate comparison with the Russian and Finnish data.

Overall, the different effects of word position in English on the one hand and Russian and Finnish on the other support the juncture hypothesis for English. The results also suggest language-specific differences in how sequential action is organized for output. If the boundaries of planning units are signaled by robust durational patterns because longer consonantal durations indicate less articulatory overlap, then the word-internal consonantal duration patterns in Russian and Finnish suggest that a sublexical unit may be important for organizing action in these languages. Further research is needed to determine whether the Russian unit is an articulatory syllable, as defined by Kozhevnikov and Chistovich (1965), and the Finnish unit is the mora, as suggested by Suomi's work (Suomi and Ylitalo, 2004; Suomi, 2005).

IV. CONCLUSION

This study contributes to a long tradition of research that uses segmental duration patterns to understand what supra-segmental units are involved in motor speech planning. Since the beginning of this tradition, the linguistic syllable has been suggested as a basic unit in the plan. Here, this suggestion was restated in terms of a generic syllable frame hypothesis, which was then investigated with two problems in mind. The principal problem with the hypothesis is that

word-internal syllable boundaries are difficult to define for particular sequences. Accordingly, the boundaries of the unit can either be defined post hoc and with respect to observed durational patterns (e.g., Kozhevnikov and Chistovich, 1965) or they can be defined a priori and with respect to word boundaries (e.g., word-final equals syllable-final), thereby confounding the role of the syllable and word in speech. In cases where word-internal syllable boundaries are less disputed (e.g., liquid-stop sequences), there is a secondary problem with the hypothesis; namely, that segmental duration is affected by intrinsic duration, coarticulatory dynamics, and the rhythm structure of a language. This means that durational patterns attributed to the syllable in English may be better understood as emergent from articulatory processes having to do with the realization of stress and particular segmental sequencing patterns.

In the current study, the syllable frame hypothesis would have been clearly supported and alternate explanations clearly refuted in the American English data if the stress manipulation on disyllabic words with medial stop-liquid sequences had resulted in distinct durational patterns: one that was consistent with an onset cluster pattern when the disyllabic word received weak-strong stress; and one that mirrored the word-final/word-initial pattern when the disyllabic word received strong-weak stress. Instead, the results on this

class of words only support the syllable frame hypothesis if word-internal syllable structure is redefined to fit the durational data.

In the absence of clear support for the syllable frame hypothesis, the results were explained with reference to stress, sequencing, and the communicative function of language. The remainder of this section outlines a production model proposed to integrate these factors into an explanation for positional effects on consonantal duration—one that heeds the emergence of cross-language differences in durational patterns.

The proposed model assumes that the rhythm and sound patterns of a linguistic unit are linked in production via jaw movement. Such an assumption follows from the work on prosodic structure and jaw movement (Stone, 1981; de Jong, Beckman, and Edwards, 1993; Harrington, Fletcher, and Roberts, 1995; Erickson, 1998; Erickson, Fujimura, and Pardo, 1998) as well as from the work on sound sequencing and jaw movement (MacNeilage, 1998; MacNeilage and Davis, 2000; MacNeilage *et al.*, 2000). Specifically, the linguistic unit specifies through its rhythmic structure an abstract open-close pattern to be used as an articulatory frame within which the open (vowel) and closed (consonants) segments of a sequence can be located in space and time. The rhythmic structure of the unit dictates both the number of open-close movements and movement amplitude. Assuming a constant speech rate, cycle number and amplitude will have temporal consequences on segmental articulation. In the case of a stress-timed language, the larger amplitude cycles associated with a stressed beat will allow more time for target attainment, resulting in longer segmental durations. Since the lengthening effects apply only within the cycle, only those gestures that are executed during the downward or upwards trajectory of the cycle will have increased durations. For this reason, stress may affect only the duration of an opening or closing gesture in segments that contain both, such as stop consonants. For instance, whereas closure for postvocalic stops can be initiated during either the upwards or downwards trajectory of the vowel, an audible stop release cannot occur unless the jaw is opening. This example also illustrates how stress conditions the stop consonantal duration patterns observed in the present study; in particular, it explains why adjacent vowel stress affects closure duration, but why only subsequent vowel stress explains aspiration duration.

An additional example illustrates how stress and sequencing interact in the proposed model to condition the pattern of short word-internal liquids in stop-liquid sequences observed for American English. A liquid that follows a stop can be articulated on the downward trajectory of the jaw because they are comprised of two gestures—an anterior and posterior one, which are sequenced differently depending on whether the liquid occurs in pre- or post vocalic position (Sproat and Fujimura, 1993; Gick *et al.*, 2006). Liquid acoustic duration is thus limited on the stop side by difficulties in regaining a transglottal pressure differential, but it is also limited on the vowel side. This is because the liquid is being articulated while the jaw continues in its downward

trajectory and, at a certain point, the open jaw configuration may become inconsistent even with the posterior constriction gesture that precedes the vocalic gesture.

At a higher level of control, the meaningful units of language create structure within the speech plan, which interact with the size-limits of the motor programs (i.e., executable chunks of planned speech action). Future work could determine the maximal size of such programs by investigating long-distance coarticulation, but work on vowel-to-vowel coarticulation already suggests that they can include more than one cycle (e.g., Öhman, 1966). In light of such work, and the evidence presented in the current study, it is reasonable to assume that the word could be the most basic unit—at least in English. The unit may be different in a more highly synthetic-inflexional language like Finnish, where more than one programming unit may be needed to execute the average word, which is significantly longer and morphologically more complex than the average English word.

Although future work will be aimed at testing and elaborating the proposed production model, the model is useful in that it highlights two broad conclusions drawn from the current study. First, positional effects on consonantal duration may or may not refer to linguistic units, which suggests that more basic articulatory explanations for such patterns should be excluded before linguistic units such as the syllable are invoked. Second, cross-language differences in durational patterns may reflect more than language-specific phonetic tuning: such differences may reflect interesting interactions between the different ways in which languages encode meaning and universal limits on the maximal size of the motor programs in speech motor control.

ACKNOWLEDGMENTS

The author is grateful to Tatiana Furrow and Lewis Notestine for help with subject recruitment and data collection. The author also thanks (in alphabetic order) Dani Byrd, Edward Flemming, Susan Guion, Scott Myers, Kari Suomi, Marija Tabain, and Doug Whalen for their constructive criticisms of earlier versions of this paper.

¹For example, biomechanical dependencies between the articulators may influence action sequencing and thus intergestural timing. This idea is developed further in the conclusion. It is also possible that habitual sequencing patterns may be transformed through extensive practice into syllable-sized speech motor programs. Although such programs could inform judgments on syllable boundaries [e.g., when listeners subvocalize in a segmentation task (e.g., Sato *et al.*, 2006)], they need not be specified in the phonology. Instead, the phonology need only represent abstract language-specific knowledge that emerges at the interface between language production and comprehension, which is the lexicon in some models (e.g., Bybee, 2001).

Boersma, P., and Weenink, D. (2002). "Praat: Doing phonetics by computer (version 4.0.34)" (Computer program). Retrieved November 2002 from <http://www.praat.org/>.

Boucher, V. (1988). "A parameter of syllabification for VstopV and relative timing invariance," *J. Phonetics* 16, 299–326.

Browman, C., and Goldstein, L. (1988). "Some notes on syllable structure in articulatory phonology," *Phonetica* 45, 140–155.

Bybee, J. (2001). *Phonology and Language Use* (Cambridge University Press, Cambridge).

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

- Campbell, W. N. (1992). "Syllable-based segmental duration," in *Talking Machines: Theories, Models, and Designs*, edited by G. Bailly, C. Benoit, and T. R. Sawallis (Elsevier, North Holland) pp. 211–224.
- Campbell, W. N., and Isard, S. D. (1991). "Segment durations in a syllable frame," *J. Phonetics* **19**, 37–47.
- Christie, W. M. (1977). "Some multiple cues for juncture in English," *General Linguistics* **17**, 212–222.
- de Jong, K. J. (2003). "Temporal constraints and characterizing syllable structuring," in *Phonetic Interpretation: Papers in Laboratory Phonology 6*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 253–268.
- de Jong, K. J. (1995). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.* **97**, 491–504.
- de Jong, K., Beckman, M. E., and Edwards, J. (1993). "The interplay between prosodic structure and coarticulation," *Lang Speech* **36**, 197–212.
- Derwing, B. L., and Nearey, T. M. (1991). "The 'vowel-stickiness' phenomenon: Three experimental sources of evidence," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, France (ICPhS, Aix-en-Provence); Vol. **2**, pp. 210–213.
- Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents* (Foris, Berlin).
- Edwards, J., and Beckman, M. E. (1988). "Articulatory timing and the prosodic interpretation of syllable duration," *Phonetica* **45**, 156–174.
- Erickson, D. (1998). "Effects of contrastive emphasis on jaw opening," *Phonetica* **55**, 147–169.
- Erickson, D., Fujimura, O., and Pardo, B. (1998). "Articulatory correlates of emotion control: Emotion and emphasis," *Lang Speech* **41**, 399–417.
- Fougeron, C., and Keating, P. (1997). "Articulatory strengthening at the edges of prosodic domains," *J. Acoust. Soc. Am.* **101**, 3728–3740.
- Fowler, C. A. (1979). "'Perceptual centers' in speech production and perception," *Percept. Psychophys.* **25**, 375–388.
- Gick, B., Campbell, F., Oh, S., and Tamburri-Watt, L. (2006). "Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids," *J. Phonetics* **34**, 49–72.
- Haggard, M. (1973). "Correlations between successive segment durations: values in clusters," *J. Phonetics* **1**, 111–116.
- Harrington, J., Fletcher, J., and Roberts, C. (1995). "Coarticulation and the accented/unaccented distinction: evidence from jaw movement data," *J. Phonetics* **23**, 305–322.
- Kahn, D. (1976). "Syllable-based generalizations in English phonology," Doctoral dissertation, Massachusetts Institute of Technology, Distributed by Indiana University Linguistics Club.
- Keating, P., Cho, T., Fougeron, C., and Hsu, C-S. (2003). "Domain-initial articulatory strengthening in four languages," in *Phonetic Interpretation: Papers in Laboratory Phonology 6*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press; New York) pp. 143–161.
- Keating, P., Wright, R., and Zhang, J. (1999). "Word-level asymmetries in consonant articulation," *University of California Working Papers in Phonetics* **97**, 157–173.
- Klatt, D. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," *J. Acoust. Soc. Am.* **59**, 1208–1221.
- Kohler, K. J. (1991). "The phonetics/phonology issue in the study of articulatory reduction," *Phonetica* **48**, 180–192.
- Kozhevnikov, V., and Chistovich, L. (1965). *Speech: Articulation and Perception* (Joint Publication Research Services, Washington, D.C.)
- Krakow, R. (1999). "Physiological organization of syllables: A review," *J. Phonetics* **27**, 23–54.
- Lehiste, I. (1970). *Suprasegmentals* (MIT Press, Cambridge, MA).
- MacNeilage, P. F. (1998). "The frame/content theory of evolution of speech production," *Behav. Brain Sci.* **21**, 499–546.
- MacNeilage, P. F., and Davis, B. L. (2000). "On the origin of internal structure of word forms," *Science* **288**, 527–531.
- MacNeilage, P. F., Davis, B. L., Kinney, A., and Matyear, C. L. (2000). "The motor core of speech," *Child Dev.* **71**, 153–163.
- Maddieson, I. (1985). "Phonetic cues to syllabification," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, New York), pp. 203–221.
- Maddieson, I. (2004). "Timing and alignment: A case study of Lai," *Language and Linguistics* **5**, 729–755.
- Moon, S.-J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* **96**, 27–56.
- Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). "Compensatory shortening' in monosyllables of spoken English," *J. Phonetics* **20**, 225–239.
- Öhman, S. E. G. (1966). "Coarticulation in VCV utterances: Spectrographic measurements," *J. Acoust. Soc. Am.* **39**, 151–168.
- Prieto, M. (2002). "Acoustic correlates of the syllable: evidence from Spanish," Doctoral dissertation, University of Illinois at Urbana-Champaign, *Dissertation Abstracts International* **63**, 580.
- Quené, H. (1992). "Durational cues for word segmentation in Dutch," *J. Phonetics* **20**, 331–350.
- Redford, M. A., and Diehl, R. L. (1999). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," *J. Acoust. Soc. Am.* **106**, 1555–1565.
- Redford, M. A., and Gildersleeve-Neumann, C. E. (2006). "The acquisition of two phonetic cues to word boundaries" (unpublished).
- Redford, M. A., and Randall, P. (2005). "The role of juncture cues and phonological knowledge in English syllabification judgments," *J. Phonetics* **33**, 27–46.
- Sato, M., Schwartz, J.-L., Abry, C., Cathiard, M.-A., and Loevenbruck, H. (2006). "Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect," *Percept. Psychophys.* **68**, 458–474.
- Selkirk, E. (1982). "The syllable," in *The Structure of Phonological Representation*, edited by H. van der Hulst and N. Smith (Foris, Dordrecht), pp. 337–383.
- Sproat, R., and Fujimura, O. (1993). "Allophonic variation in English/l/ and its implications for phonetic implementation," *J. Phonetics* **21**, 291–311.
- Stone, M. (1981). "Evidence for a rhythm pattern in speech production: observations of jaw movement," *J. Phonetics* **9**, 109–120.
- Suomi, K. (2005). "Temporal conspiracies for a tonal end: Segmental durations and accentual f0 movement in a quantity language," *J. Phonetics* **33**, 291–309.
- Suomi, K., and Ylitalo, R. (2004). "On durational correlates of word stress in Finnish," *J. Phonetics* **32**, 35–65.
- Tabain, M. (2003). "Effects of prosodic boundary on /aC/ sequences: Acoustic results," *J. Acoust. Soc. Am.* **113**, 516–531.
- Treiman, R., and Zukowski, A. (1990). "Towards an understanding of English syllabification," *J. Mem. Lang.* **29**, 66–85.
- Treiman, R., and Danis, C. (1988). "Syllabification of intervocalic consonants," *J. Mem. Lang.* **27**, 87–104.
- Tuller, B., and Kelso, J. A. S. (1991). "The production and perception of syllable structure," *J. Speech Hear. Res.* **34**, 501–508.
- Turk, A. (1994). "Articulatory phonetic clues to syllable affiliation: gestural characteristics of bilabial stops, in *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3*, edited by P. Keating (Cambridge University Press, New York).
- Umeda, N. (1975). "Vowel duration in American English," *J. Acoust. Soc. Am.* **58**, 434–445.
- Umeda, N. (1977). "Consonant duration in American English," *J. Acoust. Soc. Am.* **61**, 846–858.
- van Santen, J. P. H., and Shih, C. (2000). "Suprasegmental and segmental timing models in Mandarin Chinese and American English," *J. Acoust. Soc. Am.* **107**, 1012–1026.
- van Sluiter, A. M. C., and van Heuven, V. J. (1996). "Spectral balance as an acoustic correlate of linguistic stress," *J. Acoust. Soc. Am.* **100**, 2471–2485.

Noise estimation in voice signals using short-term cepstral analysis

Peter J. Murphy and Olatunji O. Akande

Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland

(Received 13 April 2005; revised 19 October 2006; accepted 4 December 2006)

Cepstral-based estimation is used to provide a baseline estimate of the noise level in the logarithmic spectrum for voiced speech. A theoretical description of cepstral processing of voiced speech containing aspiration noise, together with supporting empirical data, is provided in order to illustrate the nature of the noise baseline estimation process. Taking the Fourier transform of the liftered (filtered in the cepstral domain) cepstrum produces a noise baseline estimate. It is shown that Fourier transforming the low-pass liftered cepstrum is comparable to applying a moving average (MA) filter to the logarithmic spectrum and hence the baseline receives contributions from the glottal source excited vocal tract and the noise excited vocal tract. Because the estimation process resembles the action of a MA filter, the resulting noise baseline is determined by the harmonic resolution (as determined by the temporal analysis window length) and the glottal source spectral tilt. On selecting an appropriate temporal analysis window length the estimated baseline is shown to lie halfway between the glottal excited vocal tract and the noise excited vocal tract. This information is employed in a new harmonics-to-noise (HNR) estimation technique, which is shown to provide accurate HNR estimates when tested on synthetically generated voice signals. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2427123]

PACS number(s): 43.70.Gr, 43.70.Dn [AL]

Pages: 1679–1690

I. INTRODUCTION

The estimation of the harmonics-to-noise ratio (HNR) is shown to provide a useful index for noninvasive assessment of voice pathology (cf. Yumoto *et al.*, 1982, Kasuya *et al.*, 1986) and hence a reliable method for estimating the HNR in human voice is important for effective evaluation and management of voice disorders. HNR is defined as the ratio between the energy of the periodic component to the energy of the aperiodic component in the signal. As such it is sensitive to all forms of waveform aperiodicity. It only specifically reflects a signal to aspiration noise ratio when other aperiodicities in the signal are comparatively low.

A number of methods for HNR estimation in voiced speech have been proposed. The different methodologies employed can be broadly classified as time (cf. Yumoto *et al.*, 1982), frequency [cf. Kasuya *et al.* (1986), which provides an inverted HNR estimate], and cepstral (cf. de Krom, 1993) domain techniques. Some of these methods, however, are not validated for accuracy and instead represent a general HNR trend, i.e., they show a decrease or increase in HNR values with a corresponding decrease or increase in input noise values. Validation of a HNR method requires testing the technique against synthesis data with *a priori* knowledge of the HNR.

Time-domain methods that require individual period detection for HNR estimation can be problematic because of the difficulty in estimating the period markers for (pathological) voiced speech (cf. Yumoto *et al.*, 1982). Frequency-domain methods encounter the problem of estimating noise at harmonic locations (cf. Kasuya *et al.*, 1986). Cepstral techniques have been introduced¹ to supply noise estimates at all frequency locations in the spectrum (the cepstral pro-

cessing removes the harmonics from the spectrum). In the present study, two existing cepstral-based HNR estimation techniques (de Krom, 1993; Qi and Hillman, 1997) are briefly reviewed (Sec. II). The review highlights the fact that an unprocessed noise baseline estimated via cepstral processing does not approximate the actual noise floor and that a formal description of how the baseline estimate is produced has not been addressed. Hence a theoretical description of cepstral processing of voiced speech containing aspiration noise, together with supporting empirical data, is provided in order to illustrate the precise nature of the noise baseline estimation procedure (Sec. III). It is shown that the cepstrum-based noise baseline estimation process is comparable to applying a moving average (MA) filter to the power spectrum and hence the baseline receives contributions from the glottal source excited vocal tract and the noise excited vocal tract.

Before reviewing the two existing cepstrum-based HNR estimation techniques a summary introduction to cepstral analysis of voiced speech is provided (cf. Noll, 1967). The real cepstrum, $\hat{C}(n)$ is defined as the inverse Fourier transform of the log of the magnitude spectrum:

$$\hat{C}(n) = \int_{-\infty}^{\infty} \log|S(f)|e^{j2\pi fn} df. \quad (1)$$

With reference to Fig. 1(a) it is noted that in the logarithmic power spectrum for a segment of voiced speech the vocal tract contributes a slow variation, while the periodicity of the source manifests itself as a fast variation. Taking the inverse Fourier transform of this logarithmic power spectrum yields a prominent peak corresponding to the high frequency source component and a broader peak corresponding to the low fre-

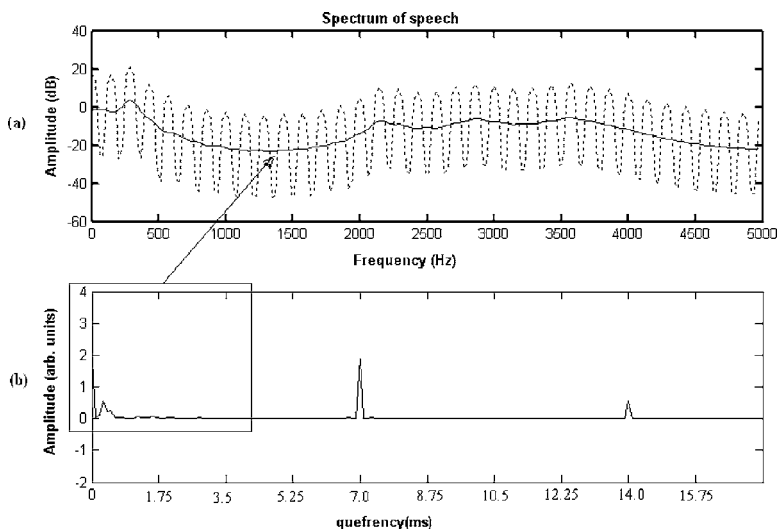


FIG. 1. (a) Spectrum of voiced speech showing original spectrum including harmonic structure and the superimposed Fourier transform of the low-pass filtered cepstrum. (b) Cepstrum of voiced speech, illustrating low quefrequencies due to the vocal tract, and periodically spaced rahmonics due to the harmonic structure.

quency formant structure [Fig. 1(b)]. To distinguish between the frequency components in a temporal waveform and the frequency in the log spectrum the term quefrequency is used to describe the log spectrum “frequency,” while the first prominent cepstral peak is termed the “first rahmonic.” Because a form of deconvolution can be achieved via the cepstrum the technique has been successfully employed to estimate a HNR in voice signals as detailed in the following section.

II. REVIEW OF EXISTING CEPSTRUM-BASED HNR TECHNIQUES

Cepstrum-based estimation of the HNR in speech signals was introduced by de Krom (1993). The essence of the de Krom (1993) technique is that the harmonics can be removed from the spectrum of voiced speech using cepstral processing and hence the noise can be estimated at all frequencies in the spectrum. The rahmonics, representing the prominent peaks at integer multiples of the fundamental period (T) in the cepstrum of voiced speech, are removed through comb-lifering. The resulting comb-liftered cepstrum is Fourier transformed to obtain a noise spectrum (log power spectrum in dB) $N_{ap}(f)$, which is subtracted from the log power spectrum, $O(f)$, of the original signal. This gives a source related log power spectrum, $H_{ap}(f)$. A baseline correc-

tion factor $B_d(f)$, defined as the deviance of harmonic peaks from the 0 dB line, is determined. This factor is subtracted from the estimated noise spectrum to yield a modified noise spectrum. The modified noise spectrum $N(f)$ is now subtracted from the original log-spectrum, in order to estimate the harmonics-to-noise ratio.

Analysis parameters, consisting of a variable temporal window length (1024, 2048, 4096) and a fixed overlap of 256 sample points for each window length are used to compute the Fourier transform of the synthesized vowel /AH/ (1-s duration, sampling frequency 10 kHz). For each temporal analysis window (Hanning), a set of HNRs is calculated. The arithmetic mean of the estimated HNR values is taken as the representative HNR value for a given signal. The technique is illustrated in Fig. 2. From the figure it can be seen that the noise floor estimate tends to underestimate the actual noise floor.

In Qi and Hillman (1997) a modification was made to the de Krom technique. In this implementation, the HNR is computed by calculating a 3200-point discrete Fourier transform for the input signal sampled at 16 kHz. Harmonic peaks are identified using a frequency delayed, peak peaking algorithm. The frequency delay is introduced to ensure that each peak located is global within a given frequency range of the

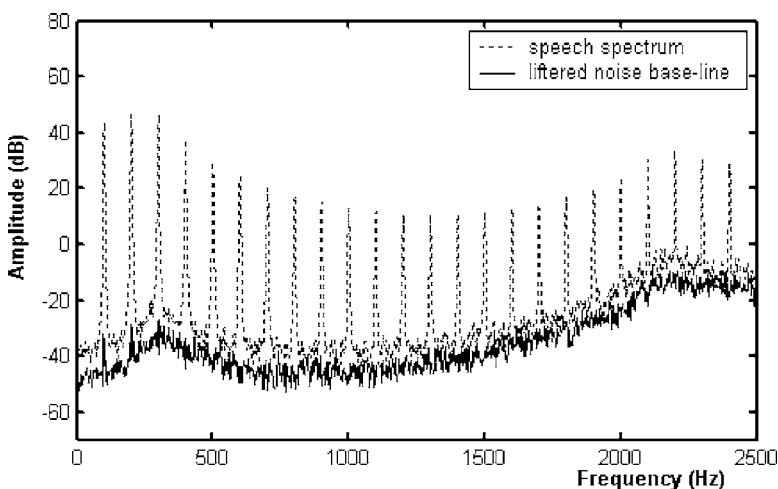


FIG. 2. HNR estimation using de Krom cepstral base-line technique (1024 sample temporal window length shown).

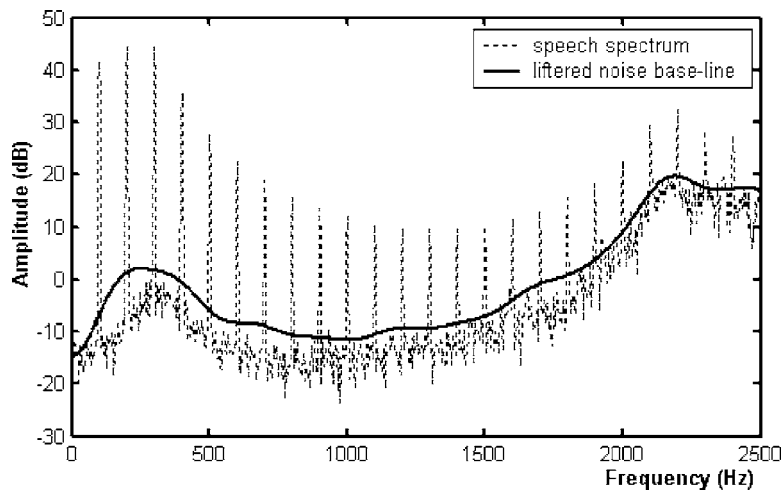


FIG. 3. HNR estimation (Qi and Hillman, 1997) cepstral baseline technique (temporal window length 3200 points).

spectrum. The spectrum of the same signal segment is then computed and a window is applied to the cepstrum to lifter out the high-quefrequency region. Low-pass liftering below the first rahmonic is employed as opposed to comb-liftering the rahmonics. Through Fourier transform of the low-pass liftered cepstrum, a smoothed [as compared to the de Krom (1993) technique] reference noise level is obtained. Finally, HNR is computed as the mean difference between the harmonic peaks and the reference levels of noise at these peak frequencies.² The technique is illustrated in Fig. 3. From the figure it can be seen that, although the baseline estimate is smoother, it overestimates the noise floor at lower frequencies. Although the Qi and Hillman (1997) technique provides a smoother baseline estimate, it does not reveal extra theoretical or experimental information over the de Krom (1993) technique, with respect to how the baseline estimate is formed.

III. ANALYSIS OF CEPSTRUM-BASED NOISE BASELINE ESTIMATION

Although the cepstrum has been used for HNR estimation, a formal description of what the noise baseline represents has not been provided. It is seen from the Introduction that harmonic and formant structure can be separated and it has been reviewed that this fact has been put to effective use in HNR estimation. However, at present no theoretical or experimental detail has been provided in order to determine where the baseline resulting from the cepstral liftering process is positioned with respect to the original spectrum. The following analysis examines the cepstrum applied to voiced speech containing aspiration noise in order to determine the nature of the baseline estimation process.

A voiced speech waveform, $s_{en}(t)$, including aspiration noise,³ $n(t)$ at the glottal source, can be approximated as

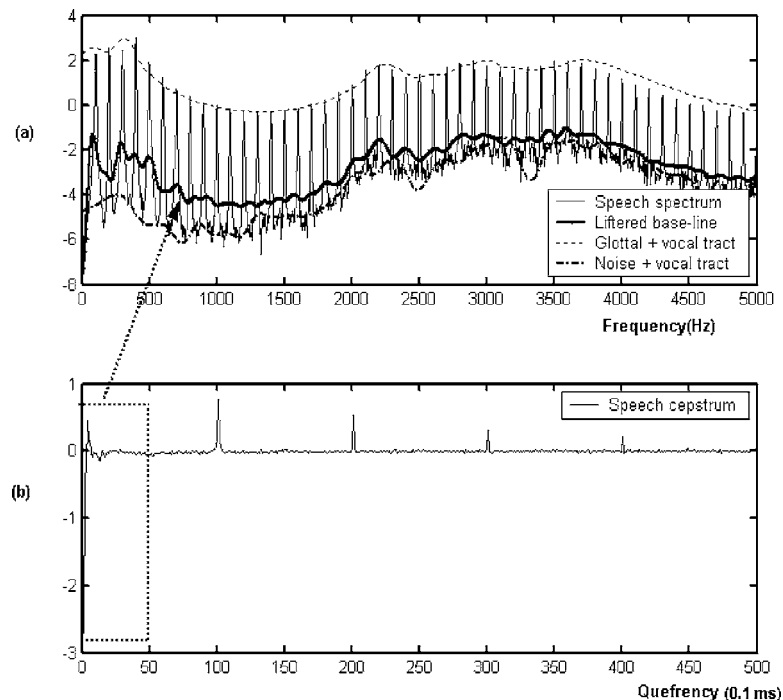


FIG. 4. (a) Spectrum of synthesized vowel /AH/ illustrating the glottal excited vocal tract and the noise excited vocal tract. The noise baseline estimate is shown superimposed. (b) Cepstrum of vowel /AH/ highlighting the signal distinction at low and high quefrequencies.

$$s_{\text{en}}(t) = [e(t) * g(t) + n(t)] * v(t) * r(t), \quad (2)$$

where $e(t)$ is a periodic impulse train, $g(t)$ is a single glottal pulse, $v(t)$ is the impulse response of the vocal tract, $r(t)$ represents the radiation load and $*$ indicates convolution.

Applying a Hanning window (w)

$$s_{\text{en}}^w(t) = \{[e(t) * g(t) + n(t)] * v(t) * r(t)\} \times w(t). \quad (3)$$

Provided the window length is sufficiently long the window function can be moved inside the convolution (Schafer and Rabiner, 1970) to give

$$s_{\text{en}}^w(t) = [(e_w(t) * g(t)) + n_w(t)] * v(t) * r(t). \quad (4)$$

Taking the Fourier transform gives

$$S_{\text{en}}^w(f) = [E_w(f) \times G(f) + N_w(f)] \times V(f) \times R(f). \quad (5)$$

Taking the logarithm of the magnitude squared values and approximating the signal energy at harmonic locations, $\log |S_{\text{en}}^w(f)|_h^2$, and at between-harmonic locations, $\log |S_{\text{en}}^w(f)|_{\text{bh}}^2$, gives

$$\log |S_{\text{en}}^w(f)|_h^2 = \log |E_w(f) \times G(f)|^2 + \log |V_R(f)|^2, \quad (6)$$

$$\log |S_{\text{en}}^w(f)|_{\text{bh}}^2 = \log |N_w(f)|^2 + \log |V_R(f)|^2, \quad (7)$$

where $V_R(f)$ is the Fourier transform of $v(t)$ and $r(t)$ combined.

Although the noise spectrum is broadband, its estimation in the presence of a harmonic signal can be concentrated at between-harmonic locations, i.e., in the spectrum of voiced speech, signal energy dominates at harmonic locations and noise energy dominates at between-harmonic locations. This approximation becomes more exact if the spectra are averaged, in which case more accurate harmonic estimates are obtained and the between-harmonics approach the true noise variance.

The cepstral technique is described with reference to Fig. 4. The spectrum of voiced speech can now be seen to consist of (a) two slowly varying spectral envelopes, one due to the glottal flow excited vocal tract $|G(f) \times V(f)|$, and the other due to the noise excited vocal tract, $|N(f) \times V(f)|$, and (b) a rapidly varying periodic structure due to $|E(f)|$. The noise excited vocal tract, $|N(f) \times V(f)|$, constitutes the spectral envelope at the bottom of speech spectrum, while the spectral envelope due to the glottal source excited vocal tract, $|G(f) \times V(f)|$, constitutes the spectral envelope at the top, as depicted in Fig. 4(a).

The two spectral envelopes are lumped together in the low quefrequency part of the speech cepstrum [highlighted segment in Fig. 4(b)]. The periodic structure due to $|E(f)|$ translates to high amplitude peaks (harmonics) at periodic locations in the cepstrum. If the low quefrequency cepstral coefficients are low-passed liftered (i.e., filtered in the cepstral domain) below the first harmonic, the Fourier transformed result represents the average of the spectral envelopes due to $|G(f) \times V(f)|$ and $|N(f) \times V(f)|$ as illustrated in Fig. 4(a) (solid line).

This Fourier transformed liftered cepstrum provides a noise baseline estimate that forms the basis for the HNR estimation techniques presented in de Krom (1993) and Qi

and Hillman (1997). The Fourier transformed liftered low quefrequency coefficients are assumed to provide an estimate of the noise excited vocal tract in voiced speech. However, the present development suggests that the transformed liftered cepstrum behaves like a moving average (MA) filter and is dependent on glottal harmonic and noise contributions as well as the temporal analysis window length.

A moving average filter is given as

$$\log |S_{\text{en}}^w[n]|_{\text{MA}}^2 = \frac{1}{M} \sum_{k=-\frac{M-1}{2}}^{\frac{M-1}{2}} \log |S_{\text{en}}^w[n-k]|^2, \quad (8)$$

where $\log |S_{\text{en}}^w[n-k]|^2$ represents spectral amplitudes at discrete points $n-k$, M is an odd integer, and $\log |S_{\text{en}}^w[n]|_{\text{MA}}^2$ is the output.

This MA filter is applied to the log spectrum of a segment of a 100 Hz synthesized voiced speech signal together with the liftered baseline [Fig. 5(a)]. The inverse Fourier transform of the result of the MA filter applied to the log spectrum is shown superimposed on the low passed liftered cepstrum in Fig. 5(b). The minimum value of M for which the smoothed average closely fits the liftered baseline depends on the fundamental frequency. For a 100-Hz signal the minimum value of M is 5 (sampling frequency=10 kHz), where the minimum value is the number of samples required to cover the frequency range from one harmonic peak to the next. The following two sections examine the MA filter (with M equal to the number of points between successive harmonics) nature of the liftered baseline in a systematic manner.

A. Influence of temporal analysis window length on the baseline estimate

The temporal analysis window length has a significant impact on the results of the cepstral analysis. In order to illustrate this effect, cepstral analysis of a windowed impulse train (100 Hz) is implemented for different temporal analysis window lengths. Figures 6(a)–6(c) show the liftered cepstral baseline (cepstral coefficients below the first harmonic) plotted against the original spectrum of the impulse train. Essentially, the cepstrum of the impulse train consists of a set of harmonics resulting from the harmonics of the original impulse spectrum and low quefrequency components that correspond to the dc content of the spectral harmonic structure. Precisely, the first cepstral coefficient in the cepstrum represents the dc content of the harmonic structure and is dependent on the size of the temporal analysis window. For a short temporal analysis window, the baseline estimate from low-pass liftering is very high, lying in the middle of the spectrum of the impulse train [Fig. 6(a)], and progressively approaches the base of the spectrum as the temporal analysis window length increases [Fig. 6(c)].

B. Influence of the glottal source spectral tilt on the baseline estimate

A noise-free, windowed, periodic glottal source is analyzed to highlight the fact that the spectral tilt due to the

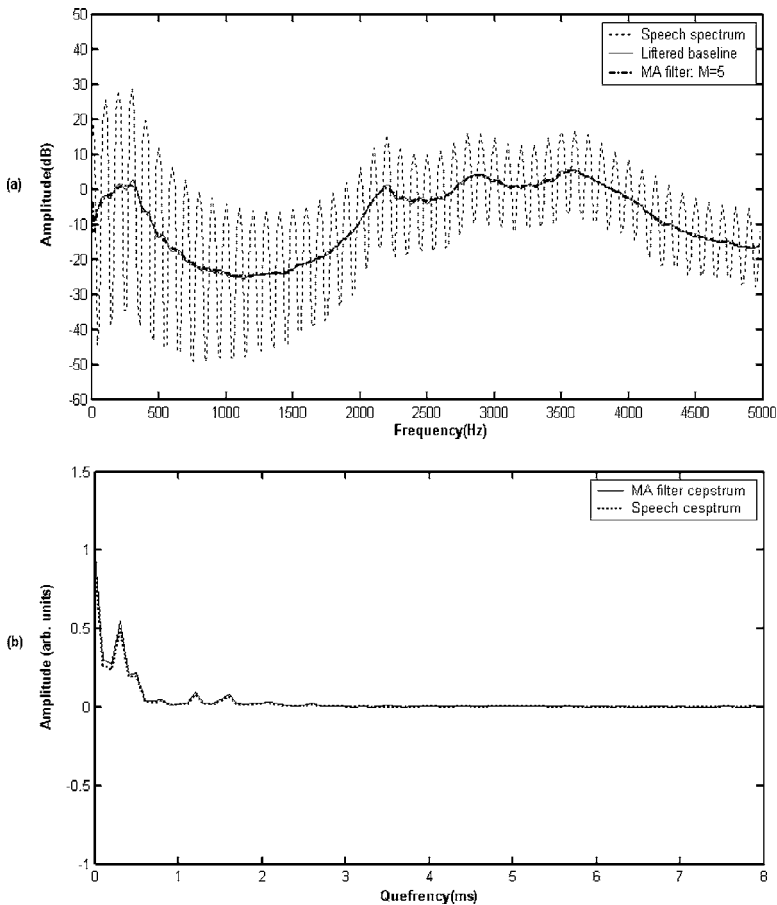


FIG. 5. (a) Log magnitude spectrum of the vowel /AH/, together with the Fourier transform of the low-pass liftered baseline and the MA filter (moving average of $M=5$ points) estimates for a 100-Hz signal. (b) Low-pass liftered cepstrum and inverse Fourier transform of the result of applying the MA filter to the log spectrum.

glottal source is the primary contributor to the skewing of the noise baseline estimate in the speech spectrum.

Employing the same analysis applied to the voiced speech waveform a windowed periodic glottal source, $g_e^w(t)$, can be written as

$$g_e^w(t) = e_w(t) * g(t), \quad (9)$$

where $e(t)$ is a periodic impulse train, $g(t)$ is a single glottal pulse, and $w(t)$ is the window function.

Taking the Fourier transform and the logarithm of the magnitude squared values gives

$$\log |G_e^w(f)|^2 = \log |E_w(f) \times G(f)|^2. \quad (10)$$

The logarithm of the product in Eq. (10) turns the multiplicative components consisting of the periodic pulse train and the glottal source into additive components comprised of $\log |G(f)|$ and a periodic pulse train $[\log |E_w(f)|]$. The inverse Fourier transform (i.e., the cepstrum) operates separately on these additive components, producing cepstral coefficients that contain information about source spectral tilt at low quefrequencies and periodic structure representing the pulse train at higher quefrequencies. An illustration of cepstral analysis of a periodic glottal source signal is given in Fig. 7. The log magnitude spectrum of the liftered cepstrum (cepstral coefficients below the first rahmonic) is plotted on the same axis as the original source spectrum (2048-point log magnitude FFT). The amplitude at zero quefrequency represents the 0 Hz or dc component and is dependent on the temporal window length as described

previously. The dc contributions become smaller as the analysis window length increases and, as a consequence, the liftered spectral envelope (spectral tilt) approaches the baseline of the source spectrum.

Examination of a periodic impulse train with noise illustrates the effect of signal periodicity and noise on the cepstral baseline estimation process, independent of the glottal wave-shape contribution (Fig. 8). Because the impulse has a flat spectrum, the liftered noise floor is not skewed due to the harmonic peaks (as per the spectrum of the periodic glottal source) and provides a close fit to the contour of the actual noise floor, residing close to the noise floor when the temporal window length is sufficiently long.

The Fourier transform of a windowed periodic glottal source with additive noise, $g_{en}^w(t)$, is given as

$$G_{en}^w(f) = E_w(f) \times G(f) + N_w(f). \quad (11)$$

Taking the logarithm of the magnitude squared values and, as stated previously, approximating the signal energy at harmonic locations, $\log |G_{en}^w(f)|_h^2$, and at between-harmonic locations, $\log |G_{en}^w(f)|_{bh}^2$, gives

$$\log |G_{en}^w(f)|_h^2 = \log |E_w(f) \times G(f)|^2 \quad (12)$$

$$\log |G_{en}^w(f)|_{bh}^2 = \log |N_w(f)|^2. \quad (13)$$

Applying the cepstrum to this signal and obtaining the liftered baseline, it can be seen that the baseline is influenced by the noise source, glottal spectral tilt, and temporal analysis window size. As the temporal window length increases the

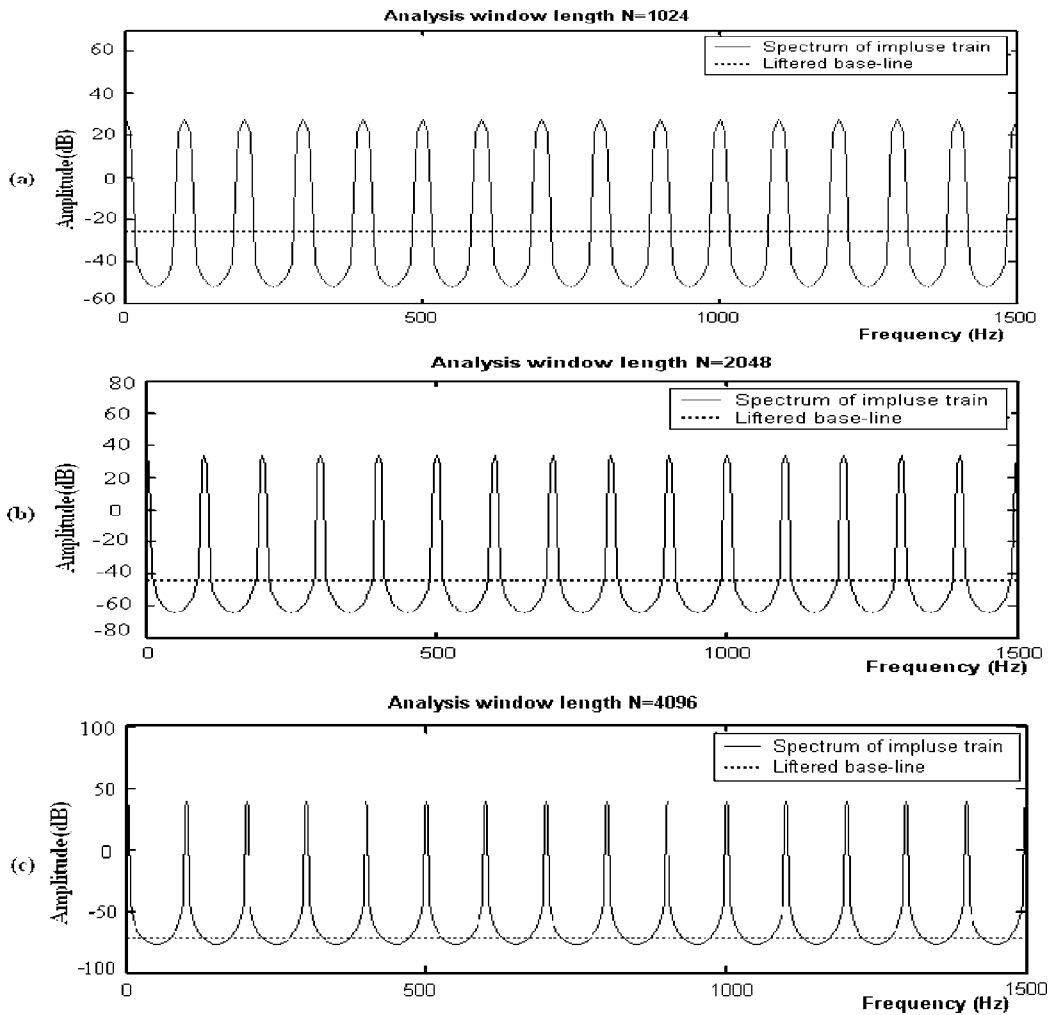


FIG. 6. Magnitude spectrum of an impulse train and lifted low cepstral components for different temporal analysis window lengths. (a) $N=1024$ sample points, (b) $N=2048$ sample points, and (c) $N=4096$ sample points.

baseline approximates the noise floor more accurately (because more estimates are available for the between harmonics as opposed to the harmonics and the Fourier transform of the lifted cepstrum behaves like a moving average filter applied to the logarithmic spectrum), but the influence of the spectral tilt remains (Fig. 9). The development of the new HNR estimator follows directly from the key result of this section—namely, that the Fourier transform of the low-pass lifted cepstrum of voiced speech is comparable to applying a moving average filter to the logarithmic power spectrum of voiced speech.

IV. HNR ESTIMATION METHOD

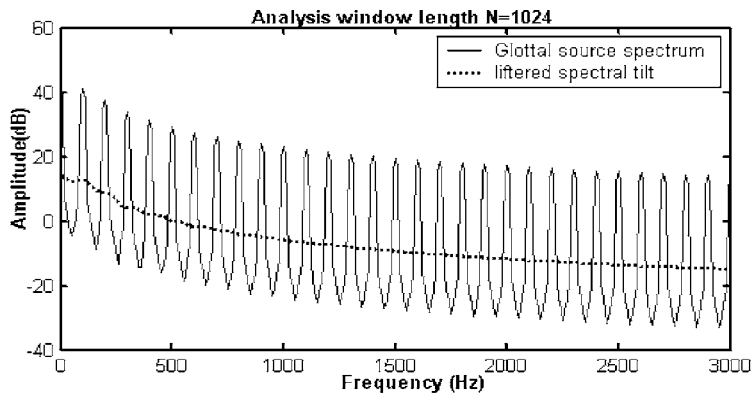
In this section, a new cepstrum based HNR method is presented. As described previously, the cepstral lifting process is equivalent to computing a moving average of the underlying spectrum, with contributions received from the noise excited vocal tract and the source excited vocal tract. The weighting given to these contributions is primarily determined by the duty cycle of the harmonic spectrum (i.e., the width of the harmonic compared to the width of the between harmonic). To a first order approximation, if these widths are equal, the baseline estimate appears in the center

between the noise and glottal excited vocal tract estimates. This concept is illustrated in Fig. 10, where a Hanning window is used to analyze synthesized speech waveforms of four different fundamental frequencies. From the figure it can be seen that the baseline estimated via low-pass lifting matches the midpoints of the spectrum waveshape. The symmetric fitting of the baseline estimate from low-pass lifting for different fundamental frequencies depends on an appropriate choice of temporal analysis window length. An empirically derived relationship between temporal analysis window length and fundamental frequency, f_0 , is given by⁴

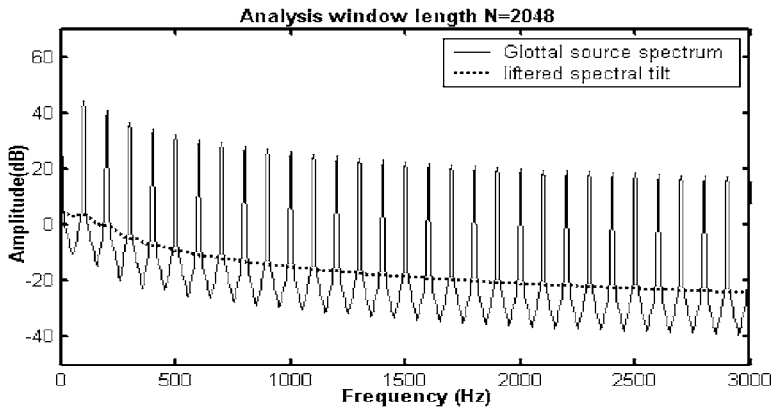
$$L = \frac{5fs}{f_0}, \quad (14)$$

where L =temporal analysis window length, f_0 is the fundamental frequency, fs is the sampling frequency.

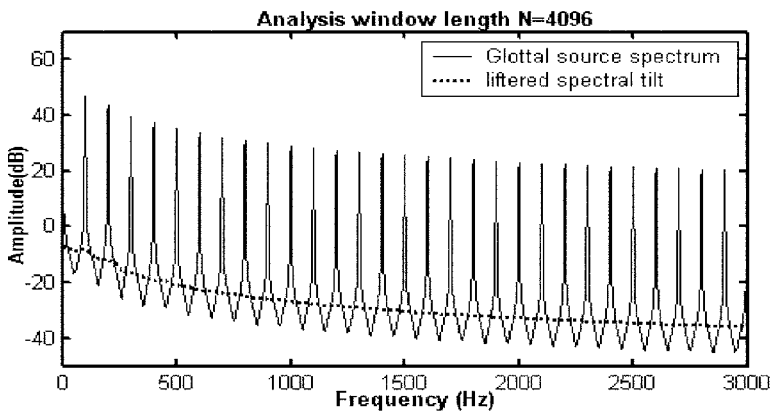
In order to calculate the noise energy, $|N_i|^2$, the symmetric baseline, $\log |C_i|^2$ (where $|C_i|^2$ is the lifted symmetric baseline estimate in linear scale) is subtracted from the harmonic energy, $\log |S_i|^2$ and this difference is then subtracted from the baseline estimate, $|\log |C_i|^2 - (\log |S_i|^2 - \log |C_i|^2)|$. This provides the noise estimate in logarithmic scale ($\log |N_i|^2$). This estimate is raised to the power of ten and the



(a)



(b)



(c)

FIG. 7. Magnitude spectrum of the glottal source and liftered spectrum approximating the spectral tilt for different temporal analysis window lengths. (a) $N=1024$ sample points, (b) $N=2048$ sample points, and (c) $N=4096$ sample points.

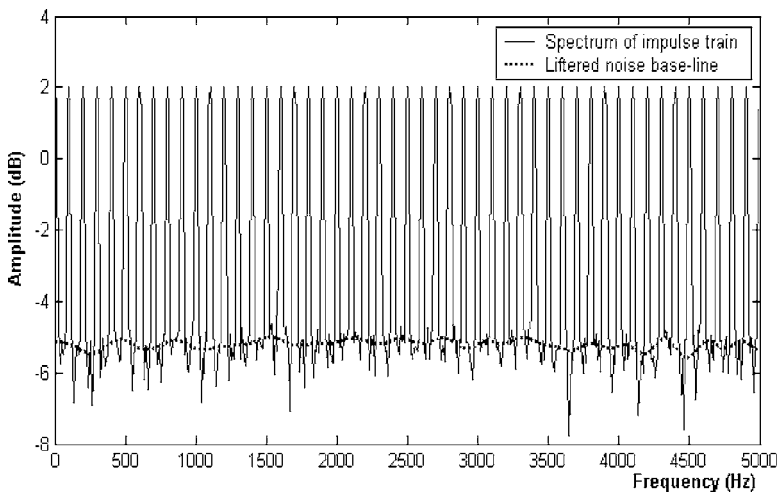


FIG. 8. Magnitude spectrum of an impulse train with 1% additive noise and liftered noise base-line (temporal window length 4096 points).

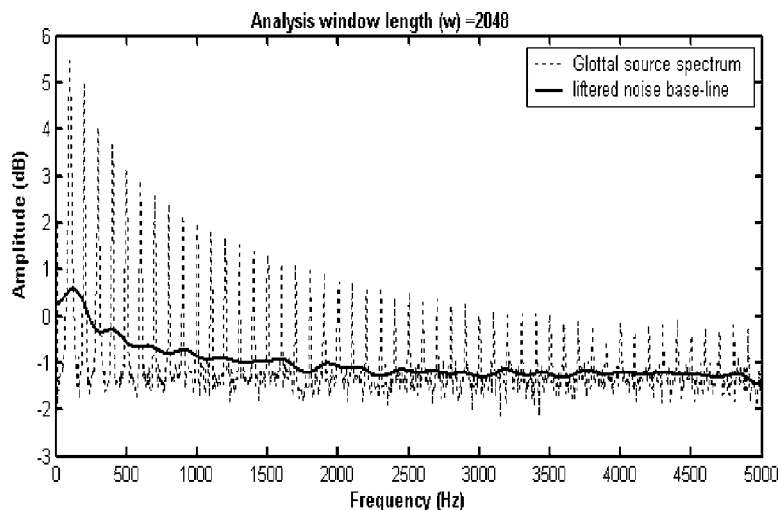


FIG. 9. Magnitude spectrum of a glottal source with aspiration noise and filtered noise baseline. The spectrum is calculated with a temporal analysis window length of 2048 points.

noise is summed. A sum, representing the signal energy, is taken of the harmonic peaks in the power spectrum of the signal and the HNR is calculated as

$$\text{HNR} = 10 \log_{10} \left[\frac{\sum_{i=0}^{L/2-1} 10^{\log |S_i|^2}}{\sum_{i=0}^{L/2-1} 10^{(\log |C_i|^2 - (\log |S_i|^2 - \log |C_i|^2))}} \right]. \quad (15)$$

The method is free from the influence of the glottal source spectral tilt. It is not necessary to estimate the noise baseline itself; all that is required is that the baseline estimate is half-

way between harmonic amplitude and noise floor estimates.⁵

V. HNR ANALYSIS

A. Synthesis data

In order to evaluate the performance of the new HNR estimation technique (termed symmetric baseline estimation, SBE), synthesized glottal source and vowel /AH/ waveforms are generated at five fundamental frequencies, beginning at 80 Hz and increasing in four steps of 60 Hz up to 320 Hz. The model, described in Fant *et al.* (1985), is adopted to synthesize the glottal flow waveform for modal register,

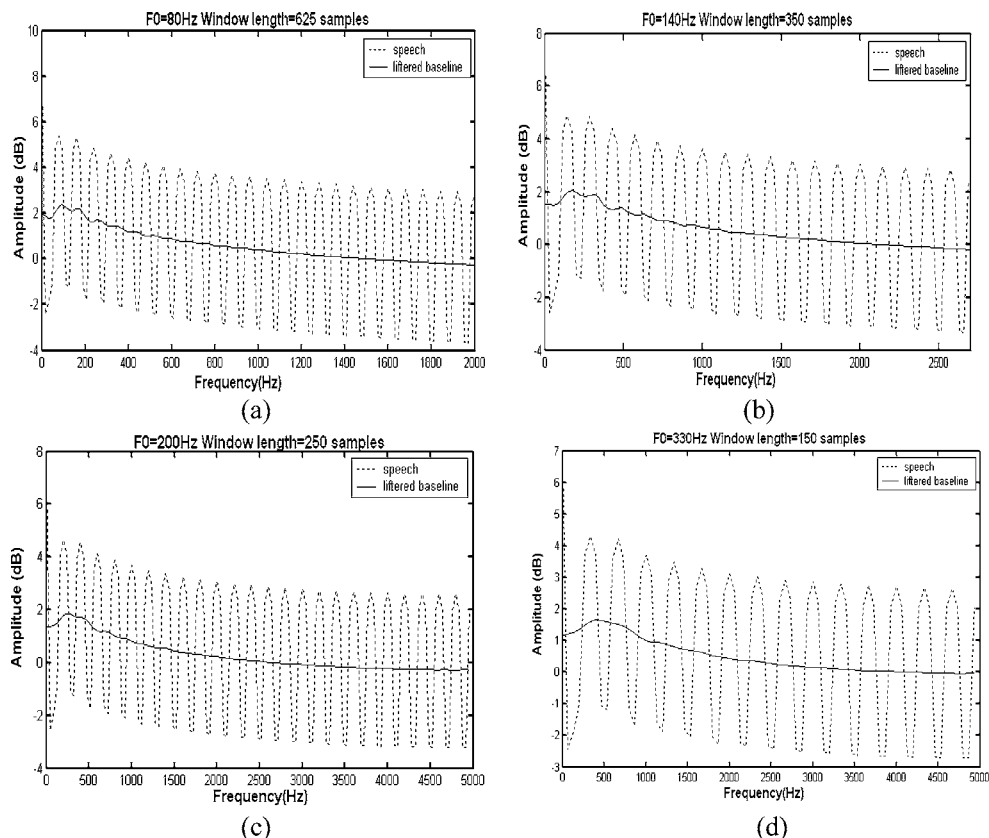


FIG. 10. Analysis window (Hanning) length for symmetric baseline estimation (SBE) for four different fundamental frequencies of the speech signal. (a) $f_0=80$ Hz, (b) $f_0=140$ Hz, (c) $f_0=200$ Hz, and (d) $f_0=330$ Hz.

while the vocal tract impulse response for the vowel /AH/ is modeled with a set of poles (Childers, 2000). Lip radiation is modeled by a first order difference operator $R(z)=1-z^{-1}$. A sampling rate of 10 kHz is used for the synthesis.

Noise is introduced by adding pseudo-random noise to the glottal pulse via a random noise generator, arranged to give additive noise of a user-specified variance (five levels from s.d. 0.125%, doubling in steps up to 2%). Two noise types are implemented. The first type, termed background noise, is calculated as a percentage of the average amplitude of the glottal pulse, but is otherwise independent of the glottal cycle. Thus the same amount of noise occurs over the whole of the glottal cycle. The second type of noise modeled, termed segment noise, is used to model noise bursts, which occur during the glottal closing phase. The HNRs for the glottal flow waveform are 58 to 34 dB, decreasing in steps of 6 dB for each noise type. The HNR for the corresponding speech signals are different from those for the glottal source because the source signal receives different resonant contributions depending on the fundamental frequency (Murphy, 1999). However, through synthesizing the speech signal with and without glottal noise, the HNR of the synthesized speech signal can also be predetermined.

In addition, jitter and shimmer are implemented. Random shimmer is introduced by adding a random variable gain factor (doubling levels from 1% to 32% s.d.) to the amplitude of the pitch period impulse train prior to convolution with the glottal pulse. Jitter is introduced via time scaling of glottal periods. The scaled periods are amplitude normalized to avoid the concomitant introduction of shimmer. Jitter is implemented at five equal intervals 1% to 5% s.d. for the 100-Hz signal.

B. Implementation

The power spectrum of the test signal is computed using a Hanning window. An appropriate temporal window length, L , is selected, based on the fundamental frequency of the underlying test signal as per Eq. (14). The inverse Fourier transform is applied to the log of the power spectrum and low cepstral coefficients (from zero up to ten samples before the first harmonic) are filtered to represent the baseline. The HNR is then calculated as per Eq. (15).

VI. HNR RESULTS

In order to evaluate the performance of the method, the estimated HNR is compared to the original HNR (dotted curve, e.g., noise of s.d. 1% corresponds to a HNR of 40 dB). Figures 11–13 show accurate SBE-based HNR estimates for the synthesized glottal source and vowel /AH/ data versus continuous random noise, segmental noise, and shimmer, respectively. Figure 14 shows HNR decreases with increasing jitter levels for both glottal and speech data [jitter (1%–5%) for a signal with fundamental frequency of 100 Hz].

VII. DISCUSSION

The present theoretical treatment only deals with (continuous) aspiration noise whereas the analysis is performed

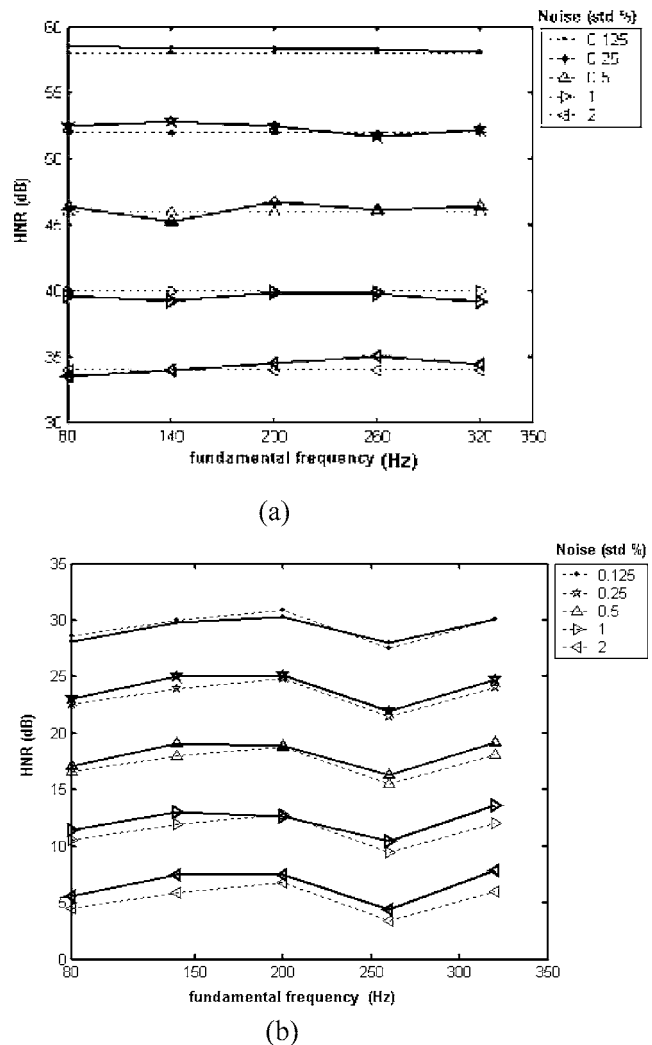


FIG. 11. Estimated HNR (solid line, SBE method) versus f_0 for (a) synthesized glottal source waveform containing continuous glottal noise (dotted line, actual HNR) and (b) synthesized vowel /AH/ with continuous glottal noise (dotted line, actual HNR).

on signals containing continuous glottal noise, segment noise, jitter, and shimmer. Through viewing HNR in the frequency domain, harmonics signify signal energy whereas “between harmonics” signify aperiodicity, no matter what the source of that aperiodicity is. Hence the development is applicable to a broader class of aperiodicity as tested under Analysis although the effects of jitter on the spectrum are considerably more complicated (Murphy, 2000) and do not naturally lend themselves to spectral averaging.

The HNR values estimated with SBE provide close estimates of the corresponding input HNRs across a range of f_0 values for the synthesized glottal and speech data containing noise and shimmer. For the jittered signals HNR is shown to decrease as jitter increases.

The HNR estimated via SBE is a conventional HNR estimator, i.e., it represents the ratio between the periodic to aperiodic energy in the signal. In order to extract this ratio via the cepstrum, the log spectral estimates must first be exponentiated (raised to the power of ten), i.e., converted to linear values. This contrasts the present method with the techniques of de Krom (1993) and Qi and Hillman (1997),

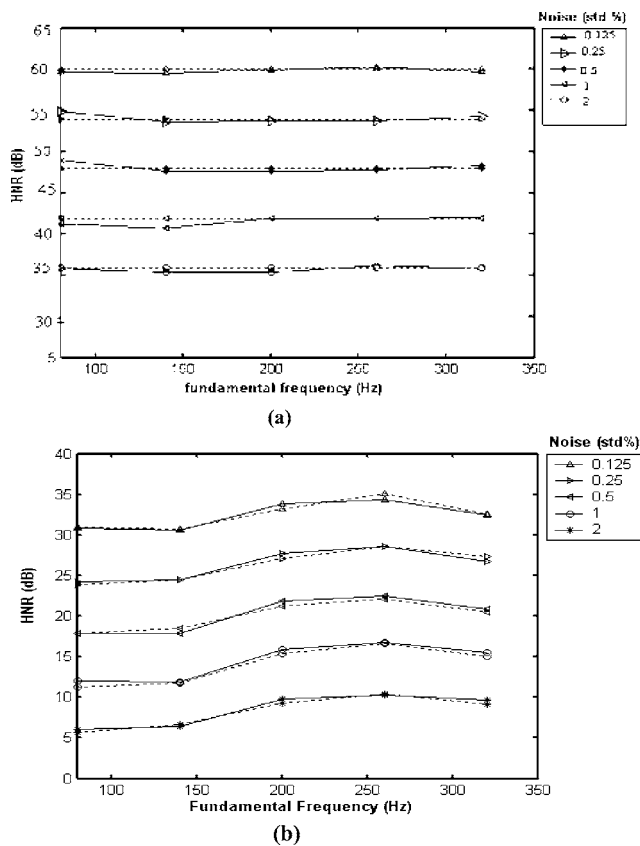


FIG. 12. Estimated HNR (solid line, new method SBE) versus f_0 for (a) synthesized glottal source waveform containing segment glottal noise (dotted line, actual HNR) and (b) synthesized vowel /AH/ with glottal segment noise (dotted line, actual HNR).

where the harmonic and noise estimates appear to be taken directly from the logarithmic power spectrum. These methods could similarly be modified (i.e., exponentiated) to derive a conventional HNR. This is, in fact, performed in Murphy and Akande (2005). The advantage of using a conventional HNR from a validation viewpoint is that the answer is known in advance from the synthesis (at least for noise and shimmer). An advantage of the de Krom (1993) and Qi and Hillman (1997) methods lies in the fact that, in taking an average of the dB ratio between each harmonic and “between-harmonic” in the spectrum, to a first approximation, the influence of the vocal tract is removed, hence providing an index more closely related to the glottal source. The present technique can easily be modified, through removing the exponentiation step, to provide an average of dB harmonic to “between-harmonic” ratios.

Figures 2 and 3 suggest that the de Krom (1993) and Qi and Hillman (1997) methods, respectively slightly overestimate and underestimate HNR, and this is confirmed in Murphy and Akande (2005). A slight modification to the de Krom (1993) technique might include taking the mean “between-harmonic” estimate as opposed to the minimum, to perform baseline shifting. The Qi and Hillman (1997) technique is modified in Murphy and Akande (2005) through preemphasizing the harmonics. This latter technique still requires a long analysis window whereas the proposed modified de Krom (1993) method could be implemented with a short analysis window as per the present study.

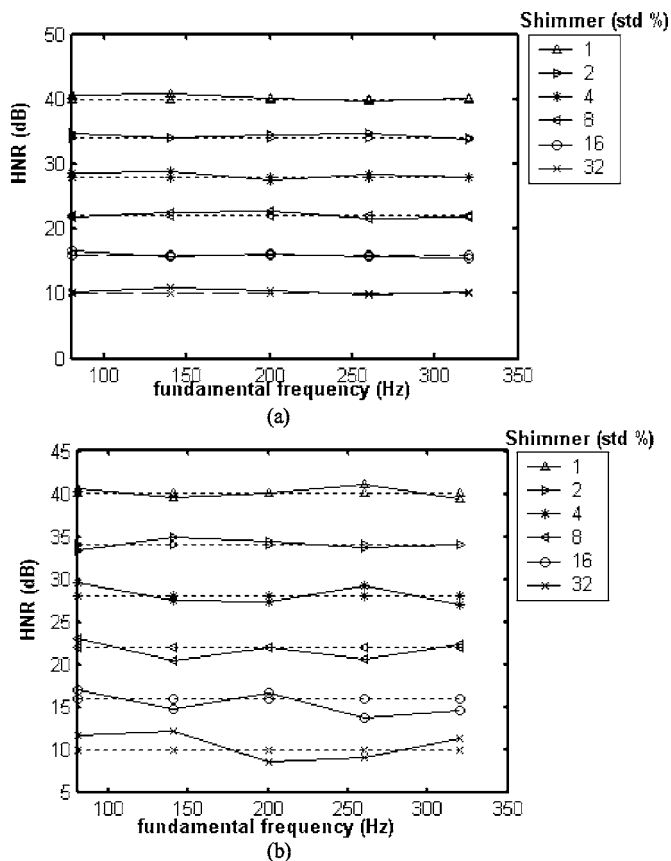


FIG. 13. Estimated HNR (solid line, new method SBE) versus f_0 for (a) synthesized glottal source waveform containing shimmer (dotted line, actual HNR) and (b) synthesized vowel /AH/ with glottal shimmer (dotted line, actual HNR).

Two important issues need to be considered with respect to HNR estimation for sustained vowel phonation when inferring glottal noise levels:

- (i) HNR is a global indicator of voice periodicity
- (ii) HNR is indirectly related to the noise level of the glottal source.

HNR provides a global estimate of signal periodicity. Hence a low value of HNR can arise from any form of

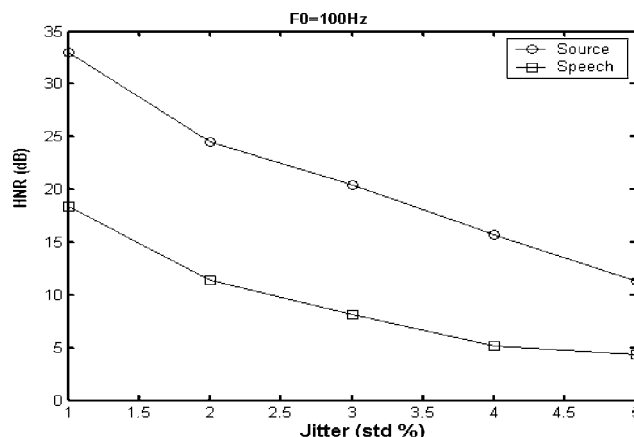


FIG. 14. Estimated HNR (SBE method) versus jitter for synthesized glottal source and voiced speech waveforms (vowel /AH/).

aperiodicity—for example, from aspiration noise, jitter, shimmer, nonstationarity of the vocal tract, or other waveform anomalies. One goal of acoustic analysis of voice is to derive more exacting measures representative of a particular aperiodicity from the speech signal. To this end methods have evolved that attempt to estimate aspiration noise independent of other waveform aperiodicities (cf. Michaelis *et al.*, 1998; Murphy, 1999). To implement the latter algorithm for real speech requires accurate extraction of the glottal flow waveform including the aspiration noise component. Recent studies (Akande and Murphy, 2005; Fu and Murphy, 2006) have focused on obtaining accurate glottal estimates, however further work is still required in order to extract the aspiration noise component. In the meantime studies such as the present one are useful in that the degree to which individual perturbations affect HNR have been correlated. Therefore, if jitter and shimmer are measured separately, and assuming other waveform aperiodicities are negligible, the degree to which HNR is due to aspiration noise can be inferred. Klingholtz and Martin (1985) have experimented with this approach, though estimating shimmer and jitter from the speech signal brings its own set of measurement problems (Hillenbrand, 1987).

As shown in the synthesis, for equal levels of glottal noise HNR is not equal across f_0 for the speech signal. If the purpose is to compare HNRs across voice patients/pathologies, then this is problematic as the signals can have equal glottal noise levels yet different HNRs due to their different f_0 's. The f_0 dependence is due to the fact that the formants are differentially excited by the harmonics as f_0 changes (Murphy, 1999). Alternative noise indices that are not f_0 or vocal tract dependent would appear merited.

VIII. CONCLUSION

It has been shown, using a theoretical development and through systematic tests on synthesis data, that the Fourier transform of the low-pass lifted cepstrum is comparable to applying a moving average (MA) filter to the logarithmic power spectrum. The noise baseline estimate (or vocal tract estimate in general speech processing) via the cepstrum technique is shown to be influenced by the glottal source excited vocal tract and the noise excited vocal tract. The influence of the glottal source excited vocal tract is reduced as the temporal window length increases. In this sense the baseline estimation via the cepstrum is equivalent to the action of a MA filter applied to the logarithmic spectrum. A new method is introduced that selects the appropriate temporal analysis window length and is shown to create a baseline estimate from low-pass lifting that symmetrically interpolates the original log spectrum of voiced speech. The technique (symmetric baseline estimation, SBE) is shown to provide accurate HNR estimates for a variety of glottal aperiodicities. The proposed technique employs a short temporal analysis window length and hence shows promise for use in the analysis of connected speech. Further development of the technique will investigate its application to real normal and pathological voice data for sustained vowels and connected speech. Future work may also include a comparison, in the quefrency

domain, of the filtering action of the MA filter with that of the low-pass lifter. As the present development provides new insights into cepstral lifting in general, the work may also find application in studies of spectral envelope estimation for use in vocoding and speech recognition.

ACKNOWLEDGMENTS

This work is supported through Enterprise Ireland research grant Nos. STR/2000/100 and RIF/2002/037.

¹Alternative cepstral based noise estimates have used the high quefrency end of the cepstrum [e.g., Yegnanarayana *et al.* (1998) use a region centered on the first harmonic peak. However, this is only used as a preprocessing step for their recursive algorithm. Hillenbrand *et al.* (1994) and Dejonckere and Wieneke (1994) investigate the first harmonic peak in the cepstrum. The noise estimate obtained in this case has recently been shown to differ from a conventional HNR (Murphy, 2006), which is the subject of the present paper.

²In de Krom (1993) and Qi and Hillman (1997) it appears that a signal-to-noise ratio is extracted directly from the power spectrum in dBs. Therefore, an average of dB values (for each harmonic/between-harmonic ratio estimate) is extracted, which contrasts with a conventional HNR estimate, which sums the signal and noise energies before taking the logarithm of their quotient.

³In this development the noise component is treated as zero mean Gaussian noise with a resulting flat spectral characteristic. In reality glottal noise may not be continuous throughout the glottal cycle. Imaizumi (1988) finds two different timing relationships between glottal configuration and turbulent noise generation in analyzing pathological voices. In certain cases he finds turbulent noise predominantly during the open phase while in others he finds the noise components are enhanced even after the major part of the glottis is closed. It seems plausible that glottal noise can have many different temporal characteristics (for example, depending on a specific pathology). In the present study noise is synthesized throughout the glottal cycle (continuous noise) and limited to the region centered around glottal closure (segment noise). From the point of view of the analysis the noise type implemented is not crucial. The Fourier transform of the low-pass lifted cepstrum provides a moving average of the noise excited vocal tract (no matter what the specific spectral pattern is due to the particular noise type) and the glottal excited vocal tract.

⁴This equation states that five periods of the waveform are found to provide the required spectral resolution to produce a baseline, which is approximately centered between the noise excited vocal tract and the glottal excited vocal tract. However, there is a range of values of temporal window length for which the corresponding spectral estimates provides approximately a symmetrical baseline. Hence the method need not necessarily estimate every pitch period, but the relationship between temporal analysis window length and average f_0 estimate over that window length should remain approximately constant.

⁵An alternative approach to baseline estimation is provided in Murphy and Akande (2005), in which the spectral tilt due to the glottal source is deemphasized. $|S_{en}^w(f)|_n$ is approximated using periodogram averaging to provide estimates for $S_e^w(t)$. The preemphasis differs from conventional approaches in that only the averaged harmonics are preemphasized while the noise floor remains unchanged. Although improvement in HNR estimation has been demonstrated through use of this technique, the necessity for a long temporal analysis window precludes the use of the method for connected speech.

Akande, O., and Murphy, P. J. (2005). "Estimation of the vocal tract transfer function for voiced speech with application to glottal wave analysis," *Speech Commun.* **46**, 15–36.

Childers, D. G. (2000). *Speech Processing and Synthesis Toolboxes* (Wiley, New York).

Dejonckere, P., and Wieneke, G. H. (1994). "Spectral, cepstral and aperiodicity characteristics of pathological voice before and after phonosurgical treatment," *Clin. Linguist. Phonetics* **8**(2), 161–169.

de Krom, G. (1993). "A cepstrum based technique for determining a harmonics-to-noise ratio in speech signals." *J. Speech Hear. Res.* **36**(2), 254–266.

Fant, G., Liljencrants, J., and Lin, Q. G. (1985). "A four parameter model of

- glottal flow," *STL-QPSR* **4**, 1–12.
- Fu, Q., and Murphy, P. J. (2006). "Robust glottal source estimation based on joint source-filter model optimization," *IEEE Trans. Audio, Speech Language Process.*, **14**, 492–501.
- Hillenbrand, J. (1987). "A methodological study of perturbation and additive noise in synthetically generated voice signals," *J. Speech Hear. Res.* **30**, 448–461.
- Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (1994). "Acoustic correlates of breathy vocal quality," *J. Speech Hear. Res.* **37**, 769–777.
- Imaizumi, S. (1988). "A preliminary study on the generation of pathological voice types," in *Vocal Fold Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 249–258.
- Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S. (1986). "Normalised noise energy as an acoustic measure to evaluate pathologic voice," *J. Acoust. Soc. Am.* **80**, 1329–1334.
- Klingholtz, M., and Martin, F. (1985). "Quantitative spectral evaluation of shimmer and jitter," *J. Speech Hear. Res.* **28**, 169–174.
- Michaelis, D., Frohlich, M., and Strube, H. W. (1998). "Selection and combination of acoustic features for the description of pathologic voices," *J. Acoust. Soc. Am.* **103**(3), 1628–1639.
- Murphy, P. J. (2006). "On first harmonic amplitude in the analysis of synthesized aperiodic voice signals," *J. Acoust. Soc. Am.* **120**(5), 2896–2907.
- Murphy, P. J. (2000). "Spectral characterisation of jitter, shimmer and additive noise in synthetically generated voice signals," *J. Acoust. Soc. Am.* **107**, 978–988.
- Murphy, P. J. (1999). "Perturbation-free measurement of the harmonics-to-noise ratio in speech signals using pitch-synchronous harmonic analysis," *J. Acoust. Soc. Am.* **105**, 2866–2881.
- Murphy, P. J., and Akande, O. O. (2005). "Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation," CD ROM Proceedings NOLISP Non-linear Speech processing, COST-277, Barcelona.
- Noll, A. M. (1967). "Cepstrum pitch determination," *J. Acoust. Soc. Am.* **41**, 293–309.
- Qi, Y., and Hillman, R. E. (1997). "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals," *J. Acoust. Soc. Am.* **102**(1), 537–543.
- Schafer, R. W., and Rabiner, L. R. (1970). "System for automatic analysis of voiced speech," *J. Acoust. Soc. Am.* **47**, 634–648.
- Yegnanarayana, B., d'Alessandro, C., and Darsinos, V. (1998). "Decomposition of the speech signal into deterministic and stochastic components," *IEEE Trans. Speech Audio Process.* **6**(1), 1–11.
- Yumoto, E., Gould, W. J., and Baer, T. (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness," *J. Acoust. Soc. Am.* **71**, 1544–1550.

Effect of spectral frequency range and separation on the perception of asynchronous speech

Eric W. Healy^{a)}

Department of Communication Sciences and Disorders, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 29208 and Psychoacoustics Laboratory, Arizona State University, Department of Speech and Hearing Science, P.O. Box 870102, Tempe, Arizona 85287-0102

Sid P. Bacon

Psychoacoustics Laboratory, Arizona State University, Department of Speech and Hearing Science, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 13 January 2006; revised 23 November 2006; accepted 23 November 2006)

The use of across-frequency timing cues and the effect of disrupting these cues were examined across the frequency spectrum by introducing between-band asynchronies to pairs of narrowband temporal speech patterns. Sentence intelligibility by normal-hearing listeners fell when as little as 12.5 ms of asynchrony was introduced and was reduced to floor values by 100 ms. Disruptions to across-frequency timing had similar effects in the low-, mid-, and high-frequency regions, but band pairs having wider frequency separation were less disrupted by small amounts of asynchrony. In experiment 2, it was found that the disruptive influence of asynchrony on adjacent band pairs did not result from disruptions to the complex patterns present in overlapping excitation. The results of experiment 3 suggest that the processing of speech patterns may take place using mechanisms having different sensitivities to exact timing, similar to the dual mechanisms proposed for within- and across-channel gap detection. Preservation of relative timing can be critical to intelligibility. While the use of across-frequency timing cues appears similar across the spectrum, it may differ based on frequency separation. This difference appears to involve a greater reliance on exact timing during the processing of speech energy at proximate frequencies. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2427113]

PACS number(s): 43.71.Es, 43.66.Mk, 43.66.Dc [KWG]

Pages: 1691–1700

I. INTRODUCTION

Traditional views of speech perception have often involved the extraction of particular features of the signal, such as frequency transitions of particular formants. While there is little doubt that these features can provide important cues, work performed within the past decade or two has dramatically increased our appreciation of the spectro-temporal analysis that is normally associated with the perception of speech. In this view, speech (or any complex dynamic signal) can be considered as an assembly of temporal patterns at different spectral loci. Although work associated with this view has generally involved basic research and normal-hearing (NH) listeners, it has been motivated in part by cochlear implants, which present a limited number of temporal patterns to corresponding fixed locations along the cochlea. This increased appreciation of a spectro-temporal analysis is related to a greater understanding of the temporal information provided by the fluctuating amplitude patterns of speech (e.g., Van Tasell *et al.*, 1987; Rosen, 1992) and the ability of listeners to understand speech represented by a small number of temporal patterns at different spectral frequencies (e.g., Dudley, 1939; Shannon *et al.*, 1995).

However, one aspect of processing that has received far less attention involves the influence that synchrony of the

temporal patterns at different loci has on speech perception. This is potentially a topic of considerable importance, because it provides information regarding the use of across-frequency timing cues during normal speech perception as well as robustness of these cues to degradation. Another potentially important implication of work involving asynchronous speech patterns involves speech perception in reverberant environments. In these environments, a desynchronization of energy across the spectrum can result from the different absorption rates of low versus high frequencies and the presence of low frequencies at longer delays (cf. Knudson and Harris, 1978). Although it is known that reverberation can interfere with intelligibility, especially for listeners with hearing impairment (cf. Moore, 1998), the mechanisms underlying this disruption are not well understood.

Another implication is more practical—work involving vocoder simulations of speech (also termed acoustic models of cochlear implants) can be influenced by across-frequency asynchrony. In this work, the speech spectrum is partitioned into bands, the temporal envelope of each band is used to amplitude modulate a corresponding carrier, and the modulated carriers are summed and presented to listeners. Although frequency-specific delays introduced during filtering to partition the speech spectrum can be corrected (e.g., Healy and Warren, 2003) or eliminated (e.g., Healy and Bacon, 2006), they are often present, causing asynchrony across the temporal patterns at various frequencies.

^{a)}Author to whom correspondence should be addressed. Electronic mail: ewh@sc.edu

That said, a small amount of work describing the influence of asynchrony on speech perception has been performed. In an early study, Flanagan (1951) examined the influence of delaying one portion of the speech spectrum relative to the remaining portions. It was found that asynchrony up to 60 ms had no effect on word recognition or speech quality. However, these results may be difficult to interpret, as either the delayed or the nondelayed portion of the spectrum was capable of providing substantial recognition in most conditions when presented alone.

Much of the current understanding of the influence of asynchrony is attributable to Greenberg and Arai (for review, see 2004). In an initial study (Arai and Greenberg, 1998), the broadband speech spectrum was partitioned into nineteen 1/4-octave bands and a random delay was introduced to each. The mean delay across frequency channels was one half of the maximum delay, and adjacent channels had delays of at least one quarter of the maximum. Sentence intelligibility was found to be relatively unaffected by maximum asynchronies up to 80 ms and to remain high (above approximately 75%) up to 140 ms. These data suggested that the auditory system was highly tolerant of across-frequency asynchrony and that across-frequency timing was relatively unimportant to sentence perception. However, a subsequent experiment was designed to examine whether this high intelligibility could be attributed to subsets of bands across the spectrum that had relatively small misalignments at any given maximum asynchrony. In that experiment (Greenberg *et al.*, 1998), the sentences were represented by only four 1/3-octave bands spanning the spectrum. Due perhaps in part to the difficulty of the sentence materials (the multiple-speaker TIMIT corpus), each band provided little intelligibility when presented individually. It was found that scores were reduced approximately 15% when the two central channels led or lagged the two lateral channels in time by 25 ms. Scores dropped approximately 50% at 75 ms. This influence of asynchrony was somewhat smaller when one central band was offset relative to the other three. Thus, when the signal was represented by a smaller number of bands, presumably increasing the role of across-frequency timing, the influence of cross-channel asynchrony was far larger.

In a similar set of experiments, Fu and Galvin (2001) partitioned broadband sentences into either four or 16 bands, and introduced delays in staggered fashion so that adjacent bands had greater misalignments than alternate bands and so that higher frequency bands were generally delayed relative to lower frequency bands. It was found that intelligibility remained high even with a maximum delay of 240 ms. However, as with the study by Arai and Greenberg (1998), this apparent tolerance for across-frequency asynchrony may be attributed to subsets of relatively synchronous information. In Fu and Galvin, as the authors correctly indicate, this information was likely provided by intact spectral information within each speech band. In another experiment, Fu and Galvin employed a speech-vocoder technique to eliminate spectral information from each band. This had the effect of eliminating the intelligibility of individual component bands and isolating comparisons across bands (analogous to filtering to an extremely narrow band). When spectral contrasts

were controlled in this manner, the influence of asynchrony was found to be substantially increased. When the spectrum was represented by only four contiguous bands, scores fell with as little as 40 ms of maximum channel delay (22 ms average). The 16-channel condition remained more tolerant of asynchrony, likely due to the increase in spectral detail and decreased reliance upon accurate across-frequency timing cues that may result with a larger number of frequency channels.

Other recent work extends this concept that fewer frequency channels produce increased reliance on relative timing. Lippmann (1996) demonstrated accurate recognition of consonants despite removal of middle-frequency energy. The resulting stimulus then provided synchronous cues across low-pass and high-pass frequencies. Healy and Warren (2003) employed a speech vocoder technique and showed that a synchronous pair of narrowband temporal patterns was sufficient to allow accurate sentence recognition, despite the fact that intelligibility was at or near zero when the patterns were presented individually. Thus, intelligibility can be attributed directly to integration across the pair. Although other cues such as those associated with overall spectral balance may be present, one obvious cue provided by the contrasting pair, which is absent when single patterns are presented, involves across-frequency timing. The synchronous pair of temporal patterns provides spectral information in its most reduced form, by allowing the direction and rate of change to be encoded as energy sweeps through the spectrum. Although these results suggest that listeners rely upon across-frequency timing when presented with stimuli that generally lack other cues, the use of this cue is not well understood.

In the current experiments, these previous observations concerning the role of across-frequency timing in speech perception and the tolerance for disruptions to this cue were extended across the speech spectrum. The use of speech-modulated carrier bands, rather than speech bands, allowed the elimination of spectral information within the bands, complete elimination of individual-band intelligibility, and the strict isolation of comparisons across bands where the effects of asynchrony could be directly observed. Further, the use of only two bands provided the maximum potential for asynchrony to disrupt intelligibility. This also allowed asynchrony to be precisely specified by a single delay value. The following questions were addressed in experiment 1: (1) Is the use of across-frequency timing cues and the influence of their disruption similar when examined separately in the low-, mid-, and high-frequency regions? Because the nature of speech cues changes dramatically across various regions of the spectrum, it is conceivable that the use of across-frequency timing will differ in different spectral regions. (2) Is the use of across-frequency timing cues similar when examined across different frequency separations? Again, due to differences in speech cues as well as potential differences in the physiological encoding of signals that are closely versus widely spaced, it is conceivable that the use of relative timing will differ across different separations. In experiments 2 and 3, the influence of frequency separation was examined in greater detail. Specifically, the effect of peripheral excitation

overlap on the influence of between-band asynchrony was investigated in experiment 2. In experiment 3, potential differences in the nature and use of timing cues were examined.

II. EXPERIMENT 1: TOLERANCE FOR ASYNCHRONY ACROSS THE SPEECH SPECTRUM

A. Method

1. Subjects

A total of 40 native-English speaking students participated and received either course credit or money in compensation. All had pure-tone audiometric thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz (American National Standards Institute, 2004). The average age of all subjects tested was 24 years and 91% were female. Care was taken to ensure that none had prior exposure to the sentence materials.

2. Stimuli

The stimuli were based upon the 100 Central Institute for the Deaf (CID) everyday speech sentences (Davis and Silverman, 1978). They were produced by a male speaker having a standard American Dialect and recorded at a 22 kHz sampling rate with 16 bit resolution. An additional set of ten practice sentences were drawn from the high-predictability subset of the Speech Perception in Noise test (SPIN, Kalikow *et al.*, 1977). In each condition, these sentences were represented by a pair of narrowband temporal patterns. The preparation of these patterns has been described in detail elsewhere (Healy and Bacon, 2002; Healy and Warren, 2003). Briefly, each pattern was created by first filtering the speech to a 1/3-octave band using a cascaded pair of eighth-order Butterworth filters (96 dB/octave). The filtered speech was then full-wave rectified and multiplied on a sample point-by-point basis with a pure tone having a frequency matching the center frequency of the speech band. The amplitude-modulated tones were then restricted to the original frequency region of origin by again filtering with the same filters used to create the corresponding speech bands. This processing, in which fluctuation rate was limited by the bandwidth of spectral filtering rather than to a fixed arbitrary value using a low-pass smoothing filter, ensured that the temporal patterns fluctuated in the same fashion as would corresponding narrowbands of speech. All of this processing was performed in MATLAB. The peak of the slow-response rms average of each sentence was scaled to play back at 70 dB SPL in a headphone coupler (Larson Davis AEC 101 and 800B). The center frequencies of the bands were 530, 1100, 2100, and 4200 Hz.

The temporal speech patterns were mixed in pairs for presentation to listeners. A total of four band pairs were created: one in the low-frequency region (530 and 1100 Hz), one in the high-frequency region (2100 and 4200 Hz), one in the mid-frequency region having a frequency separation of approximately one octave (1100 and 2100 Hz), and one having a frequency separation of approximately three octaves (530 and 4200 Hz). The frequency-specific phase shifts of the filters were evaluated by filtering a single-sample click using the cascaded Butterworth filters at each center fre-

quency and measuring the temporal offset between the centers of the resulting filtered pulses. Measured values were doubled to account for both speech and carrier filtering. These delays were corrected by delaying the high-frequency band relative to the low-frequency band, and the bands were digitally mixed in temporal synchrony. The correction for the low-frequency band pair was 10 ms, that for the high-frequency pair was 2 ms, that for the mid-frequency one-octave separation was 4 ms, and that for the three-octave separation was 16 ms. Additional conditions were created in which the bands were misaligned by 12.5, 25, 50, and 100 ms. Both conditions in which the high-frequency band was delayed relative to the low-frequency band and the opposite offset were prepared.

3. Procedure

Separate groups of ten subjects each were assigned to the four band-pair conditions, and each listener heard each of the between-band asynchronies employed. Half the subjects in each group heard the high-frequency band delayed relative to the low-frequency band and the other half heard the opposite offset. The sentence list-to-condition correspondence was balanced so that each list was heard in each condition an equal number of times, and the order in which the conditions were heard was randomized for each listener. Subjects first heard the ten SPIN practice sentences broadband, and then again in the aligned condition. The practice list was played again processed with an asynchrony prior to the presentation of 20 CID test sentences in that given asynchrony condition. Test sentences were played only once. Subjects were tested individually in a sound-attenuating booth seated with the experimenter. They were instructed to repeat as much of each sentence as they could and were encouraged to guess if unsure of the content. The experimenter controlled the delivery of sentences and recorded the number of standard scoring keywords recalled correctly. The files were played back from a PC using Edirol Gina 24 digital to analog converters and delivered diotically over Sennheiser HD 250II headphones.

B. Results

1. Experiment 1a: Direction of asynchrony

First, it is important to recall that these very temporal patterns yield CID sentence intelligibility values at or near zero when presented individually across a wide range of frequencies (Healy and Warren, 2003). Group mean intelligibility scores for the current stimuli having components at 1100 and 2100 Hz are presented in Fig. 1. As the figure shows, the two offsets produced nearly identical intelligibility scores. These data also show that intelligibility was substantially affected by across-frequency asynchronies as small as 12.5 or 25 ms, and was reduced to values near the floor at 100 ms. A two-factor (2 offset conditions X 5 asynchronies) mixed analysis of variance (ANOVA) was employed to confirm these observations: The effect of offset was nonsignificant [$F(1, 8)=0.3, p=0.63$] as was the interaction between offset and asynchrony [$F(4, 32)=0.8, p=0.53$]. The main effect of asynchrony was significant [$F(4, 32)=190.8, p<.01$] and post hoc testing using the Holm-Sidak method indicated that

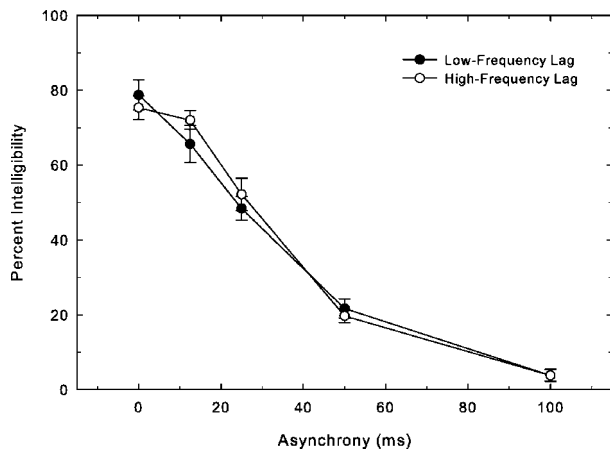


FIG. 1. Group mean intelligibilities and standard errors for CID sentences represented as a pair of narrowband temporal speech patterns at 1100 and 2100 Hz. The patterns were presented in temporal synchrony and at the asynchronies indicated. Half the subjects heard the low-frequency band lag relative to the high-frequency band, and the other half heard the opposite offset.

performance at all asynchronies differed ($p < .05$). Because they produced equivalent scores, data from the two offsets were pooled for all subsequent analyses.¹

2. Experiment 1b: Frequency region

In the upper panel of Fig. 2, data from the group hearing the band pair in the low-frequency region (530 and 1100 Hz) were compared to those hearing the pair in the high-frequency region (2100 and 4200 Hz). As shown, intelligibility as a function of asynchrony was nearly identical across the two regions. A two-factor mixed ANOVA was again used to confirm these observations: The effect of frequency region was not significant [$F(1, 18)=0.2, p=0.66$] nor was the interaction [$F(4, 72)=0.6, p=0.67$]. The main effect of asynchrony was again significant [$F(4, 72)=94.5, p < 0.01$] and post hoc testing (Holm-Sidak) indicated that performance at all asynchronies differed ($p < .05$).

3. Experiment 1c: Frequency spacing

Although there are three sets of bands separated by approximately one octave (those in the low-, mid-, and high-frequency regions), the most appropriate for comparison to the three-octave separation conditions is that in the mid-frequency range. These sets of band pairs both straddle the approximate center of the spectrum at 1500 Hz, and both produced similar overall levels of performance. Shown in the lower panel of Fig. 2 are scores for the mid-frequency band pair separated by approximately one octave (1100 and 2100 Hz), averaged across the two offsets and replotted from Fig. 1, and the pair separated by approximately three octaves (530 and 4200 Hz). Although intelligibility produced by the one-octave pair drops considerably over the initial 25 ms of asynchrony, scores for the three-octave pair are reduced by a smaller amount. A two-factor mixed ANOVA revealed that the interaction between frequency region and asynchrony was significant [$F(4, 72)=4.3, p < 0.01$]. Post hoc analysis of this interaction (Holm-Sidak) indicated that the different frequency spacings produced equivalent scores at asynchronies

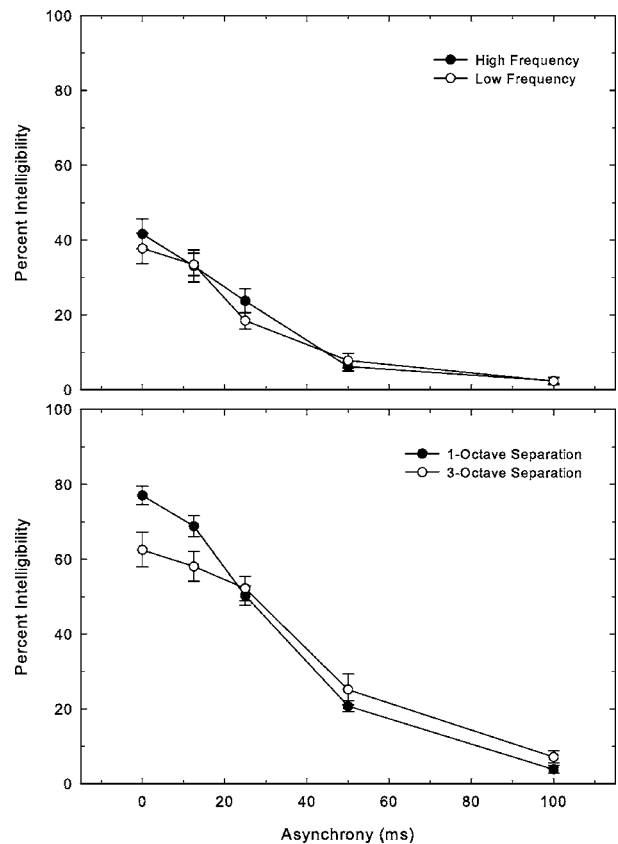


FIG. 2. The upper panel shows mean intelligibilities and standard errors for separate groups of listeners, one who heard a pair of narrowband temporal speech patterns from the low spectral region (530 and 1100 Hz) and another who heard a pair from the high spectral region (2100 and 4200 Hz). The patterns were presented in temporal synchrony and at the asynchronies indicated. The lower panel shows corresponding scores for two additional groups of listeners who hear pairs of narrowband temporal speech patterns having spacings of approximately one octave (1100 and 2100 Hz) or three octaves (530 and 4200 Hz), both straddling the center of the speech spectrum at approximately 1500 Hz.

of 25 ms and above, but different scores ($p < .05$) in the synchronous condition and at 12.5 ms. The main effect of asynchrony was also significant [$F(4, 72)=166.0, p < 0.01$] and all asynchronies differed (Holm-Sidak, $p < 0.05$). The main effect of frequency spacing was not significant [$F(1, 18)=2.5, p=0.13$].

An additional analysis examined the influence of asynchrony across all three band pairs having a separation of approximately one octave. When scores for all four conditions were expressed as proportions of scores in the aligned conditions to eliminate differences in overall level of performance, a two-factor mixed ANOVA revealed a significant main effect of band pair [$F(3, 36)=4.2, p < 0.05$], but post hoc testing (Holm-Sidak) indicated that performance only differed ($p < 0.05$) for the band pair having the three-octave separation. Thus, the influence of asynchrony was equivalent for conditions having a separation of one octave, but different for the three-octave separation.²

C. Discussion

In accord with Greenberg *et al.* (1998) and Fu and Galvin (2001), these data indicate that NH listeners can be

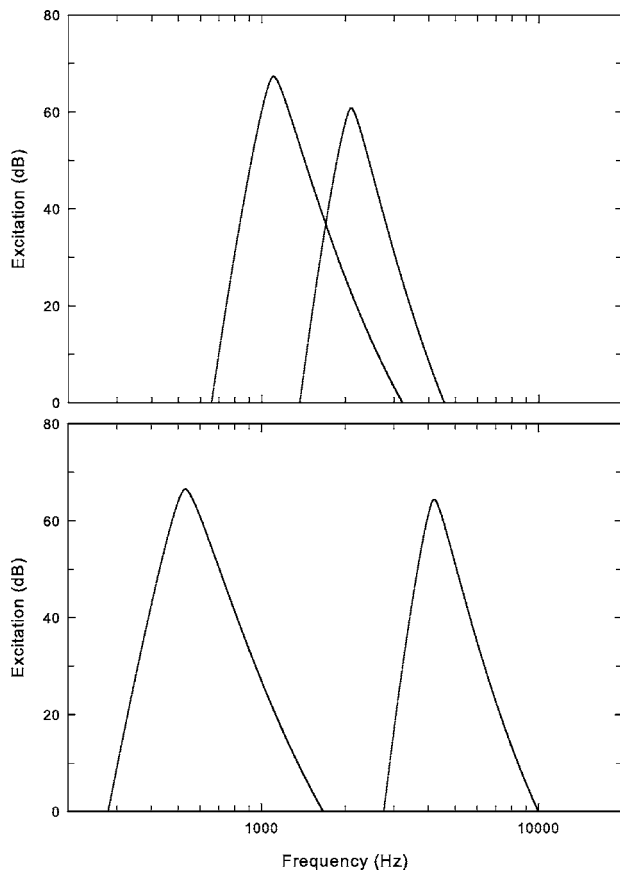


FIG. 3. Patterns of peripheral excitation produced by 70 dB SPL pure tones at 1100 and 2100 Hz (top panel) and 530 and 4200 Hz (bottom panel).

highly dependent upon preservation of across-frequency synchrony under conditions in which this information is emphasized. Also in accord with Greenberg *et al.*, performance was not affected by the direction in which across-frequency asynchrony occurred. The current results extend these previous observations by indicating that the influence of asynchrony is similar across various regions of the speech spectrum, suggesting that the use of relative timing cues is similar. Further, it was found that speech information that was more widely spaced in frequency was more resistant to the disruptive effects of small amounts of asynchrony relative to patterns having a smaller spacing. This latter result suggests that the reliance upon detailed across-frequency timing cues might be less for patterns having larger frequency separations than it is for patterns that are more closely spaced.

A number of possible interpretations exist for this observed effect of frequency spacing, in which the band pair having the wider frequency separation is less affected by small amounts of asynchrony. One possibility is that asynchrony is especially disruptive to overlapping portions of excitation, where the individual fluctuations normally combine to form complex patterns. An obvious difference between separations involves overlap: While the excitation produced by the band pair having the large frequency separation presumably does not overlap in the auditory periphery, excitation of the other band pair does substantially (see Fig. 3).³ In the experiment that follows, this effect of frequency spacing is explored in more detail. Specifically, it is examined

whether the effect can be attributed to overlap of excitation in the auditory periphery of the one-octave spaced band pair, which is absent in the three-octave pair.

III. EXPERIMENT 2: EFFECT OF FREQUENCY SPACING AND PERIPHERAL EXCITATION OVERLAP

A. Method and results

An additional 24 NH listeners were recruited using qualification and compensation procedures employed in experiment 1. They were divided into three groups of eight subjects each. Each listener heard the mid-frequency bands spaced by one octave and three octaves in synchrony and also at an asynchrony of 25 ms. Thus, in this experiment, the comparison between frequency separations was made within subjects. Each listener heard 20 CID sentences in each condition, and the low-frequency band was delayed relative to the high-frequency band for all listeners. The presentation order of conditions was balanced so that each was heard an equal number of times in each serial position, and the condition-to-sentence list correspondence was balanced so that each list appeared in each condition an equal number of times. Half the subjects in each group heard the one-octave separation conditions followed by the three-octave separation conditions, and the other half heard the opposite order.

The procedure was modified in this experiment to provide additional practice prior to hearing the test conditions. The session started with the SPIN sentences, heard first broadband, then again in synchrony as either the one- or three-octave pair, depending upon which test conditions were heard first. Subjects then heard ten CID sentences represented as a one- or three-octave pair at an asynchrony of 12.5 ms. Following this practice, subjects heard the 0- and 25 ms asynchronies in balanced order. Practice was repeated prior to the remaining frequency-spacing condition. Test sentences were presented only once. Other procedures were identical to experiment 1.

It was first important to confirm the presence of the frequency spacing effect observed across groups in experiment 1 within a single group of subjects. A first group of subjects heard the stimuli in a quiet background. As panel (b) of Fig. 4 shows, a pattern of results similar to that seen in experiment 1 was observed: The band pair having the smaller frequency separation was more affected by asynchrony than was the pair having the larger separation. A two-factor repeated-measures ANOVA revealed, most importantly, that the interaction between frequency separation and asynchrony was significant [$F(1,7)=5.6, p<0.05$]. The main effects were also significant [for frequency separation, $F(1,7)=10.1, p<0.05$, for asynchrony $F(1,7)=133.9, p<0.01$].

Performance was then assessed as the overlapping portion of excitation was masked by a continuous narrowband noise masker. Low-noise noise (LNN, Pumplin, 1985; Hartmann and Pumplin, 1988) was employed to eliminate the large fluctuations in intensity that accompany narrowband filtering of Gaussian noise. The method employed to generate LNN (Method 1 from Kohlrausch *et al.*, 1997) involved iteratively (1000 in this case) dividing a noise waveform by its envelope. Sinusoidal components having flat amplitude

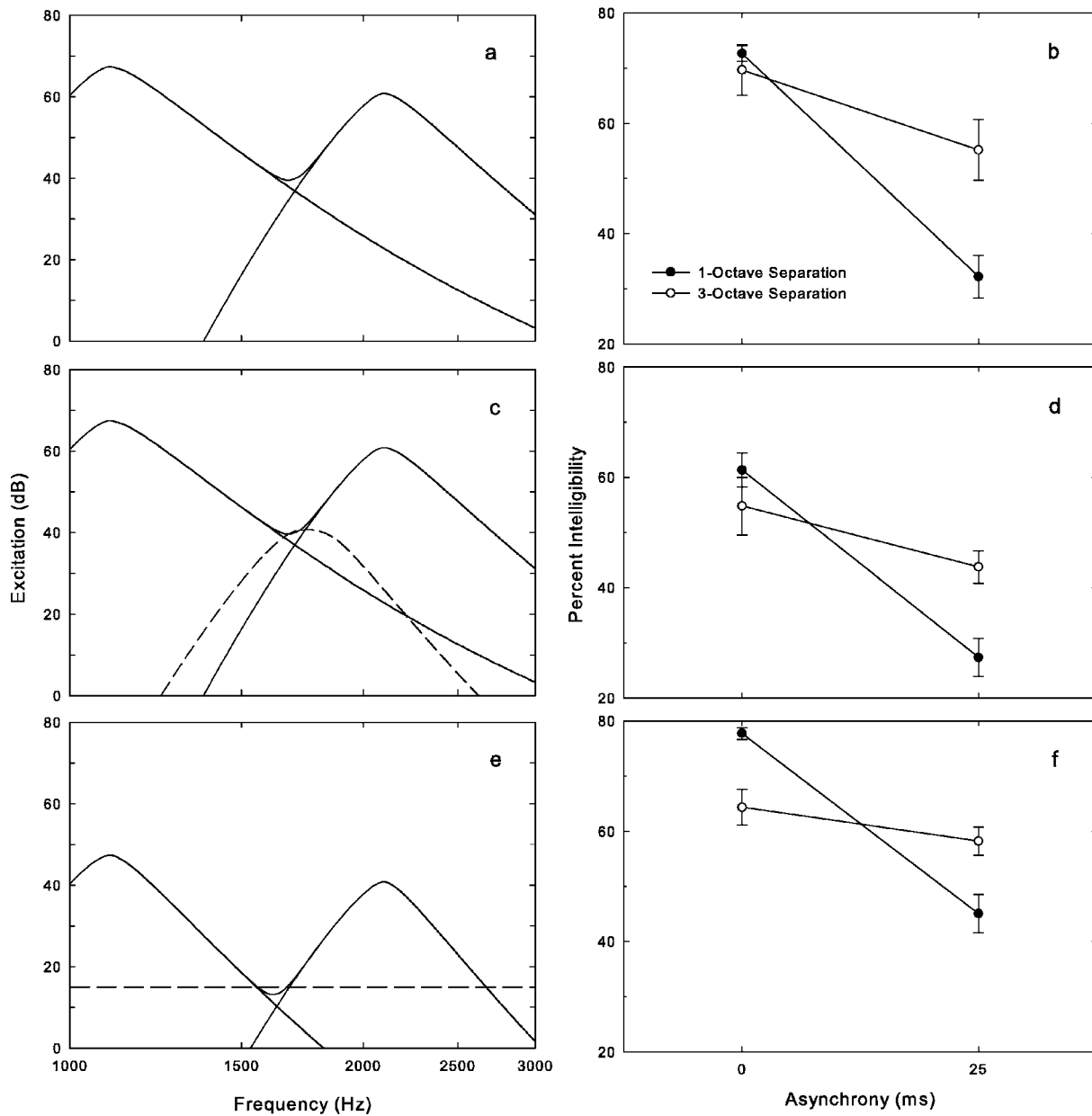


FIG. 4. Panel (a) shows patterns of peripheral excitation (Glasberg and Moore, 1990) for pure tones at 1100 and 2100 Hz when presented individually and simultaneously at 70 dB SPL. This level of excitation overlap was found to be within 1 dB of that produced by the same tones having the dominant AM sidebands representing the speech-modulated tones in experiment 2. Panel (b) shows mean intelligibilities and standard errors for listeners who heard pairs of narrowband temporal speech patterns having spacings of approximately one (1100 and 2100 Hz) and three octaves (530 and 4200 Hz), and between-band asynchronies of 0 and 25 ms. In panel (c), the patterns elicited by the tones are shown again, along with excitation produced by a narrowband low-noise noise masker (dashed line). In panel (d) are corresponding scores from listeners hearing the speech-modulated stimuli along with this noise. In panel (e) are excitation patterns for the pure tones when presented individually and simultaneously at a level of 50 dB SPL. The dashed line in this panel is at 15 dB SPL, which indicates approximate audiometric thresholds. Panel (f) shows corresponding scores from listeners hearing these stimuli.

and appropriate phase were assembled at a spacing of 0.5 Hz. This LNN masker had a bandwidth of one equivalent rectangular band (ERB_N , Glasberg and Moore, 1990; Moore, 2003) and spanned 1664 to 1879 Hz. It was presented continuously along with the speech patterns at a spectrum level of 25 dB SPL, which yielded a masker-to-signal overlap excitation ratio of 0 dB [see Fig. 4(c)]. Despite the presence of this masker, the characteristic pattern of results in which the smaller frequency separation was more affected by asynchrony was again observed [Fig. 4(d)]. A two-factor repeated-measures ANOVA revealed that the interaction was significant [$F(1, 7)=8.5, p<0.05$]. The main effect of asyn-

chrony was also significant [$F(1, 7)=29.0, p<0.01$], but the main effect of band separation was not [$F(1, 7)=3.8, p=0.09$].⁴

An additional group of listeners was employed to confirm this result. In these conditions, the speech-modulated stimuli were presented without noise, but at a slow-response rms peak level of 50 dB SPL. This lower level provided both less spread of excitation of the signal and an attenuation of the excitation overlap. As Fig. 4(e) shows, the peak of the peripheral excitation overlap was reduced in this condition to a level approximating that of normal audiometric thresholds

in these middle frequencies (15 dB SPL). Scores from the low-level conditions are presented in Fig. 4(f). As before, the smaller separation was more affected by asynchrony [interaction: $F(1,7)=20.0, p<0.01$]. The main effect of asynchrony was significant [$F(1,7)=61.2, p<0.01$], but the main effect of band separation was not [$F(1,7)=0.0, p=0.96$].

B. Discussion

The effect of interest in these experiments was the interaction between frequency spacing of the band pair and the influence of asynchrony. In accord with the across-subjects results of experiment 1, it was observed in experiment 2 that the band pair having the smaller frequency separation was more disrupted by asynchrony than was the pair having the larger separation. Excitation from signals that are close in frequency overlaps to form complex patterns representing the simultaneous usually synchronous signals. Asynchrony may potentially disrupt this normal mixing and affect signals that are spaced more closely differently from signals having a wider separation and lacking mixed patterns of peripheral excitation. However, the characteristic interaction between frequency spacing and asynchrony was observed when the overlapping portion of excitation was masked or when it was presumably inaudible. These results therefore indicate that the greater impact of asynchrony observed for the bands having the smaller frequency separation is not attributable to disruption of the complex mixture that occurs in the region of peripheral excitation overlap.

The question therefore remains regarding the mechanism accounting for this observed effect of frequency separation. One possibility involves the nature of speech information across the different frequency regions. It is possible that the speech cues that exist at more proximate frequencies are based more on exact timing (e.g., formant transitions encoding place of articulation) and thus are more disrupted by small amounts of asynchrony than are cues across more disparate frequencies. This difference in the existence of speech cues can be thought of as a *nature-of-cues* interpretation. An alternative involves differences not in the nature or existence of these cues, but instead in their use (or perhaps their availability). It is possible that the acoustic cues themselves are not less disrupted at larger separations, but instead that the mechanism employed to process disparate information is one which does not rely as heavily upon (or perhaps does not have available) the details of across-frequency timing. This can be thought of as a *use-of-cues* interpretation.

Although it is potentially difficult to address the first interpretation empirically, a quantitative measure of acoustic signal degradation resulting from across-frequency asynchrony has been proposed (Healy *et al.*, 2005). This measure involves the change in correlation of the fluctuating acoustic envelopes across two speech frequencies. Although the exact correlation value was not of particular interest, the reduction in this correlation as a function of between-band asynchrony was hypothesized to reflect signal degradation resulting from desynchronization of the patterns and, indeed, was found to correspond closely to reductions in sentence intelligibility for NH listeners. This correlation analysis was employed in ex-

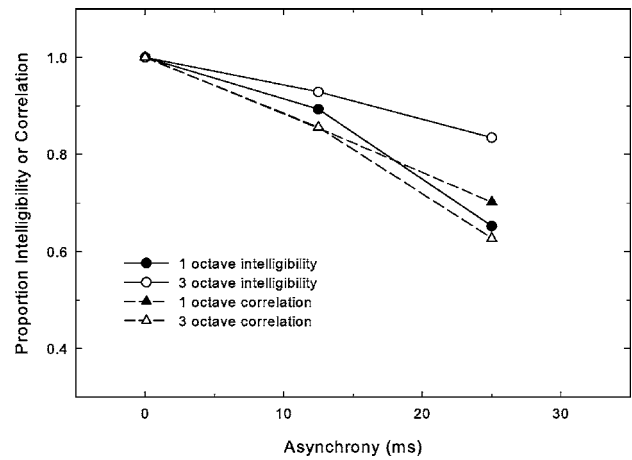


FIG. 5. Intelligibility at two band separations from experiment 1 and correlations of the corresponding acoustic temporal patterns. All values are expressed as proportions of the values in the synchronous conditions.

periment 3a. In experiment 3b, a use-of-cues interpretation was assessed using dichotic presentation in which the nature of the acoustic speech cues was held constant but their presentation changed.

IV. EXPERIMENT 3: USE AND NATURE OF TIMING CUES ACROSS FREQUENCY SEPARATIONS

A. Experiment 3a: Correlation analysis

The analysis employed by Healy *et al.*, 2005 was applied to the current conditions by correlating on a sample point-by-point basis the acoustic envelopes of the temporal speech patterns comprising the mid-frequency one-octave and three-octave conditions. Envelopes of the modulated patterns at each frequency were extracted using full-wave rectification and low-pass filtering at 100 Hz (100 dB/octave). The correlation was averaged for each sentence and direction of offset and was performed with the pair of speech patterns in synchrony and also at 12.5 and 25 ms of asynchrony (see Healy *et al.*, 2005 for details). Figure 5 shows intelligibility data replotted from the lower panel of Fig. 2, along with the corresponding acoustic correlations, all plotted as proportions of the values in the synchronous conditions to highlight changes in performance or correlation. Two effects are apparent. First, the reduction in acoustic correlation resulting from desynchronization is similar for the two frequency separations. Thus, according to this particular measure, asynchrony was no more or less disruptive to the acoustic signal at the one-octave separation than at the three-octave separation. Second, the manner in which the correlation was reduced closely resembled the reduction in intelligibility for the one-octave separation. This suggests that detailed across-frequency timing cues were used in this condition, as their disruption was accompanied by a corresponding reduction in performance. In contrast, performance at the wider frequency separation was more stable, despite the apparent disruptions to the acoustic signal over the first 25 ms of asynchrony. Because this relative stability can be attributed to lower performance in the synchronous condition, where across-frequency timing was intact (see the lower panel of Fig. 2), it

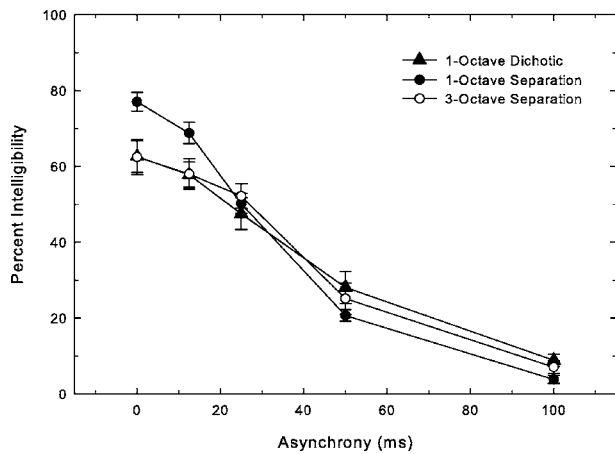


FIG. 6. As Fig. 2 lower panel, with the addition of mean intelligibility scores and standard errors for a group of ten listeners who heard the bands comprising the one-octave separation condition delivered to opposite ears.

is suggested that the processing of the disparate patterns did not rely as heavily on details of across-frequency timing to improve performance and, correspondingly, was not hindered by their disruption.

B. Experiment 3b: Dichotic presentation

In this experiment, bands comprising the mid-frequency one-octave separation were split for delivery to opposite ears. These conditions provide a test of the use-of-cues interpretation, as the nature of the cues provided by the acoustic patterns did not change. A group of ten listeners was recruited using selection criteria employed previously and experimental procedures were identical to those of experiment 1. In addition, the presentation ear was balanced such that half the subjects heard the low-frequency band in the right ear and the high-frequency band in the left ear. The results are displayed in Fig. 6. Under dichotic stimulation, the very same one-octave stimuli produced lower scores in the synchronous condition and were far less sensitive to small disruptions in across-frequency timing. These results suggest that listeners exposed to dichotic stimulation were less able to use the intact across-frequency timing cues to improve performance and were less hindered by their disruption.

V. GENERAL DISCUSSION

When speech is reduced to energy fluctuations occurring in only two narrow bands of frequencies, either one of which provides no intelligibility individually, the remaining cues are likely associated to some extent with across-frequency timing. The use of these cues and the tolerances for their disruption were the focus of the current study. For bands in the middle of the spectrum and separated by one octave, it was found that intelligibility fell sharply when relatively small disruptions were made in across-band synchrony, indicating that listeners were relying heavily upon the detailed across-frequency timing information provided by the synchronous temporal patterns. Scores fell when as little as 12.5 ms of asynchrony was introduced, and 100 ms was sufficient to disrupt intelligibility almost completely. This latter result indicates that an asynchrony value of 100 ms appears

capable of completely disrupting the across-frequency timing mechanism in NH listeners. Due to greater absorption of higher frequencies, reverberation causes lower frequencies to be present at greater delays. Because of this experience with asynchrony of direct and indirect sounds, it might be assumed that listeners would be more tolerant of conditions in which low frequencies were delayed relative to highs. However, no difference was observed when either the high-frequency band or the low-frequency band of the pairs was delayed, indicating symmetry in the effect of asynchrony.

The current study also allowed an examination of the use of timing cues across various frequency regions and separations. Disruptions in across-frequency timing had similar influences on performance for bands separated by roughly one octave, whether these bands were located together in the low-, mid- or high-frequency region, suggesting that the way in which across-frequency timing cues are used is similar across these varied regions. This result may be considered surprising given the large differences that exist in the nature of speech cues in different frequency regions.

Some information regarding the influence of across-frequency asynchrony on specific classes of speech cues is available. Stone and Moore (2003) found that delaying speech frequencies below approximately 1000 Hz by as little as 15 ms reduced transmission of both manner and place of articulation cues for hearing-impaired listeners. However, voicing cues were unaffected by these small disruptions. This latter result is in accord with those of Yuan *et al.* (2004) who reported that a cue to voicing involves differences in the timing of envelope onsets across disparate frequency regions (below 350 Hz and above 3000 Hz), but that the average difference across the voiced/unvoiced contrasts was relatively large (approximately 150 ms).

There were differences in the influence of small amounts (up to 25 ms) of asynchrony when the frequency spacing between bands straddling the center of the speech spectrum was increased from one to three octaves. While performance for bands separated by approximately one octave fell sharply with small amounts of asynchrony, performance for bands separated by approximately three octaves changed far less.⁵

While peripheral excitation overlap appeared to be a viable interpretation, the results of experiment 2 suggest it is not correct. Instead, the results of experiment 3 suggest that the processing of speech patterns at different separations may differ in the extent to which cues associated with details of across-frequency timing are employed. The lack of correspondence between acoustic signal degradation and performance, together with an apparent insensitivity to the slight disruptions that exist in the acoustic cues at the larger frequency separation (Fig. 5) provide little support for a nature-of-cues interpretation, but instead is in accord with a use-of-cues interpretation.

These results, in which speech-like bands having larger frequency separation are less sensitive to disruptions in fine timing, are at least somewhat in accord with results employing nonspeech stimuli. It has been shown that the comparison of relative timing of acoustic events can be poorer when that comparison is made across disparate frequencies. For example, gap detection thresholds are larger when the mark-

ers preceding and following the gap have different frequencies (e.g., Perrott and Williams, 1971; Formby *et al.*, 1998; Grose *et al.*, 2001a, 2001b; Lister *et al.*, 2002; Lister and Roberts, 2005; Heinrich and Schneider, 2006). Although frequency separations smaller than those used in the current study are often employed, data describing large frequency separations have shown correspondingly large thresholds (e.g., Fig. 2 in Phillips *et al.*, 1997). Because the acoustic stimuli employed in those tasks do not change as a function of frequency (the nature of cues does not change), these results are also in accord with a use-of-cues interpretation.

It has been proposed that gap detection is accomplished using two mechanisms. A *within-channel* mechanism accounts for detection across markers separated by as much as half an octave to an octave, and a *between-channel* process accounts for detection across larger frequency differences (see Formby *et al.*, 1998). Support for the between-channel process was provided by Formby *et al.* by examining dichotic gap detection. It was found that thresholds were relatively unaffected by the frequency of dichotic markers and were virtually identical to those observed for monaural presentation of markers having disparate frequencies. The authors concluded that the between-channel mechanism involves comparison of independent channels across which exact timing information is not available or not used.

The results of experiment 3b are in accord with this view. A pattern of results more closely resembling those of the three-octave separation was observed when the one-octave separation was presented dichotically. Because the nature of the acoustic cues is similar under diotic and dichotic presentation, differences in performance as a function of asynchrony must result from difference in the use or availability of the cues.

Together, these results support a use-of-cues interpretation for the observed difference in the impact of between-band asynchrony across different frequency separations. The processing of speech energy that is relatively close in frequency appears to rely more heavily upon detailed across-frequency timing cues than does the processing of speech information that is more widely separated. Further, these results suggest that the integration of speech information across various frequencies may be accomplished using dual mechanisms like those proposed for gap detection. Comparisons in which exact timing of events is available and utilized, such as those associated with within-channel tasks, may be associated with processing of relatively adjacent speech frequencies. However, the between-channel comparison of disparate speech frequencies appears to proceed without this heavy reliance on, and same strict regard for, exact timing.

Because intelligibility in the current conditions requires across-frequency integration, these conditions provide the potential to observe the maximally disruptive effects of asynchrony. As such, these values form a basis on which to evaluate the potentially disruptive influences of cross-channel delays. The current results indicate that signal processing delays of 10 ms or more, such as those associated with cascaded filters, could potentially have an appreciable effect on performance. Of course these values represent the potential

for disruption, and the actual influence on listener performance would depend upon a number of factors, including the amount of other concurrently available speech cues.

VI. SUMMARY AND CONCLUSIONS

In the current study, pairs of temporal speech patterns were employed that produced little or no intelligibility individually and therefore required across-frequency integration for accurate recognition. An obvious cue to recognition of these stimuli therefore involves across-frequency timing, and it seems reasonable to assume that this cue was used to some extent. However, the current results showed that tolerances for disruption to exact timing varied, suggesting that the use of these cues differed across conditions in systematic ways. For bands having a relatively small frequency separation (approximately one octave), as little as 12.5 ms of across-frequency asynchrony was sufficient to significantly degrade performance, and 100 ms was sufficient to completely disrupt across-frequency integration. Further, the influence of disruption to across-frequency timing was equivalent across various spectral regions having a fixed (log) frequency separation, suggesting similarity in the use of these cues across the spectrum. In contrast, speech patterns having different frequency separations were found to have different tolerances for disruption to exact across-frequency timing, suggesting that differences exist in the extent to which these cues are used. This effect is not attributable to differences in peripheral excitation overlap. Rather, it appears that the processing of speech patterns at different separations may take place using dual mechanisms that differ in the extent to which they rely upon detailed across-frequency timing cues.

ACKNOWLEDGMENTS

This work was supported in part by NIH/NIDCD Grant Nos. DC05795 (E.W.H.) and DC01376 (S.P.B.) and by the NOHR Foundation (E.W.H.). The authors are grateful for the comments of Ken Grant, Steven Greenberg, and an anonymous reviewer on a previous version of this manuscript.

¹This effect of offset direction was confirmed by examining data from all 40 listeners and all four dual-band conditions of experiment 1. The intelligibility resulting from the opposite offsets was within 1% on average across the five asynchronies.

²If the three-octave separation conditions are instead compared to the low- or high-frequency one-octave pairs (and scores again normalized as proportions), numerical differences between the one- and three-octave separations are slightly larger.

³Excitation patterns throughout were created using the executable program *excit2005*, by Brian R. Glasberg and Brian C. J. Moore, which is available for download at <http://hearing.psychol.cam.ac.uk/>. The program allows the calculation of excitation patterns according to the procedure described in Moore *et al.* (1997), which is in turn a modification of Glasberg and Moore (1990). The program runs on an IBM PC or compatible.

⁴The results were not substantially different when broadband noise was employed to mask the overlapping portion of excitation overlap.

⁵Additional conditions were prepared to ensure that the current results involving speech-modulated carriers could be extended to speech. Narrow speech bands at the center frequencies employed in experiment 2 were created by filtering the sentences to 1/3 octave using extremely steep slopes (6000-order finite-duration impulse response filters). The bands were assembled into one-octave and three-octave pairs having asynchronies of 0 and 25 ms. Four NH listeners having the same qualification criteria employed in the formal experiments were recruited and experienced the pro-

cedures of experiment 2. The difference in mean intelligibility scores between the 0 and 25 ms asynchrony conditions was 29% at the one-octave separation, but 4% at the three-octave separation [interaction $F(1,3) = 99.9, p < 0.01$].

Arai, T., and Greenberg, S. (1998). "Speech intelligibility in the presence of cross-channel spectral asynchrony," *Proceedings of the IEEE International Conference on Acoustics and Speech Signal Processing*, pp. 933–936.

American National Standards Institute. (2004). *ANSI-S3.6-2004, Specifications for Audiometers* (New York).

Davis, H., and Silverman, S. R. (1978). *Hearing and Deafness*, 4th ed. (Holt, Rinehart, and Winston, New York).

Dudley, H. (1939). "The vocoder," *Bell Lab. Rec.*, **17**, 122–126.

Flanagan, J. L. (1951). "Effect of delay distortion upon the intelligibility and quality of speech," *J. Acoust. Soc. Am.* **23**, 303–307.

Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (1998). "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection," *J. Acoust. Soc. Am.* **103**, 3554–3560.

Fu, Q.-J., and Galvin, J. J. III. (2001). "Recognition of spectrally asynchronous speech by normal-hearing listeners and Nuclues-22 cochlear implant users," *J. Acoust. Soc. Am.* **109**, 1166–1172.

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

Greenberg, S., and Arai, T. (2004). "What are the essential cues for understanding spoken language?," *IEICE Trans. Inf. Syst.* **E87-D(5)**, 1059–1069.

Greenberg, S., Arai, T., and Silipo, R. (1998). "Speech intelligibility derived from exceedingly sparse spectral information," *Proceedings of the Fifth International Conference on Spoken Language Processing*, pp. 74–77.

Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. (2001a). "Gap detection for similar and dissimilar gap markers," *J. Acoust. Soc. Am.* **109**, 1587–1595.

Grose, J. H., Hall, J. W., and Buss, E. (2001b). "Gap duration discrimination in listeners with cochlear hearing loss: Effects of gap and marker duration, frequency separation, and mode of presentation," *J. Assoc. Res. Otolaryngol.* **2**, 388–398.

Hartmann, W. M., and Pumplin, J. (1988). "Noise power fluctuations and the masking of sine signals," *J. Acoust. Soc. Am.* **83**, 2277–2289.

Healy, E. W., and Bacon, S. P. (2002). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," *J. Speech Lang. Hear. Res.* **45**, 1262–1275.

Healy, E. W., and Bacon, S. P. (2006). "Measuring the critical band for speech," *J. Acoust. Soc. Am.* **119**, 1083–1091.

Healy, E. W., Kannabiran, A., and Bacon, S. P. (2005). "An across-frequency processing deficit in listeners with hearing impairment is supported by acoustic correlation," *J. Speech Lang. Hear. Res.* **48**, 1236–1242.

Healy, E. W., and Warren, R. M. (2003). "The role of contrasting temporal amplitude patterns in the perception of speech," *J. Acoust. Soc. Am.* **113**,

1676–1688.

Heinrich, A., and Schneider, B. (2006). "Age-related changes in within- and between-channel gap detection using sinusoidal stimuli," *J. Acoust. Soc. Am.* **119**, 2316–2326.

Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337–1351.

Knudson, V. O., and Harris, C. M. (1978). *Acoustical Designing in Architecture* (Acoustical Society of America, Melville, NY).

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (1997). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," *Acust. Acta Acust.* **83**, 659–669.

Lippmann, R. P. (1996). "Accurate consonant perception without mid-frequency speech energy," *IEEE Trans. Speech Audio Process.* **4**, 66–69.

Lister, J., Besing, J., and Koehnke, J. (2002). "Effects of age and frequency disparity on gap discrimination," *J. Acoust. Soc. Am.* **111**, 2793–2800.

Lister, J., and Roberts, R. A. (2005). "Effects of age and hearing loss on gap detection and the precedence effect: Narrow-band stimuli," *J. Speech Lang. Hear. Res.* **48**, 482–493.

Moore, B. C. J. (1998). *Cochlear Hearing Loss*. (Whurr, London).

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).

Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.* **45**, 224–240.

Perrott, D. R., and Williams, K. N. (1971). "Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity," *Psychonomic Soc.* **25**, 73–74.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). "Detection of silent intervals between noise activating different perceptual channels: Some properties of "central" auditory gap detection," *J. Acoust. Soc. Am.* **101**, 3694–3705.

Pumplin, J. (1985). "Low-noise noise," *J. Acoust. Soc. Am.* **78**, 100–104.

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

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

Stone, M. A., and Moore, B. C. J. (2003). "Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay," *Ear Hear.* **24**, 175–183.

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.

Yuan, H., Reed, C. M., and Durlach, N. I. (2004). "Envelope-onset asynchrony as a cue to voicing in initial English consonants," *J. Acoust. Soc. Am.* **116**, 3156–3167.

Effects of stimulus variability and adult aging on adaptation to time-compressed speech

Julie D. Golomb,^{a)} Jonathan E. Peelle, and Arthur Wingfield^{b)}

Volen National Center for Complex Systems, Brandeis University, Waltham, Massachusetts 02454

(Received 4 February 2006; revised 13 September 2006; accepted 1 January 2007)

With as few as 10–20 sentences of exposure, listeners are able to adapt to speech that is highly distorted compared to that which is encountered in everyday conversation. The current study examines the extent to which adaptation to time-compressed speech can be impeded by disrupting the continuity of the exposure sentences, and whether this differs between young and older adult listeners when they are equated for starting accuracy. In separate sessions conducted one week apart, the degree of adaptation was assessed in four exposure conditions, all of which involved exposure to the same number of time-compressed sentences. A continuous exposure condition involved presentation of the time-compressed sentences without interruption. Two alternation conditions alternated time-compressed speech and uncompressed speech by single sentences or groups of four sentences. A fourth condition presented sentences that were separated by a period of silence but no uncompressed speech. For all conditions, neither young nor older adults' overall level of learning was influenced by disruptions to the exposure sentences. In addition, participants' performance showed reliable improvement across the first and subsequent sessions. These results support robust learning mechanisms in speech perception that remain functional throughout the lifespan.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2436635]

PACS number(s): 43.71.Es, 43.71.Lz, 43.70.Dn [MSS]

Pages: 1701–1708

I. INTRODUCTION

Speaking rates vary considerably in the course of everyday listening. Although speech rates typically range from 140 to 180 words per minute (wpm), speakers reading from a prepared script often exceed these rates (Stine *et al.*, 1990). Even a single talker will vary his or her speaking rate considerably within a given conversation (Miller, Grosjean, and Lomanto, 1984). These variations in speech rate influence, among other things, listeners' perception of phonemic boundaries (Miller, Aibel, and Green, 1984; Miller and Liberman, 1979), which play a key role in speech recognition. To effectively process speech, listeners must therefore be able to adjust to changes in speech rate extremely rapidly. When variations in speech rate are minimal, as is typically the case in natural speech, listeners can accomplish this task without apparent effort. However, when speech rates are extremely fast, adaptation becomes considerably more difficult.

Despite the challenge posed by rapid speech rates, listeners are able to adapt to speech time compressed to rates that are substantially faster than that to which they are accustomed. Studies consistently show that exposure to 10–20 sentences of time-compressed speech is sufficient to significantly improve listeners' comprehension of such materials (Dupoux and Green, 1997; Pallier *et al.*, 1998; Sebastián-Gallés *et al.*, 2000). This perceptual learning is a robust phenomenon that is largely preserved in healthy older adults

(Peelle and Wingfield, 2005) in spite of other perceptual (Schneider, 1997) and cognitive (Wingfield and Stine-Morrow, 2000) changes.

Perceptual learning is often characterized as the shifting of attention toward cues that are relevant for a task and away from those which are irrelevant (Francis *et al.*, 2000; Goldstone, 1998; Nosofsky, 1986). For example, Francis *et al.* (2000) used category-level feedback to train listeners to distinguish speech cues that varied across two dimensions: (1) the frequency spectrum occurring during burst onset, and (2) the origin of the formant transitions. Different groups of listeners were provided feedback that emphasized the frequency or burst components of the stimuli. Listeners' judgment scores reflected this feedback, with decisions about stimuli with conflicting burst and formant information following the dimension on which they were trained. In addition, listeners made less use of cues present in the dimension on which they were not trained. These results support a mechanism of perceptual learning of speech sounds that involves a reallocation of attention during the learning process.

Given the above-mentioned view of perceptual learning, it is reasonable to think that adaptation to time-compressed speech occurs as listeners recalibrate phonemic boundaries to accommodate the much more rapid speech rate. An outstanding question is whether the redirection of attentional focus necessary for this adjustment can be disrupted if listeners hear speech at two speech rates that are far removed from each other. If increasing attention to cues salient at one speech rate comes at a cost to attention to cues at a second rate, alternating between the two rates over the course of an adaptation period should produce shifts in attention that hinder learning.

^{a)}Current affiliation: Interdepartmental Neuroscience Program, Yale University, New Haven, Connecticut 06520.

^{b)}Electronic mail: wingfield@brandeis.edu

An alternate line of research has examined stimulus variability as a means of increasing the overall learning and generalizability of phonetic training (e.g., Barcroft and Sommers, 2005; Lively *et al.*, 1993). Logan *et al.* (1991) trained Japanese listeners on identification of /t/ and /l/, a discrimination that previous training paradigms had little success in teaching. The authors effected a significant improvement in identification scores by employing training words that varied in placement and context of the consonant sounds (word-initial and final positions, singleton and cluster environments, and in intervocalic positions), spoken by five talkers. Follow-up studies using a single talker for training found that listeners improved in the /t/-/l/ distinction for that talker, but that this learning did not generalize to novel talkers (Lively *et al.*, 1993). These results indicate that, at least in some contexts, variability in training stimuli may actually be beneficial to listeners' perceptual learning. This presumably occurs because, rather than merely forming associations with a single idiosyncratic token of a phoneme, with sufficient variability listeners are able to define a perceptual space that includes both a prototype and several variants (e.g., Kuhl, 1991).

Dupoux and Green (1997) asked whether adaptation to time-compressed speech might be affected by interrupting a series of time-compressed sentences with presentation of sentences heard at a different speech rate. The authors presented listeners with 10 sentences time-compressed to 38% of their original duration followed by 5 uncompressed sentences, and a final 5 sentences again compressed to 38% of their original duration. Perceptual learning was assessed by asking listeners, following each sentence, to report as many words as possible from that sentence. Listeners' recall accuracy improved over the course of the time-compressed sentences, but showed a moderate drop for the two time-compressed sentences immediately following the uncompressed speech. However, listeners' accuracy quickly regained (and surpassed) the level achieved at the end of the initial 10-sentence set of compressed sentences. Dupoux and Green (1997) concluded that there was a short-term, local adjustment to the uncompressed speech that caused the drop in performance, but that perceptual learning persisted through a longer-term component, as evidenced by the rapid recovery.

The initial drop in accuracy following a change in speech rate reported by Dupoux and Green (1997) suggests that listeners' attentional focus may have indeed shifted toward speech cues salient at the slower speech rate, hindering their performance on the time-compressed sentences. It is possible that, had there been more interruptions between the exposure sentences, a summation of the small performance drops associated with each interruption would have resulted in a significant decline in overall learning. This would be consistent with attentional processes that tune in to the current or local speech rate at the expense of other speech rates. Alternatively, it may be the case that such narrow local attention is not necessary to adapt to speech at a new rate. The primary goal of the current study was to investigate whether frequent shifts in speech rate might have a significant impact on perceptual learning.

A second goal of this experiment was to examine the effect of shifts in speech rate on perceptual learning in older adults. There are many reasons one would expect age to adversely affect adaptation to rapid speech. First, older adults typically have poorer hearing acuity than do young adults (Morrell *et al.*, 1996), with this decline in auditory acuity contributing to poorer performance on a variety of language tasks (Sommers, 1997; Wingfield *et al.*, 2006). Accompanying these changes in peripheral hearing are age-related declines in auditory temporal processing ability. In gap detection tasks, older adults typically require longer gaps between sounds than do young adults to correctly discriminate one from two tones (Schneider *et al.*, 1994; Strouse *et al.*, 1998; Schneider and Hamstra, 1999). Fitzgibbons and Gordon-Salant (1995) investigated listeners' temporal resolution ability by comparing discrimination thresholds for changes in the duration of tones embedded in sequences to those for the same tones presented in isolation. They found that older adults' discrimination ability was significantly worse for tones embedded in sequences compared to that of the young adults. These findings have been upheld in studies using modified natural speech sounds (Gordon-Salant *et al.*, 2006).

Above and beyond these age-related changes in sensory processing, older adults exhibit declines in several cognitive domains important for language processing. These include declines in attentional resources (McDowd and Shaw, 2000), slower speed of information processing (Salthouse, 1996), a diminished working memory capacity (Zacks *et al.*, 2000), and difficulty inhibiting irrelevant or distracting information (Tun *et al.*, 2002). The result is that older adults perform more poorly than their younger counterparts on a wide variety of language tasks, including exhibiting a greater difficulty processing time-compressed speech (Gordon-Salant and Fitzgibbons, 1993; Wingfield *et al.*, 1999, 2003).

In addition to auditory and linguistic processing, adaptation to rapid speech involves learning new acoustic representations of stored phonemic categories, and older adults exhibit declines on a number of tasks involving learning and memory. These include word generation, paired associates, cued recall, and free recall (Craik, 1977; Craik *et al.*, 1987; Kausler, 1994). There is also evidence from the episodic memory literature that supports age-related weakening of associations between items (Kahana *et al.*, 2002; Wingfield and Kahana, 2002). Although the literature on age-related changes in perceptual learning is limited, the available evidence would lead one to expect age-related deficits on this front as well. For example, older adults show significant impairments in prism adaptation (Fernandez-Ruiz *et al.*, 2000) and semantic category visual search tasks (Gilbert and Rogers, 1996; Rogers *et al.*, 1994).

To the degree that adaptation to time-compressed speech depends on the above-listed processes, such adaptation should be adversely affected in adult aging. Contrary to this expectation, it has recently been demonstrated that under ideal training conditions—that is, an uninterrupted series of training sentences in an otherwise quiet environment—older adults initially adapt to time-compressed speech in a manner comparable to young adults (Peelle and Wingfield, 2005). However, this does not rule out age-related declines in per-

ceptual learning. In numerous other language paradigms, older adults can compensate to keep pace with young adults in easier conditions, but their performance falls off as difficulty is increased through perceptual or cognitive manipulations (e.g., Tun, 1998; Wingfield *et al.*, 2003). Older adults' potential difficulty with perceptual learning is also supported by age-related declines in the maintenance and transfer of perceptual learning with time-compressed speech (Peelle and Wingfield, 2005). It is therefore important to examine whether variations in speech rate exert a differential effect on older adults' adaptation.

A better understanding of the dynamics of perceptual learning in older age is vital for a number of reasons. In practical terms, as older adults have more difficulty understanding certain kinds of speech, they stand to benefit from well-designed training programs. To be effective, such programs will need to reflect a deep knowledge about the dynamics of perceptual learning and how easily it can be disrupted. Second, as indicated previously, existing research on perceptual learning in older adults is scant, particularly with respect to speech sounds. This study thus helps to address an important gap in our understanding of cognitive aging, particularly as it relates to language comprehension. Finally, given known age-related declines in cognitive resources, assessing older adults' performance on this task provides additional guidance in identifying cognitive systems that support perceptual learning. That is, age-related differences in task performance are likely attributable to a cognitive system which shows significant age-related change. Conversely, a lack of age differences would suggest that the processes do not rely on such systems.

In the current experiment, young and older adult listeners were presented with time-compressed sentences to assess perceptual learning in four conditions. In the first condition, time-compressed sentences were presented without any intervening uncompressed sentences, comparable to previous studies (Dupoux and Green, 1997; Pallier *et al.*, 1998; Sebastián-Gallés *et al.*, 2000; Peelle and Wingfield, 2005). In addition, adaptation was tested in three alternation conditions. Two of these conditions involved alternating compressed sentences with uncompressed sentences. Equal numbers of compressed and uncompressed sentences were used, but switches between the two rates occurred with different frequencies. In the final alternation condition, compressed sentences were alternated with a silent period to determine whether disruptions in exposure continuity, without experiencing sentences at a different speech rate, would influence learning.

II. METHOD

A. Participants

All participants were native English speakers and reported themselves to be in good health. The young participants (8 men and 8 women) consisted of university undergraduates aged 18–22 years ($M=19.6$, $s.d.=1.8$). They had a mean of 14.0 years of education at the time of testing ($s.d.=1.3$) and a mean WAIS-III vocabulary score (Wechsler, 1997) of 47.2 ($s.d.=5.2$). The older participants (6 men

and 10 women) were healthy, community-dwelling adults aged 65–85 years ($M=72.4$, $s.d.=5.0$). They had a mean of 15.8 years of education ($s.d.=1.4$) and a mean WAIS-III vocabulary score of 49.1 ($s.d.=7.2$). Both groups were thus well-educated, with the older group having an average of 1.8 more years of formal education [$t(30)=3.74$, $p<0.001$]. There was no difference in vocabulary score between the young and the older adults [$t(30)=0.84$, $n.s.$]. Participants were compensated with a small monetary sum for their involvement.

Audiometric testing was conducted to ensure that all participants had clinically normal hearing within the speech range. The mean pure-tone average [(PTA)—taken as the mean threshold for tones at 500, 1000, and 2000 Hz] in the better ear was 3.5 dB HL ($s.d.=3.6$) for the young adults and 13.7 dB ($s.d.=8.2$) for the older adults. Mean speech reception thresholds [(SRTs)—the intensity at which two-syllable words can be correctly identified 50% of the time] for the better ear were 4.4 dB HL ($s.d.=2.5$) for the young adults and 12.8 dB ($s.d.=8.4$) for the older adults. Although age differences in both PTA [$t(30)=4.54$, $p<0.001$] and SRT [$t(30)=3.89$, $p<0.001$] were significant, both groups fell within the range of clinically normal hearing, defined as a PTA of 25 dB HL or less in the speech frequency range (Hall and Mueller, 1997).

B. Stimuli

The stimuli consisted of 136 sentences, each of which contained 10 words (7 content words and 3 function words) and 14–16 syllables (e.g., *Sarah took her dirty work clothes to the neighborhood cleaners*). Sentences were recorded by a female native speaker of American English at a comfortable speech rate of approximately 200 wpm.

Because older adults are differentially affected by rapid speech (Gordon-Salant and Fitzgibbons, 1993; Wingfield *et al.*, 1999, 2003), it would have been problematic to use a single speech rate for all participants: Speech that would have been appropriate for young adults would have proven too difficult for older adults, whereas slowing the speech rate to accommodate the older adults would not have posed a sufficient challenge for the young adult listeners. Different speech rates were therefore used for the young and older adults so that they would be approximately equated for starting accuracy level. These speech rates were chosen based on pilot data to equate both participant groups for overall accuracy levels prior to any adaptation.

Sound-editing software (SoundEdit, Macromedia Inc., San Francisco) was used to compress 96 of the stimulus sentences to 30% of their original duration (approximately 680 wpm) for the young adults and 40% of their original duration (approximately 510 wpm) for the older adults. The software uses a type of PSOLA algorithm for time compression, which compresses both speech and silence equally and preserves information from short sounds such as stop consonants and formant glides through temporal averaging. The remaining 40 sentences were left uncompressed. Sentences were presented binaurally over headphones at a comfortable

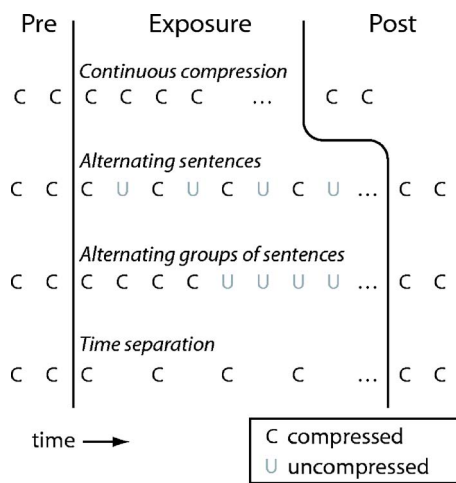


FIG. 1. Schematic illustration of the four experimental conditions employed. In all four conditions, two time-compressed sentences were presented before and after exposure to assess the effects of exposure on sentence intelligibility. All four conditions contained 20 time-compressed sentences during the exposure portion, with no interruption (*continuous compression*), alternating with uncompressed speech in single sentences (*alternating sentences*) or groups of four (*alternating groups of sentences*), or with a silent pause between compressed sentences (*time separation*).

listening level, determined for each participant, which remained unchanged for that participant throughout the experiment.

C. Procedure

The four presentation conditions used in this study are schematically displayed in Fig. 1. In each of the four conditions, the first two sentences (preexposure) and the last two sentences (postexposure) were always presented at the compressed rate. The sentences in between varied in composition according to condition. The *continuous compression* condition consisted of 20 time-compressed exposure sentences presented without any intervening uncompressed sentences and is comparable to previous studies involving perceptual adaptation to time-compressed sentences (Dupoux and Green, 1997; Pallier *et al.*, 1998; Sebastián-Gallés *et al.*, 2000; Peelle and Wingfield, 2005). The *alternating sentences* and *alternating groups of sentences* conditions consisted of presentation of 20 compressed exposure sentences and 20 uncompressed sentences alternating either every other sentence or in groups of four sentences, respectively. A fourth *time separation* condition consisted of 20 compressed exposure sentences presented with a 5 s pause inserted between them to approximate the spacing of the *alternating sentences* condition without presenting any uncompressed speech.

Participants were individually tested in four sessions. Each session involved one of the four above-described conditions, with the order of conditions counterbalanced across participants. The particular sentences used in each condition were also counterbalanced across participants, such that across all participants each sentence occurred equally often in each condition. Sessions were separated by 7–8 days to minimize possible carryover in learning from previous sessions.

Participants were instructed to recall verbatim as much of each sentence as possible immediately following its presentation. They were encouraged to guess if they were unsure, and they were informed that there would be no penalty for words recalled incorrectly or in the wrong order. No feedback was given regarding the correctness of responses. Participants pressed a key to indicate when they were finished recalling; a second keypress initiated presentation of the next sentence. Participants' spoken responses were recorded onto cassette tape for later analysis. The first session included a single time-compressed practice sentence to orient participants to the task; subsequent sessions included no practice.

D. Scoring

Sentence recall was scored as the number of correct content words out of seven possible for each sentence. Words with added or eliminated suffixes (e.g., -s, -ed, -ing) were counted as correct, as were verbs given in the wrong tense. If a participant recalled half of a compound word (e.g., "school" instead of "schoolwork") it was marked as half credit. Portions of noncompound words (e.g., "let" instead of "letters") were not given any credit. These scoring criteria are consistent with previous studies (Dupoux and Green, 1997; Pallier *et al.*, 1998; Sebastián-Gallés *et al.*, 2000; Peelle and Wingfield, 2005).

III. RESULTS

A. Effects of exposure condition

All analyses focused on listeners' comprehension of time-compressed sentences. Performance on any intervening uncompressed sentences, which occurred in the two alternating speech conditions, was near perfect and was not analyzed further. For each condition, listeners' performance on preexposure and postexposure sentences was compared to evaluate the magnitude of adaptation.

Recall accuracy for participants' pre- and postexposure sentences in all conditions is displayed in Fig. 2. As intended, by using a somewhat slower speech rate for the older adults than for the young adults, the preexposure levels were approximately equal across condition and age group. These preexposure data were submitted to a 4 (Exposure condition: continuous compression, alternating sentences, alternating groups of sentences, time separation) \times 2 (Age: young, older) mixed design ANOVA, with exposure condition as a within-participants variable. The main effect of exposure condition was not significant [$F(3, 90) < 1$], as was expected, given that for each session the preexposure sentences occurred before any experimental manipulations. Although older adults showed a slight advantage in starting accuracy levels due to the specific speech rates that were used, there was no significant effect of age [$F(1, 30) < 1$], or an age by condition interaction [$F(3, 90) < 1$]. Thus, the speech rates were appropriately chosen to equate the two groups on starting accuracy.

Visual inspection of Fig. 2 shows that exposure to time-compressed speech improved listeners' accuracy scores in all conditions. To confirm this conclusion, the data were submitted to a 2 (Exposure: preexposure, postexposure) \times 4 (Ex-

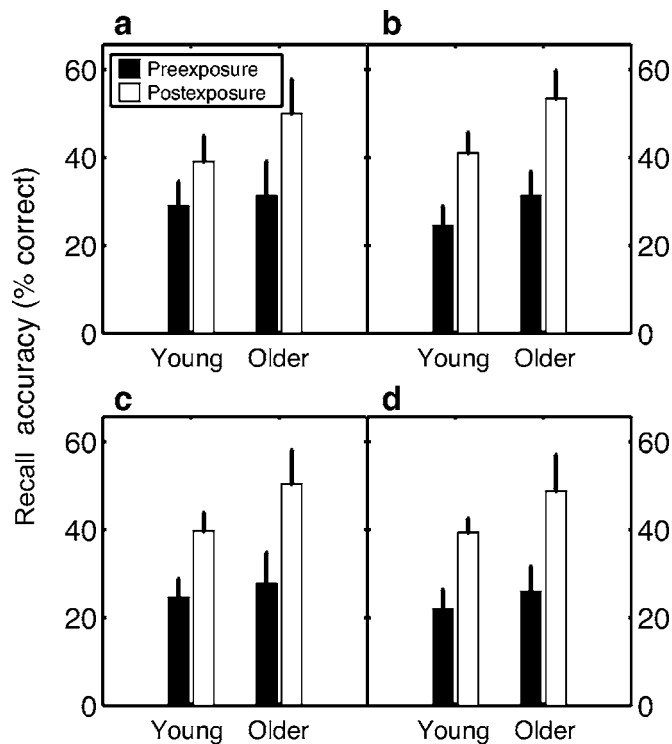


FIG. 2. Response accuracy for young and older listeners before (black bars) and after (white bars) exposure to time-compressed sentences as a function of exposure condition. (a) Continuous compression condition (no interruption); (b) alternating sentences condition (alternating single compressed and uncompressed sentences); (c) alternating groups of sentences condition (alternating compressed and uncompressed sentences in groups of 4); and (d) time separation condition (5 s between sentences). Error bars represent 1 standard error.

posure condition: continuous compression, alternating sentences, alternating groups of sentences, time separation) \times 2 (Age: young, older) ANOVA. There was a main effect of exposure [$F(1, 30) = 53.87, p < 0.001$], confirming that listeners' accuracy improved with exposure to the time-compressed speech. On average, the young group improved from 25.1% correct in the preexposure test to 39.8% correct in the postexposure test, an average improvement of 14.7%. The older group improved from an average of 29.0% correct in the preexposure test to 50.7% correct in the postexposure test, an average improvement of 21.7% correct. There was no effect of exposure condition [$F(3, 90) < 1$], supporting the observation that participants' improvement was not affected by stimulus spacing or by speech rate variability. In addition, the older adults' performance did not differ significantly from that of the young adults, producing no main effect of age [$F(1, 30) = 1.41, n.s.$]. None of the interactions was significant (all F 's < 2 , all p 's > 0.17).

It is conceivable that, although overall improvement due to exposure to time-compressed speech was equivalent across conditions, learning might still have been affected by the exposure condition manipulation. For example, participants' recall accuracy may have fluctuated more from trial to trial in the alternating conditions relative to the continuous compression condition due to the inclusion of uncompressed sentences, which have been shown to transiently reduce recall accuracy (Dupoux and Green, 1997). To determine

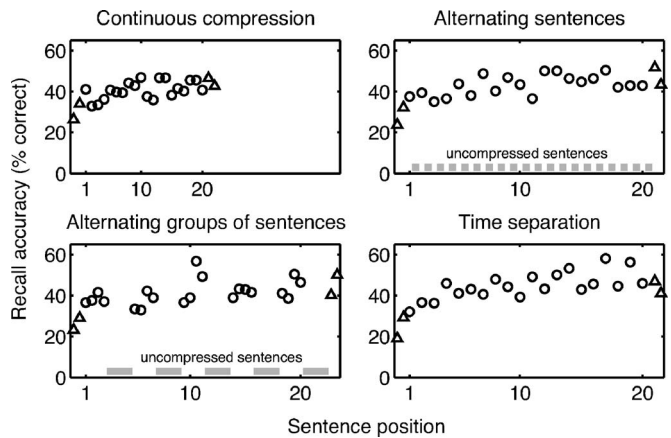


FIG. 3. Response accuracy for time-compressed sentences during each of four exposure conditions (circles), collapsed across young and older adults. The gray bars along the abscissa indicate when uncompressed sentences occurred for the alternating sentences and alternating groups of sentences conditions. The preexposure and postexposure sentences are replotted here for reference (triangles). Sentence position refers to the location of each time-compressed sentence within the exposure sentences.

whether this was the case, the exposure data were analyzed for each condition trial-by-trial. These exposure data, along with those for pre- and postexposure data, are displayed in Fig. 3 (the gray bars along the abscissa indicate the location of uncompressed sentences or groups of sentences in those conditions). To allow comparison across conditions, each exposure sentence was assigned a position number 1–20 based on the amount of exposure, regardless of any intervening sentences or pauses. Thus, equivalent sentence positions indicate that comparable exposure had occurred.

The exposure data were submitted to a 4 (Exposure Condition: continuous compression, alternating sentences, alternating groups of sentences, time separation) \times 20 (Exposure sentence position: 1–20) ANOVA. As expected, recall accuracy improved with increased exposure, indicated by a main effect of exposure sentence position [$F(19, 589) = 2.75, p < 0.001$]. Visual inspection of Fig. 3 might suggest that the alternating group condition resulted in slightly less steady learning, with brief drops in accuracy every four time-compressed sentences, corresponding to the inserted groups of uncompressed sentences. However, despite these minor fluctuations, there was no consistent effect of exposure condition on learning, evidenced by a lack of a main effect of exposure condition [$F(3, 93) = 1.43, n.s.$]. The results of the learning data analysis are thus consistent with our analysis of the pre- and postexposure sentences, and do not indicate any significant effect of exposure condition.

B. Effects of repeated exposure

Although sessions were conducted a minimum of 7 days apart in order to minimize carryover of learning from one session to another, the possibility that some carryover occurred could not be ruled out. Because the first set of analyses failed to reveal any effect of exposure condition, it was possible to analyze the data as a function of session, collapsing across exposure condition.

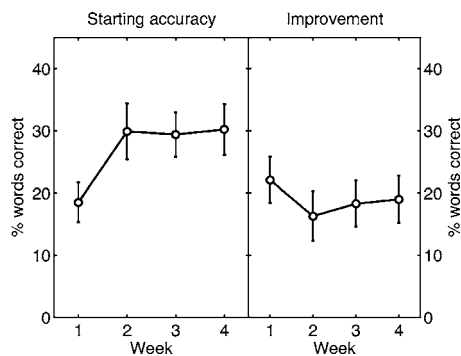


FIG. 4. Starting accuracy (left panel) and amount of improvement (right panel) for the four experimental sessions, collapsed across age and exposure condition. Error bars represent 1 standard error.

Participants' preexposure recall accuracy at the start of each of the four sessions is shown in the left panel of Fig. 4. Because there were no age differences in our analyses of exposure condition, data for young and older listeners have been combined when looking at session-to-session improvement. The lack of any age differences was also verified using uncorrected t-tests (all p 's > 0.14).

Visual inspection of the left panel of Fig. 4 suggests that starting accuracy increased between the first session and subsequent sessions, which is consistent with retention of learning from one session to subsequent sessions. An ANOVA confirmed this difference in starting accuracy levels, showing a significant effect of session [$F(3,93)=5.11, p < 0.05$]. Post-hoc t-tests confirmed that this effect was attributable to the significant increase in starting accuracy from week 1 to week 2 [$t(31)=3.08, p < 0.01$]. There was no significant change between weeks 2–3 or 3–4 (both p 's > 0.8).

In addition to examining listeners' starting accuracy levels, the question of whether total amount of improvement changed over time was addressed by analyzing improvement scores (difference between postexposure and preexposure scores) over sessions. These data are plotted in the right panel of Fig. 4. Visual inspection of the data suggests listeners' accuracy improved a comparable amount each week. Consistent with this observation, an ANOVA on these data failed to show a significant effect of session [$F(3,93) < 1$], indicating that the amount of improvement indeed remained stable across sessions. As with the starting accuracy data, t-tests for each week's accuracy levels relative to baseline confirmed that there were no differences in performance between young and older adults (all p 's > 0.29).

IV. DISCUSSION

Although there are a number of possible ways to characterize perceptual learning, researchers in the speech literature have generally opted for models involving the allocation of attention to relevant stimulus dimensions (e.g., Francis and Nusbaum, 2000; Logan *et al.*, 1991; Nittrouer and Miller, 1997). In the current study we investigated whether wide fluctuations in speech rate would disrupt this attentional process. We found that interrupting time-compressed training sentences with brief pauses or uncompressed speech had no

effect on perceptual learning. Furthermore, there were no differences in performance between young and older adult listeners.

Our results suggest that listeners' attentional allocation to a particular speech rate is not disrupted by the use of a second speech rate, as would be predicted from previous studies of learning with speech sounds (Francis, 2000). This finding is especially surprising given that the current experiment employed a fairly marked contrast, alternating time-compressed speech (510 or 680 wpm) with uncompressed speech (200 wpm). Because listeners take longer to adapt to speech that is further removed from unaltered speech (Dupoux and Green, 1997; Peelle and Wingfield, 2005), these extreme choices of speech rates should present a worst-case scenario for the listener.

This discrepancy might seem to imply that the paradigm of attentional allocation is incorrect, and that adapting to a new speech rate does not in fact require attentional resources. If this were the case, it would not be surprising if shifts in speech rate should fail to hinder learning. An alternative framework within which to view adaptation to time-compressed speech might be closer to that of stimulus imprinting, in which repeated exposure to novel stimuli lead to strengthened internalized traces of that stimulus (e.g., Logan, 1988; Palmeri *et al.*, 1993; Palmeri, 1997). Under such a view, learning would be governed most strongly by the amount of exposure to novel sounds, regardless of intervening sentences. However, this and other alternative frameworks seem to be poor fits for the generalizability shown in speech perception, and are not consistent with previous studies of phoneme discrimination.

A more likely explanation, then, is that natural speech represents a special case in speech processing. Although we used two very different speech rates, only the faster rate required adaptation by the listeners. As the normal speech rate did not require assimilation of any new information, it was not dependent on attentional resources, and thus did not take these resources away from the novel, faster speech rate. Restated more formally, the normal speech rate presumably contains phoneme prototypes (Kuhl, 1991). At the beginning of training the time-compressed speech tokens are too far away from the prototype to be correctly identified, but through training they are brought closer in perceptual space (Iverson and Kuhl, 1995). Within such a framework the prototypes remain relatively stable, and would not be disrupted by the small number of sentences employed in our training regime. Under this interpretation, attention is indeed directed toward relevant dimensions of the novel stimuli, but is not diverted when the uncompressed stimuli are presented. This description fits best with prior studies of phoneme discrimination and the current data.

Successful speech comprehension relies upon a number of more general cognitive processes, including auditory processing, working memory, attention, and executive control (e.g., Grossman *et al.*, 2003; Peelle *et al.*, 2004; Price *et al.*, 2005; Wingfield and Stine-Morrow, 2000). Given the notable age-related declines in the cognitive abilities presumed to be necessary for perceptual learning, the lack of age differences in the current study is particularly striking. Two possible ex-

planations exist for this difference. First, it is possible that portions of the language processing system function in a mostly autonomous and encapsulated manner, and that because of their importance and continuous use are relatively resistant to age-related cognitive declines seen in the larger language-processing network. A second possibility is that, although there are age-related declines in cognitive systems important for perceptual learning, older adults are able to compensate for these declines through increased expertise (Wingfield and Stine-Morrow, 2000) and neural recruitment (Cabeza, 2002; Wingfield and Grossman, 2006). Regardless of the explanation, the current results support the hypothesis that, within the temporal limits of the aging auditory system, the initial stage of adaptation to time-compressed speech is preserved in adult aging (Peelle and Wingfield, 2005).

The design of the current study does not allow us to rule out an alternate interpretation of our results. Specifically, the two alternating conditions contained an additional 20 uncompressed sentences. It is conceivable that the alternation did in fact hinder performance, but that this effect was obscured by a commensurate *increase* in performance due to the additional numbers of stimuli employed in these two conditions. Previous studies have indicated the adaptation to time-compressed speech is not simply due to general task practice (Dupoux and Green, 1997; Peelle and Wingfield, 2005). However, because all of the stimuli were spoken by the same speaker, it is possible that there was adaptation to speaker-specific acoustic cues apart from speech rate aided by increased exposure to uncompressed sentences.

It is notable that both young and older participants' pre-exposure recall accuracy showed a significant increase between the first and second weeks. Although the duration of perceptual adaptation has not yet been fully specified (see Janse, 2003), the current data provide strong evidence that learning persists for at least one week. It is also striking that no significant age differences were observed in week-to-week retention. It has been previously demonstrated that older adults' learning reaches an asymptote after 20 time-compressed sentences, whereas young adults' performance continues to improve (Peelle and Wingfield, 2005). The current results indicate that if training is conducted in multiple sessions, older adults' performance can improve beyond this initial limit and keep pace with that of their younger counterparts. The breaks between weekly sessions may be necessary to combat fatigue or to allow time for memory consolidation.

Finally, these results have practical implications for training regimes in both experimental and real-world situations in which acoustically altered speech stimuli are encountered (e.g., Davis *et al.*, 2005; Rosen *et al.*, 1999; Shannon *et al.*, 1995; Vongphoe and Zeng, 2005). First, the precise timing of exposure sentences is not important for effective learning. Second, interrupting training sentences with uncompressed speech (e.g., conversation with the researcher) will not influence adaptation.

The current study demonstrates that adaptation to time-compressed speech is not hindered by brief pauses between training sentences or the insertion of uncompressed sentences. Most important, the lack of differences in young and

older adults' performance points to learning processes that are highly resistant to age-related cognitive decline, or that such declines are effectively countered in older adults by compensatory processes.

ACKNOWLEDGMENTS

The authors acknowledge support from NIH Grant Nos. AG04517 and AG019714 from the National Institute on Aging (A.W.) and fellowship F31 DC006376 from the National Institute on Deafness and other Communicative Disorders (J.P.). We also gratefully acknowledge support from the W. M. Keck Foundation.

- Barcroft, J., and Sommers, M. S. (2005). "Effects of acoustic variability on second language vocabulary learning." *Stud. Second Lang. Acquis.* **27**, 387–414.
- Cabeza, R., (2002). "Hemispheric asymmetry reduction in older adults: The HAROLD model." *Psychol. Aging* **17**, 85–100.
- Craik, F. I. M. (1977). "Age differences in human memory," in *Handbook of the Psychology of Aging*, edited by J. E. Birren and K. W. Schaie (Von Nostrand Reinhold, New York), pp. 384–420.
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., and Anderson, N. D. (1987). "Patterns of memory loss in three elderly samples," *Psychol. Aging* **2**, 79–86.
- Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," *J. Exp. Psychol. Gen.* **134**, 222–241.
- Dupoux, E., and Green, K. (1997). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 914–927.
- Fernandez-Ruiz, J., Hall, C., Vergara, P., and Diaz, R. (2000). "Prism adaptation in normal aging: Slower adaptation rate and larger aftereffect," *Cognit. Brain Res.* **9**, 223–226.
- Fitzgibbons, P. J., and Gordon-Salant, S. (1995). "Age effects on duration discrimination with simple and complex stimuli," *J. Acoust. Soc. Am.* **98**, 3140–3145.
- Francis, A. L., Baldwin, K., and Nusbaum, H. C. (2000). "Effects of training on attention to acoustic cues," *Percept. Psychophys.* **62**, 1668–1680.
- Gilbert, D. K., and Rogers, W. A. (1996). "Age-related differences in perceptual learning," *Hum. Factors* **38**, 417–424.
- Goldstone (1998). "Perceptual Learning," *Annu. Rev. Psychol.* **49**, 585–612.
- Gordon-Salant, S., and Fitzgibbons, P. J. (1993). "Temporal factors and speech recognition performance in young and elderly listeners," *J. Speech Hear. Res.* **36**, 1276–1285.
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (2006). "Age-related differences in identification and discrimination of temporal cues in speech segments," *J. Acoust. Soc. Am.* **119**, 2455–2466.
- Grossman, M., Cooke, A., DeVita, C., Lee, C., Alsop, D., Detre, J., Gee, J., Chen, W., Stern, M. B., and Hurtig, H. I. (2003). "Grammatical and resource components of sentence processing in Parkinson's disease: An fMRI study," *Neurology* **60**, 775–781.
- Hall, J., and Mueller, G. (1997). *Audiologist Desk Reference* (Singular, San Diego).
- Iverson, P., and Kuhl, P. K. (1995). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," *J. Acoust. Soc. Am.* **97**, 553–562.
- Janse, E. (2003). *Production and Perception of Fast Speech* (Landelijke Onderzoekschool Taalwetenschap, Utrecht, The Netherlands).
- Kahana, M. J., Howard, M., Zaromb, F., and Wingfield, A. (2002). "Age dissociates recency and lag-recency effects in free recall," *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 530–540.
- Kausler, D. H. (1994). *Learning and Memory in Normal Aging* (Academic, San Diego).
- Kuhl, P. (1991). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Percept. Psychophys.* **50**, 93–107.
- Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). "Training Japanese listeners to identify English /t/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories." *J.*

- Acoust. Soc. Am. **94**, 1242–1255.
- Logan, G. D. (1988). "Toward an instance theory of automatization," *Psychol. Rev.* **95**, 492–527.
- 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.
- McDowd, J. M., and Shaw, R. J. (2000). "Attention and aging: A functional perspective," in *The Handbook of Aging and Cognition*, 2nd ed, edited by F. I. M. Craik and T. A. Salthouse (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 221–292.
- Miller and Liberman (1979). "Some effects of later-occurring information on the perception of stop consonant and semivowel," *Percept. Psychophys.* **25**, 457–465.
- Miller, J. L., Aibel, I. L., and Green, K. (1984). "On the nature of rate-dependent processing during phonetic perception," *Percept. Psychophys.* **35**, 5–15.
- Miller, J. L., Grosjean, F., and Lomanto, C. (1984). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," *Phonetica* **41**, 215–225.
- Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (1996). "Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level," *J. Acoust. Soc. Am.* **100**, 1949–1967.
- Nittrouer, S., and Miller, M. E. (1997). "Predicting developmental shifts in perceptual weighting schemes," *J. Acoust. Soc. Am.* **101**, 2253–2266.
- Nosofsky, R. M. (1986). "Attention, similarity, and the identification-categorization relationship," *J. Exp. Psychol. Gen.* **115**, 39–57.
- Pallier, C., Sebastián-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (1998). "Perceptual adjustment to time-compressed speech: A cross-linguistic study," *Mem. Cognit.* **26**, 844–851.
- Palmeri, T. J. (1997). "Exemplar similarity and the development of automaticity," *J. Exp. Psychol. Learn. Mem. Cogn.* **23**, 324–354.
- Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). "Episodic encoding of voice attributes and recognition memory for spoken words," *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 309–328.
- Peelle, J. E., McMillan, C., Moore, P., Grossman, M., and Wingfield, A. (2004). "Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI," *Brain Lang* **91**, 315–325.
- Peelle, J. E., and Wingfield, A. (2005). "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech," *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 1315–1330.
- Price, C., Thierry, G., and Griffiths, T. (2005). "Speech-specific auditory processing: Where is it?," *Trends Cogn. Sci.* **9**, 271–276.
- Rogers, W. A., Fisk, A. D., and Hertzog, C. (1994). "Do ability-performance relationships differentiate age and practice effects in visual search?," *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 710–738.
- Rosen, S., Faulkner, A., and Wilkinson, L. (1999). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," *J. Acoust. Soc. Am.* **106**, 3629–3636.
- Salthouse, T. A. (1996). "The processing-speed theory of adult age differences in cognition," *Psychol. Rev.* **103**, 403–428.
- Schneider, B. A. (1997). "Psychoacoustics and aging: Implications for everyday listening," *Journal of Speech-Language Pathology and Audiology* **21**, 111–124.
- Schneider, B. A., and Hamstra, S. J. (1999). "Gap detection thresholds as a function of tonal duration for younger and older listeners," *J. Acoust. Soc. Am.* **106**, 371–380.
- Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (1994). "Gap detection and the precedence effect in young and old adults," *J. Acoust. Soc. Am.* **95**, 980–991.
- Sebastián-Gallés, N., Dupoux, E., Costa, A., and Mehler, J. (2000). "Adaptation to time-compressed speech: Phonological determinants," *Percept. Psychophys.* **62**, 834–842.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Sommers, M. S. (1997). "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," *J. Acoust. Soc. Am.* **101**, 2278–2288.
- Stine, E. L., Wingfield, A., and Myers, S. D. (1990). "Age differences in processing information from television news: The effects of bisensory augmentation," *J. Gerontol.* **45**, 1–8.
- Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). "Temporal processing in the aging auditory system," *J. Acoust. Soc. Am.* **104**, 2385–2399.
- Tun, P. A. (1998). "Fast noisy speech: Age differences in processing rapid speech with background noise," *Psychol. Aging* **13**, 424–434.
- Tun, P. A., O'Kane, G., and Wingfield, A. (2002). "Distraction by competing speech in young and older adult listeners," *Psychol. Aging* **17**, 453–467.
- Vongphoe, M., and Zeng, F.-G. (2005). "Speaker recognition with temporal cues in acoustic and electric hearing," *J. Acoust. Soc. Am.* **118**, 1055–1061.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (Psychological Corporation, New York).
- Wingfield, A., and Grossman, M. (2006). "Language and the aging brain: Patterns of neural compensation revealed by functional brain imaging," *J. Neurophys.* **96**, 2830–2839.
- Wingfield, A., and Kahana, M. J. (2002). "The dynamics of memory retrieval in older adulthood," *Can. J. Exp. Psychol.* **56**, 187–199.
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, L. C. (2006). "Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity," *J. Am. Acad. Audiol.* **17**, 487–497.
- Wingfield, A., Peelle, J. E., and Grossman, M. (2003). "Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults," *Aging, Neuropsychology, and Cognition* **10**, 310–322.
- Wingfield, A., and Stine-Morrow, E. A. L. (2000). "Language and speech," in *The Handbook of Aging and Cognition*, 2nd ed., edited by F. I. M. Craik and T. A. Salthouse (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 359–416.
- Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (1999). "Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech," *Psychol. Aging* **14**, 380–389.
- Zacks, R. T., Hasher, L., and Li, K. Z. H. (2000). "Human memory," in *The Handbook of Aging and Cognition*, edited by F. I. M. Craik and T. A. Salthouse (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 293–357.

Evaluation of companding-based spectral enhancement using simulated cochlear-implant processing

Andrew J. Oxenham,^{a)} Andrea M. Simonson, Lorenzo Turicchia, and Rahul Sarpeshkar
*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139*

(Received 4 January 2006; revised 14 December 2006; accepted 21 December 2006)

This study tested a time-domain spectral enhancement algorithm that was recently proposed by Turicchia and Sarpeshkar [IEEE Trans. Speech Audio Proc. **13**, 243–253 (2005)]. The algorithm uses a filter bank, with each filter channel comprising broadly tuned amplitude compression, followed by more narrowly tuned expansion (companding). Normal-hearing listeners were tested in their ability to recognize sentences processed through a noise-excited envelope vocoder that simulates aspects of cochlear-implant processing. The sentences were presented in a steady background noise at signal-to-noise ratios of 0, 3, and 6 dB and were either passed directly through an envelope vocoder, or were first processed by the companding algorithm. Using an eight-channel envelope vocoder, companding produced small but significant improvements in speech reception. Parametric variations of the companding algorithm showed that the improvement in intelligibility was robust to changes in filter tuning, whereas decreases in the time constants resulted in a decrease in intelligibility. Companding continued to provide a benefit when the number of vocoder frequency channels was increased to sixteen. When integrated within a sixteen-channel cochlear-implant simulator, companding also led to significant improvements in sentence recognition. Thus, companding may represent a readily implementable way to provide some speech recognition benefits to current cochlear-implant users. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2434757]

PACS number(s): 43.71.Ky, 43.66.Ts [DOS]

Pages: 1709–1716

I. INTRODUCTION

In a recent paper, Turicchia and Sarpeshkar (2005) proposed a new strategy for time-domain spectral enhancement, based on relatively broadband compression followed by more frequency-selective expansion. This compressing-and-expanding (companding) approach leads to certain properties shared by the peripheral auditory system. In particular, it can produce a suppression of the response to one tone by the presence of another, more intense, tone at a nearby frequency—an effect known as two-tone suppression. At a more global level, the companding scheme can lead to the enhancement of spectral peaks in a stimulus, relative to nearby spectral valleys. Spectral enhancement techniques have often been used in an attempt to improve speech intelligibility in noise for listeners with hearing loss. Despite many different approaches and solutions (e.g., Bunnell, 1990; Simpson *et al.*, 1990; Stone and Moore, 1992; Baer *et al.*, 1993; Franck *et al.*, 1999; Lyzenga *et al.*, 2002; Yang *et al.*, 2003), the results so far, in terms of improved speech intelligibility, have been not been particularly promising (for a review, see Moore, 2003).

In contrast to earlier studies using spectral enhancement techniques, an initial evaluation of Turicchia and Sarpeshkar's (2005) algorithm has shown that it can provide some improvement in speech intelligibility for cochlear-implant

users and for normal-hearing subjects listening to acoustic simulations of cochlear-implant processing (Bhattacharya and Zeng, 2007): In both groups, recognition of words in sentences increased on average by between 10 and 20 percentage points. These new findings are encouraging and stimulate a number of questions. In particular, it is not clear what properties of the new algorithm (Turicchia and Sarpeshkar, 2005) or the evaluation method (Bhattacharya and Zeng, 2007) allow it to provide some benefit, where others have shown little or no improvement. A number of possible explanations exist. One reason may be that, in contrast to most other schemes, this was tested with cochlear-implant users, rather than hearing-impaired listeners. Frequency selectivity is generally poorer in cochlear-implant users, who typically only have access to about eight broadly tuned channels of spectral information (Friesen *et al.*, 2001), than in listeners with moderate-to-severe hearing loss. It is thus possible that cochlear-implant users might benefit more than hearing-impaired listeners from spectral enhancement schemes. For instance, it has been shown that the benefits of enhanced spectral contrast can increase with decreasing initial spectral resolution (Loizou and Poroy, 2001).

Another potential reason is that the scheme is implemented in the time domain, using logarithmically scaled and spaced filters. This is closer to the processing carried out by the normal auditory system than the more typical linear-frequency FFT processing, in which the windowing can lead to temporal distortions that may be more audible than those produced by a logarithmically scaled system. One other potential reason is the number of analysis channels used. One

^{a)} Author to whom correspondence should be addressed. Current address: Department of Psychology, University of Minnesota, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

other earlier study used a filterbank approach, which was implemented in real time with 16 contiguous frequency channels, where the outputs from adjacent channels were used to adjust the gain of each channel, thereby producing some enhancement of spectral peaks (Stone and Moore, 1992). The authors suggested that the lack of improvement they found might be due to the small number of frequency analysis bands (16) used in their implementation. In contrast, Bhattacharya and Zeng's (2007) implementation of the companding algorithm had 50 analysis channels.

The present study had three main aims. The first aim was to test the robustness of the improvement provided by the companding algorithm. This was done by using very similar parameters to those used by Bhattacharya and Zeng (2007). The second aim was to identify the parameters that were critical in providing the improvement in speech intelligibility. Specifically, changes to the companding time constants and filter bandwidths were made and their effects on performance were measured. The third aim was to test whether a "pared-down" version of the companding algorithm, which might be suitable for implementation in cochlear-implants in the near future, might also provide improved speech intelligibility. This aim was also able to test the effect of reducing the number of analysis channels from 50 to 16, in order to test whether this was a critical difference between the evaluation of the companding algorithm (Bhattacharya and Zeng, 2007) and the earlier filter-bank study of Stone and Moore (1992) in hearing-impaired listeners.

The aims were achieved by measuring speech intelligibility in normal-hearing listeners using a steady-state background noise, after the stimuli (speech and noise) had been processed through a noise-excited envelope vocoder, which is designed to simulate certain aspects of cochlear-implant processing, such as the poor frequency selectivity and the lack of temporal fine-structure information (e.g., Shannon *et al.*, 1995; Dorman *et al.*, 1998). The study is divided into five separate experiments, each involving eight different normal-hearing listeners. The first three experiments test a 50-channel version of the companding algorithm by passing its output through an 8-channel cochlear-implant simulator and varying the filter bandwidths and time constants of the algorithm. The next experiment tests the same 50-channel algorithm after passing it through a 16-channel cochlear-implant simulator. The final experiment presents a modified version of the algorithm that uses only 16 analysis channels, in a way that could be readily implemented with current implant technology.

II. GENERAL METHODS

A. Stimuli

Listeners were tested on their ability to correctly report words from the Hearing in Noise Test (HINT) sentence lists (Nilsson *et al.*, 1994), spoken by a male talker. The background noise was steady Gaussian noise that was spectrally shaped to approximate the long-term spectrum of speech, and was generated by filtering white noise with a filter composed of two poles at 200 Hz, two poles at 8 kHz, and a zero

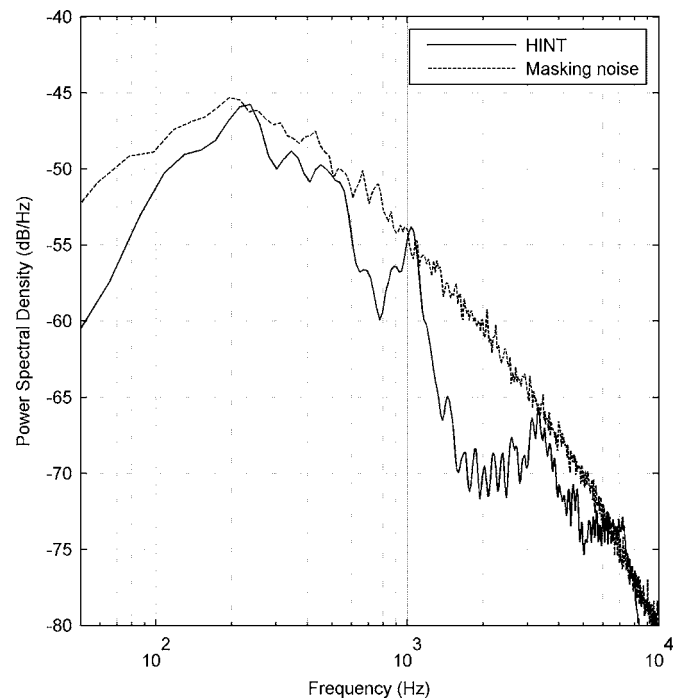


FIG. 1. Long-term power spectrum of the speech used in the HINT data base (solid line) and the steady-state masking noise (dashed line).

in the origin (see Fig. 1). The speech was presented at average signal-to-noise ratios of 0, 3, and 6 dB. The noise and speech were always added prior to any processing.

1. Companding processing

We implemented the companding algorithm presented in Turicchia and Sarpeshkar (2005). The strategy uses a non-coupled filter bank and compression-expansion blocks as shown in Fig. 2. Every channel in the companding architecture has a relatively broad prefilter, a compression block, a relatively narrowband postfilter, and an expansion block. The prefilter and postfilter in each channel have the same center frequency. The pre- and postfilter banks have logarithmically spaced center frequencies that span the desired spectral range. Finally, the channel outputs of this nonlinear filter bank are summed to generate an output with enhanced spectral peaks.

The broadband filter determines the frequency range that can affect the gain of the compressor in a given channel. The narrowband filter determines a narrower subset of these fre-

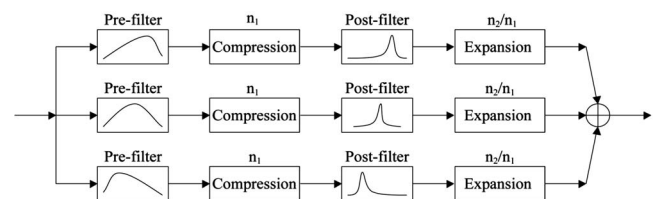


FIG. 2. Block diagram of the companding architecture, showing the stimulus analyzed in a bank of broad prefilters. The output of each prefilter is subjected to compression, is filtered again using sharper postfilters, and then expanded. The outputs from each channel are then summed to produce the processed broadband stimulus. For more details, see Turicchia and Sarpeshkar (2005).

quencies that can affect the gain of the expander. The compressor and expander are normally complementary, such that a tone at the center frequency of the filters will undergo sequential compression and expansion and emerge unchanged in level. However, sufficiently intense frequencies outside the narrowband filter passband but within the broadband filter passband can determine the gain of the compressor only to be filtered out by the narrowband filter. This can result in the suppression of weak frequency components at or near the filters' center frequency by stronger more remote frequency components. Consequently, the output spectrum of the filter bank will have a local winner-take-all-like characteristic with strong spectral peaks in the input suppressing weaker neighboring ones, and high signal-to-noise-ratio channels being emphasized over weaker ones. There are certain similarities between the properties exhibited by the companding algorithm and phenomena such as two-tone suppression that are found at early stages of the auditory periphery (e.g., Sachs and Kiang, 1968; Ruggero *et al.*, 1992). However, there are also differences, such as the differential growth of suppression for low- and high-side suppressors that is found in the auditory periphery (e.g., Delgutte, 1990), but not in the companding algorithm.

The compression and expansion blocks can be instantaneous or use feedforward automatic gain control (AGC), which sets the gain based on the input level of the signal. In this paper, we only use AGCs as in the original description of Turicchia and Sarpeshkar (2005). The time constants of the compression and expansion AGCs are set at $w/(2\pi f_c)$ s with $w=40$ unless otherwise mentioned, where f_c is the center frequency of the channel filters in Hertz. The compression and expansion functions are compressive and expansive power laws of $n_1=0.3$ and $1/n_1=1/0.3$, respectively. The n_2 parameter of the algorithm, which sets the overall compression in the signal after companding, had a value of 1 throughout this paper such that there was no overall compression. Further parameters are listed in Sec. III A. A more detailed analysis of the potential benefits and operation of the algorithm may be found in Turicchia and Sarpeshkar (2005).

2. Noise vocoding

All stimuli (whether enhanced by companding or not) were processed by a noise-excited envelope vocoder in the following manner: The stimuli were first bandpass filtered (with sixth-order Butterworth filters) into 8 or 16 contiguous frequency bands between 250 and 6000 Hz. The frequency range was divided equally in terms of the ERB_N scale (Glasberg and Moore, 1990). For the 8-channel conditions, the Q values of the resulting filters ranged from 1.8 for the lowest-frequency filter to around 3.0 for the mid- and high-frequency filters (above 1 kHz, where the ERB_N is roughly constant as a proportion of the center frequency); for the 16-channel conditions, the Q values were higher, ranging from around 3.5 at the lowest center frequency to between 5 and 6 for the mid- to high-frequency filters. The envelopes of the signals were extracted by half-wave rectification and lowpass filtering using a second-order Butterworth filter with a cutoff frequency of 200 Hz. The envelopes were then used to modulate narrowband noises, filtered by the same band-

pass filters that were used to filter the original stimuli. The modulated narrowband noise waveforms were summed together. This simulation preserves the temporal envelope cues in speech, but not the fine structure (e.g., Shannon *et al.*, 1995).

B. Procedure

Listeners were instructed that they would hear distorted sentences in a noisy background. The listeners' task was to type what they heard via a computer keyboard. Listeners were told that some of the utterances would be very hard to understand, and that they should make their best guess at as many words as possible.

Approximately 1 h of practice was provided before testing in order to acclimatize listeners to the sound of the noise-excited envelope vocoder. For this, listeners were presented with eighteen lists, each comprising seven sentences from the IEEE/TIMIT corpus (IEEE, 1969), spoken by a male talker. The sentences were initially presented without noise and then at signal-to-noise ratios of 10 and then 5 dB. At each signal-to-noise ratio a list of seven sentences was presented without companding, followed by a list with companding. After each practice sentence, listeners were required to type what they had heard via a computer keyboard. Feedback was then provided by displaying the sentence on the computer monitor, and listeners had the option to repeat the sentence as often as they wished before moving on to the next sentence.

The actual experiment comprised the three different signal-to-noise ratios (0, 3, and 6 dB SNR), each presented with and without companding, resulting in a total of six conditions. A list of 10 HINT sentences was presented in a block for each condition. The order of presentation of the conditions was randomized across listeners, and each condition was presented once before any was repeated. A total of 24 HINT sentence lists were presented in each experiment, resulting in a total of four lists (i.e., 40 sentences) per condition per listener. No feedback was given during the actual experiment, which took approximately 1 h and immediately followed the practice session.

For both the practice and actual experiments, the stimuli were presented diotically at an overall level of 75 dB SPL (after processing) in each ear over Sennheiser 580 headphones. All processing (companding and cochlear-implant simulation) was carried out in advance and the sentences were stored as 16-bit sound files on the computer hard drive and were played out via a LynxOne (LynxStudio) sound card. After the session, the listeners' responses were scored off-line. All words were counted, and the percent of correct words in each condition was calculated. Obvious misspellings were counted as correct.

C. Listeners

Eight listeners participated in each of the experiments. No listener participated in more than one experiment, and no listener had taken part in any previous study involving HINT sentences. All listeners were paid for their participation, and

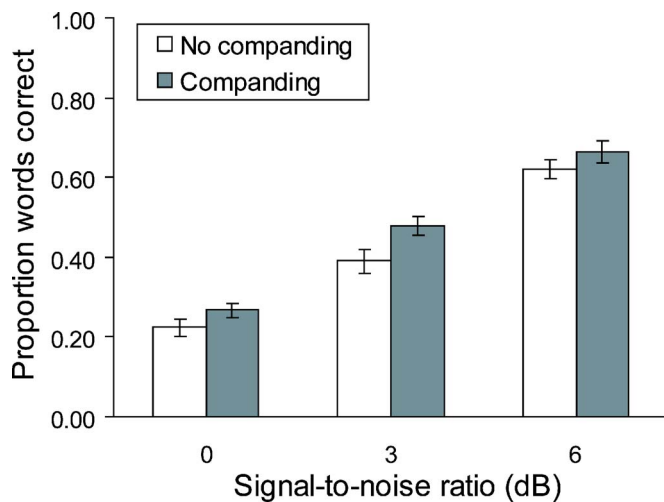


FIG. 3. Initial speech intelligibility results with and without companding. Proportion of words correctly reported are plotted for the three signal-to-noise ratios tested. Error bars represent ± 1 standard error of the mean across the eight listeners.

had audiometric thresholds of 25 dB HL or less at octave frequencies between 250 and 8000 Hz. The ages of the listeners ranged from 18 to 60, with a mean age of 29.

III. EXPERIMENT 1: PERFORMANCE USING AN 8-CHANNEL IMPLANT SIMULATOR WITH AND WITHOUT COMPANDING

A. Companding parameters

The first experiment involved testing the companding algorithm with parameters essentially identical to those used by Bhattacharya and Zeng (2007). The Q values of the pre-filters and postfilters were set to 2.3 and 13.8, respectively (corresponding to Q_{10} values of 1.1 and 6.8, respectively).¹ The time-constant parameter for the AGCs, w , was set to 40, which corresponds to an overall time constant of about 13 ms at 500 Hz and 1.3 ms at 5000 Hz. Stimuli were processed using a 50-channel implementation of the companding algorithm, with the lowest and highest filter center frequencies set to 100 and 8000 Hz, respectively. The filters were spaced at equal distances along a log scale.

B. Results and discussion

The results, averaged across the eight listeners, are shown in Fig. 3. The percentage of words correctly reported is shown with and without companding at the three signal-to-noise ratios tested. The companding provides a small but reliable improvement in performance. This was confirmed using a repeated-measures analysis of variance (ANOVA), with companding (on or off) and signal-to-noise ratio as fixed factors, which showed that both factors were statistically significant [companding: $F(1,7)=13.65$, $p<0.01$; signal-to-noise ratio: $F(2,7)=436.252$, $p<0.001$], with no significant interaction. The improvement in performance, averaged across subjects and signal-to-noise ratios, was 6 percentage points. This corresponds roughly to a 1-dB improvement in signal-to-noise ratio at a given level of performance. These results are broadly consistent with those reported by

Bhattacharya and Zeng (2007). They showed an average improvement of around 5 and 15 percentage points in sentence word recognition at speech-to-noise ratios of 0 and +5 dB, respectively, with normal-hearing subjects listening through a 16-channel implant simulator. More importantly, they also showed an average improvement of between 5 and 20 percentage points at speech-to-noise ratios of between 0 and +10 dB for cochlear-implant users. Thus, the present experiment, together with the study of Bhattacharya and Zheng (2007) confirms the small, but robust, improvements produced by the algorithm in both normal-hearing subjects (through a noise-excited vocoder simulation of implant processing) and cochlear-implant users. The remaining experiments probe the effects of various parameters within the algorithm, such as filter bandwidth, compression time constants, and the number of analysis and presentation filter channels.

IV. EXPERIMENT 2: EFFECTS OF ALTERING ANALYSIS FILTER BANDWIDTH AND COMPRESSION AND EXPANSION TIME CONSTANTS

A. Rationale

One difference between the parameters used by us and Bhattacharya and Zeng (2007), and those proposed in the original paper (Turicchia and Sarpeshkar, 2005) is the bandwidth of the filters. The original paper proposed using relatively broad filters, with overall Q values of 2.3 and 4.6 (corresponding to Q_{10} values of 1.1 and 2.3) for the pre- and postfilter, respectively. The narrower postfilters used in experiment 1 are closer to those found in normal hearing, where estimates of Q range from between about 9 and 12 at 1 kHz, depending on how the bandwidths are estimated (Oxenham and Shera, 2003). On the other hand, broader (lower-order) filters are less computationally expensive, have less inherent group delay, and are amenable to very low-power analog implementation, which may make them of practical value in cochlear implants. Experiment 2A investigated the effect of decreasing the Q value of the postfilters by a factor of about 3.

Another potentially important set of parameters within the companding algorithm are the time constants associated with compression and expansion. The question of what time constants are most appropriate for compression in hearing aids has been studied extensively. On one hand, very short attack time constants can be deleterious to intelligibility (e.g., Stone and Moore, 2004). On the other hand, the compression found in the normal auditory system at the level of the basilar membrane tends to be very fast acting, and can be thought of as acting quasi-instantaneously on the temporal envelope (e.g., Ruggero *et al.*, 1997). In addition, studies using companding in automatic speech recognition, suggested potential benefits of shorter time constants (Guinness *et al.*, 2005). The time constant of $40/(2\pi f_c)$ s, used in experiment 1 corresponds to around 6 cycles of the center frequency, which is longer than that associated with cochlear compression and suppression effects. Experiment 2B tested the effect of decreasing the time constants.

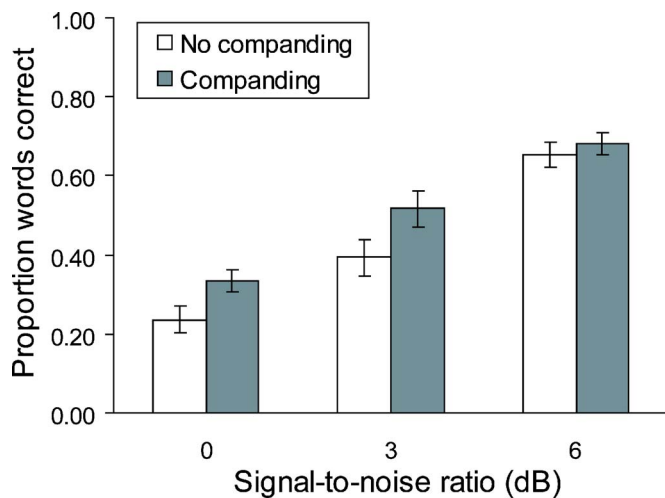


FIG. 4. Speech intelligibility with broader filter tuning in the postfilter stage of the compander. Bars show the mean and standard error from eight new subjects. Other details are as in Fig. 3.

B. Companding parameters

In experiment 2A, both compression and expansion AGC time constants were $40/(2\pi f_c)$ s, as in experiment 1, but the postfilter Q value was changed. Specifically, in experiment 2A, the prefilter and postfilter Q values were 2.3 and 4.6. In experiment 2B, the Q values were the same as in experiment 1, but the compression time constant was reduced to $5/(2\pi f_c)$ s and the expansion time constant was reduced to $20/(2\pi f_c)$ s. All other parameters were left unchanged from experiment 1.

C. Results and discussion

The average results from experiments 2A and 2B are shown in Figs. 4 and 5, respectively. Consider first the results using the broader filters (Fig. 4). As in experiment 1, a repeated-measures ANOVA showed significant effects of both companding [$F(1, 7)=15.883, p=0.005$] and signal-to-noise ratio [$F(2, 7)=168.491, p<0.001$], with no interaction.

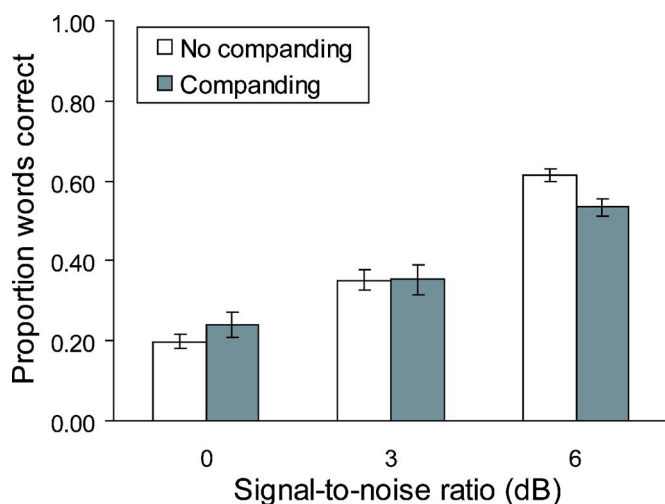


FIG. 5. Speech intelligibility with reduced time constants for both the compression and expansion stages of the compander. Bars show the mean and standard error from eight new subjects. Other details are as in Fig. 3.

A comparison of these results with those from experiment 1 (Fig. 3) suggests that decreasing the sharpness of tuning of the companding filters by a factor of between 2 and 3 produced no decrease in the improvement in intelligibility over the original parameters [$F(2, 14)<1, n.s.$].

Consider next the results using the shorter time constants. Here, an ANOVA revealed no significant main effect of companding [$F(1, 7)=0.359, p=0.568$], with a significant interaction [$F(2, 7)=9.333, p=0.003$], suggesting that decreasing the time constants actually worsened performance, especially at the higher signal-to-noise ratio, relative to the original parameter choices. A between-subjects comparison between the companding conditions of experiment 1 and experiment 2B indicated a significant effect [$F(2, 14)=5.323, p=0.012$], further confirming the significantly poorer performance with the shorter time constants.

Overall, the results confirm that companding can improve speech intelligibility somewhat and show that the improvement is not critically dependent on the filter bandwidths of the companding scheme. This is encouraging in terms of the prospects for implementing the algorithm in actual cochlear implants, because broader filters are less computationally expensive. Furthermore, the architecture is amenable to very low-power analog VLSI implementations, which may result in a much longer battery life than is currently available. The results do, however, seem to depend on the compression and expansion time constants, with time constants of $40/(2\pi f_c)$ s producing a significant improvement, and shorter time constants of $5/(2\pi f_c)$ s and $20/(2\pi f_c)$ s for compression and expansion, respectively, producing no significant benefit relative to no processing. Thus, in contrast to what was found in automatic speech recognition (Guinness *et al.*, 2005), shorter, perhaps more physiologically realistic, time constants do not appear to improve intelligibility, at least for the conditions tested here. It remains to be seen what effect would result from time constants longer than those used in experiment 1.

V. EXPERIMENT 3: COMPANDING FOLLOWED BY 16-CHANNEL VOCODER PROCESSING

A. Rationale

All the results so far were obtained using an 8-channel noise-excited envelope vocoder. In an earlier study using vowel stimuli, Loizou and Poroy (2001) found that the minimum spectral contrast necessary for vowel recognition increased with decreasing number of spectral channels. Thus, it is possible that if our simulations had used a larger number of more narrowly tuned channels, no further benefit of companding would have been observed. Given that current cochlear implants are rarely able to deliver more than eight independent channels of information, the caveat associated with numbers of channels may not be an important one for the present. However, it is possible that technical improvements will lead to a larger number of channels becoming available for implant patients in the future. It is therefore of interest to test the extent to which the improvement in performance found for companding will generalize to cochlear-

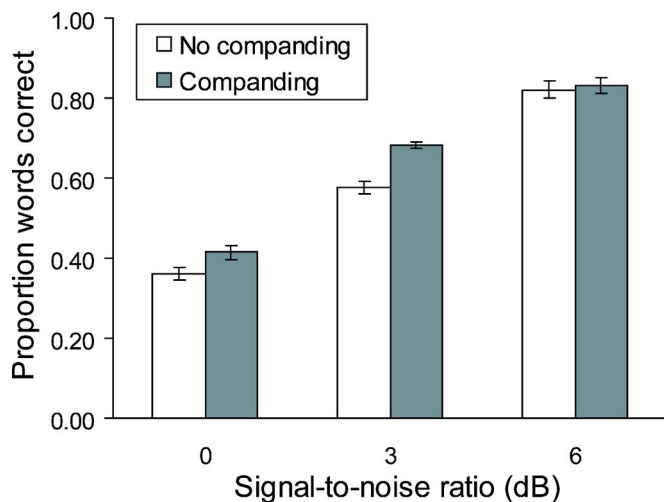


FIG. 6. Speech intelligibility with a 16-band vocoder. Frequency resolution was improved by presenting subjects with 16 bands of spectral information. The companding algorithm itself was the same as used for experiment 1 (see Fig. 3). Again, eight new subjects participated.

implant simulations with larger numbers of spectral channels. This was tested in experiment 3 using a 16-channel implant simulator.

B. Companding and vocoder parameters

The parameters used in the companding algorithm were the same used in experiment 1. The only difference is that the sounds were then processed by a cochlear-implant simulator implemented with 16 bandpass filter (sixth-order Butterworth filters) between 250 and 6000 Hz rather than 8, as in experiments 1 and 2. The frequency range was again divided equally in terms of the ERB_N scale (Glasberg and Moore, 1990).

C. Results

The results using a 16-channel implant simulator are shown in Fig. 6. A repeated-measures ANOVA showed significant effects of both companding [$F(1, 7)=18.611, p < 0.005$] and signal-to-noise ratio [$F(2, 7)=308.605, p < 0.001$]. This time, the interaction also reached significance [$F(2, 7)=7.883, p=0.005$], reflecting the fact that little or no improvement was observed at the highest signal-to-noise ratio. The improvement in the conditions with companding, averaged across listeners and signal-to-noise ratios, was 5 percentage points. Thus, despite the improved spectral resolution of the cochlear-implant simulations (16 instead of 8 channels), some benefit was still provided by the companding algorithm.

VI. EXPERIMENT 4: SPECTRAL ENHANCEMENT WITH ONLY 16 ANALYSIS CHANNELS

A. Rationale

The experiments so far have demonstrated the ability of the companding system described by Turicchia and Sarpeshkar (2005) to provide small but significant benefits in speech intelligibility when the speech is passed through 8- and 16-channel cochlear-implant simulators. However, the algorithm

as tested so far still comprises 50 independent frequency analysis channels, the outputs of which are combined before being fed to the cochlear-implant simulator. It would not be possible to implement such processing within current implants without increasing their complexity. Another possibility is to simply modify existing filters within cochlear implants to incorporate the benefits of companding in each of the channels of a cochlear implant. The final experiment tests whether it is possible to reduce the number of analysis channels within the companding algorithm from 50 to 16 while still retaining the beneficial effects of companding that were observed in the original experiments.

B. Algorithm

Stimuli were processed by first implementing companding using only a 16-channel filter bank, as shown in Fig. 2 but without the final summation. The outputs of the resulting companding channels were combined as in a cochlear-implant simulator by extracting the final envelope outputs of each companding channel, using these envelopes to modulate bandpass noise corresponding to that channel, and then finally summing these outputs. In effect, we directly use companded spectral-analysis outputs, rather than speech enhanced by companding, as inputs to the noise vocoder.

The parameters used for companding were as in experiment 1 except that the prefilter was a second-order Butterworth filter and the postfilter was a fourth-order Butterworth filter, so that the overall filter order was the same as the sixth-order Butterworth filter used in the original vocoder. To provide a quantitative comparison of the filters used here and in Experiment 1, the Q_{10} values were derived. In experiment 1, the Q_{10} values for the pre- and postfilter were 1.1 and 6.8, respectively. Here, for the Butterworth filters, the Q_{10} values were between 1.2 and 2.1 for the prefilters and between 2 and 3.5 for the postfilters, depending on the center frequency. All envelope detectors for compression and expansion were implemented as second-order lowpass Butterworth filters, with cut-off frequencies of 70 and 12 Hz, respectively.

C. Results

The results for the 16-channel companding system are shown in Fig. 7. A repeated-measures ANOVA showed significant effects of both companding [$F(1, 7)=13.825, p < 0.01$] and signal-to-noise ratio [$F(2, 7)=190.8, p < 0.001$], with no significant interaction. The mean increase in intelligibility, pooled across subjects and signal-to-noise ratios, was 4 percentage points. Thus, even with a reduction in the number of analysis frequency channels in the companding algorithm from 50 to 16, a small benefit of companding was still observed.

VII. DISCUSSION AND SUMMARY

The results in many of the conditions show that companding can result in a small but significant improvement in the intelligibility of speech in a background of steady-state noise. The improvements proved robust to some, but not other, changes in the companding parameters. In particular, it was possible to reduce the sharpness of tuning of the post-

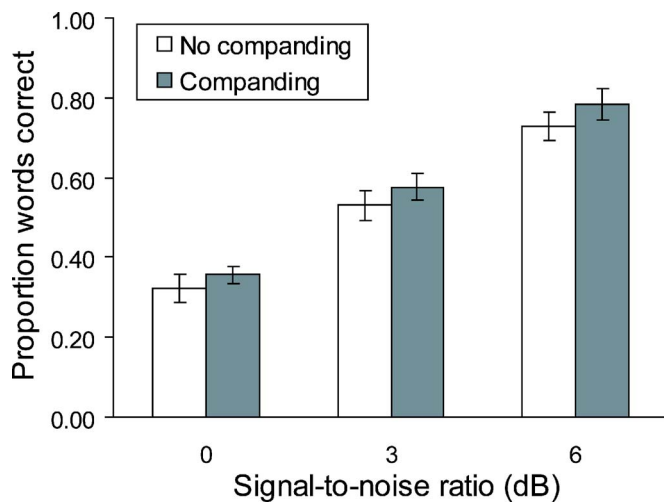


FIG. 7. Speech intelligibility with 16-channel companding, followed by a 16-band vocoder. Instead of the 50 channels used in all previous experiments, this graph shows results using only 16 analysis channels, the outputs of which are fed directly to 16 presentation channels within the implant simulator. Other graph details are as in Fig. 3.

filter by a factor of 3 without altering the benefit provided by the companding algorithm. On the other hand, shortening the time constants to more closely resemble the instantaneous compression and suppression provided by the basilar membrane resulted in the loss of any benefit provided by the algorithm. The benefit of companding was observed for both 8-channel and 16-channel simulations, suggesting that the spectral enhancement continues to be of benefit even with a spectral resolution that exceeds what is currently available to most cochlear-implant users. Perhaps most encouragingly, a simplified version of the companding algorithm, using only 16 analysis channels in a way that could be implemented into current implants, still resulted in a small improvement in speech intelligibility.

The present study tested speech intelligibility only in a steady-state noise background. Some of the most challenging acoustic environments for cochlear implants and implant simulations are those that involve maskers with spectrotemporal fluctuations, such as a competing talker (Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004). Further study will be required to determine whether the companding algorithm described here continues to provide an advantage in more complex masking situations.

There are obvious limits to the applicability of the current research using acoustic stimulation to actual cochlear implants. Nevertheless, as shown by a number of studies (e.g., Shannon *et al.*, 1995; Friesen *et al.*, 2001; Turner *et al.*, 2004) such simulations can capture many important aspects of cochlear implant processing, without some of the confounding individual differences often observed in patients, and at much lower effort and cost. In further support of our approach, the results of experiment 1 are in good agreement with those of Bhattacharya and Zeng (2007) in actual cochlear-implant users, which provides some optimism that the new results from this study—in particular the finding that a readily implementable system can provide some benefit—may generalize to cochlear-implant users.

ACKNOWLEDGMENTS

This work was supported primarily by the Advanced Bionics Corporation. A.J.O. and A.M.S. are also supported by the National Institutes of Health [No. R01 DC 03909 and R01 DC 05216 (awarded to A.J.O.)]. We thank Johannes Lyzenga, Michael Stone, and one anonymous reviewer for their helpful comments on an earlier version of this manuscript.

¹Turicchia and Sarpeshkar (2005) used the parameters q_1 and q_2 to describe the pre- and postfilter properties. Here for simplicity we use the overall Q of the filters. The original filters were built with 4 stages and therefore the overall Q value is 1.15 times the value of q_1 or q_2 . As an alternative measure, Q_{10} values refer to the center frequency of the filter, divided by the bandwidth of the filter at its 10-dB (rather than 3-dB) down points.

- Baer, T., Moore, B. C. J., and Gatehouse, S. (1993). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality and response times," *J. Rehabil. Res. Dev.* **30**, 49–72.
- Bhattacharya, A., and Zeng, F. G. (2007). "Companding to improve cochlear-implant speech recognition in noise," *J. Acoust. Soc. Am.* (submitted).
- Bunnell, H. T. (1990). "On enhancement of spectral contrast in speech for hearing-impaired listeners," *J. Acoust. Soc. Am.* **88**, 2546–2556.
- Delgutte, B. (1990). "Two-tone suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," *Hear. Res.* **49**, 225–246.
- 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.
- Franck, B. A., van Kreveld-Bos, C. S., Dreschler, W. A., and Verschuure, H. (1999). "Evaluation of spectral enhancement in hearing aids, combined with phonemic compression," *J. Acoust. Soc. Am.* **106**, 1452–1464.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Guinness, J., Raj, B., Schmidt-Nielsen, B., Turicchia, L., and Sarpeshkar, R. (2005). "A companding front end for noise-robust automatic speech recognition," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 249–252.
- IEEE. (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Loizou, P. C., and Poroy, O. (2001). "Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners," *J. Acoust. Soc. Am.* **110**, 1619–1627.
- Lyzenga, J., Festen, J. M., and Houtgast, T. (2002). "A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity," *J. Acoust. Soc. Am.* **112**, 1145–1157.
- Moore, B. C. J. (2003). "Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms," *Speech Commun.* **41**, 81–91.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Nilsson, M., Soli, S., 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.
- Oxenham, A. J., and Shera, C. A. (2003). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," *J. Assoc. Res. Otolaryngol.* **4**, 541–554.
- 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.
- 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.
- Ruggero, M. A., Robles, L., and Rich, N. C. (1992). "Two-tone suppression

- in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," *J. Neurophysiol.* **68**, 1087–1099.
- Sachs, M. B., and Kiang, N. Y. S. (1968). "Two-tone inhibition in auditory nerve fibers," *J. Acoust. Soc. Am.* **43**, 1120–1128.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Simpson, A. M., Moore, B. C. J., and Glasberg, B. R. (1990). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," *Acta Oto-Laryngol., Suppl.* **469**, 101–107.
- 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. (1992). "Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality," *J. Rehabil. Res. Dev.* **29**, 39–56.
- Stone, M. A., and Moore, B. C. J. (2004). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," *J. Acoust. Soc. Am.* **116**, 2311–2323.
- Turicchia, L., and Sarpeshkar, R. (2005). "A bio-inspired companding strategy for spectral enhancement," *IEEE Trans. Speech Audio Process.* **13**, 243–253.
- Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (2004). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," *J. Acoust. Soc. Am.* **115**, 1729–1735.
- Yang, J., Luo, F., and Nehorai, A. (2003). "Spectral contrast enhancement: Algorithms and comparisons," *Speech Commun.* **39**, 33–46.

Sinusoidal modeling for nonstationary voiced speech based on a local vector transform

Masashi Ito and Masafumi Yano

Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

(Received 26 May 2006; revised 12 December 2006; accepted 13 December 2006)

A voiced speech signal can be expressed as a sum of sinusoidal components of which instantaneous frequency and amplitude continuously vary with time. Determining these parameters from the input, the time-varying characteristics are crucial error sources for the algorithms, which assume their stationarity within a local analysis segment. To overcome this problem, a new method is proposed, local vector transform (LVT), which can determine instantaneous frequency and amplitude for nonstationary sinusoids. The method does not assume the local stationarity. The effectiveness of LVT was examined in parameter determination for synthesized and naturally uttered speech signals. The instantaneous frequency for the first harmonic component was determined with an accuracy almost equal to that of the time-corrected instantaneous frequency method and higher accuracy than that of spectral peak-picking, autocorrelation, and cepstrum. The instantaneous amplitude was also determined accurately by LVT while considerable errors were left in the other algorithms. The signal reconstructed from the determined parameters by LVT agreed well with the corresponding component of voiced speech. These results suggest that the method is effective for analyzing time-varying voiced speech signals. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2431581]

PACS number(s): 43.72.Ar, 43.60.Ac, 43.60.Wy [DOS]

Pages: 1717–1727

I. INTRODUCTION

Acoustic analysis provides basic data for a wide range of speech-related studies, such as speech synthesis, perception, coding, and modification. In the analysis of voiced speech signals, at least two features are known to be crucial: the fundamental frequency (F0) and the spectral envelope. The F0 reflects continuous vibration of the vocal folds, which is closely related to perceived pitch. Cepstrum and autocorrelation are widely accepted methods for F0 analysis (Noll, 1967; Rabiner, 1977). The spectral envelope reflects the shape of the vocal tract, which is mainly related to perceived vowel quality. Linear predictive coding is a well-known analysis method for the spectral envelope (Atal and Hanauer, 1971).

Another important acoustical analysis method that has been extensively studied is the sinusoidal model, which can potentially analyze both F0 and the spectral envelope at the same time (Hedlin, 1981; McAulay and Quatieri, 1986; Marques and Almeida, 1989; George and Smith, 1997; Laroche, 2000; Hermus *et al.*, 2005). In this method, voiced speech is expressed as a sum of sinusoidal components of which frequency and amplitude are continuously varied with time. Then, the F0 corresponds to the frequency for the first harmonic component whereas the envelope corresponds to the amplitude for each component. One of the crucial advantages of the sinusoidal model is reversibility, which means that the signal can be uniquely reconstructed from the parameters (frequency and amplitude). This enables the evaluation of the accuracy of the analyzed parameters, where the reconstructed signal agrees with the input if the parameters are accurately estimated.

The parameters in the sinusoidal model cannot be uniquely determined from the input without any constraints because the process is a kind of ill-posed problem, in which frequency and amplitude for every signal component must be estimated from only the waveform. To solve this problem, it is effective to assume stationarity of the parameters within a local analysis period. On this local stationarity assumption (LSA), the frequency and amplitude for each component can be determined from the corresponding peak in the power spectrum. McAulay and Quatieri (1986) proposed a system for sinusoidal modeling based on the spectral peak-picking and claimed the reconstructed signal to be indistinctive from the input speech for listeners.

However, the reconstructed signal in the method by McAulay and Quatieri does not physically agree with the input. In actual speech signals, frequency and amplitude for every component continuously vary with time as a result of articulatory movements. In these signals, the periods where the LSA may be valid are very limited. Thus, inevitable errors are left in the parameter determination based on the LSA. This is a crucial drawback for the reversibility in the sinusoidal model.

To overcome this problem, there have been two noteworthy approaches. First, Marques and Almeida (1989) proposed a method for nonstationary sinusoidal modeling, in which the instantaneous frequency was not assumed stationary but was assumed to vary linearly with time. The frequency was determined by iterative calculations to minimize the error between the input and modeled spectra. George and Smith (1997) proposed a similar iterative method in which the error was evaluated in the time domain. These algorithms are effective for nonstationary sinusoids when the optimiza-

tion processes are successfully converged. However, it is not certain that the process converges for an arbitrary input signal because it may fall into a local minimum if the initial parameters are not appropriate.

Second, instantaneous frequency estimation is also important in another line of studies. Since the original proposal by Kodera and colleagues, extensive studies have been conducted on the instantaneous frequency estimation (Kodera *et al.*, 1976; Auger and Flandrin, 1995; Nelson, 2001). Recently, the method was clearly verified by Fulop and Fitz (2006) and called as a time-corrected instantaneous frequency (TCIF). In TCIF, the instantaneous frequency of the separable amplitude and/or frequency-modulated line component is clearly expressed on a remapped representation, which is uniquely calculated from a phase spectrogram of the input. Since the method does not assume the local stationarity, the instantaneous frequency can be accurately determined for time-varying signals. For voiced speech signals, the TCIF can determine the instantaneous frequency at least for the lower harmonic components, which are well separated in frequency in most cases. However, the TCIF may not determine the instantaneous amplitude accurately because it is only used for frequency estimation. A reliable method that can determine the accurate instantaneous amplitude for time-varying components has not yet been proposed, while it is indispensable for the reversibility of the sinusoidal model.

In the present study, we propose a new method that can accurately determine instantaneous frequency, phase, and amplitude for time-varying sinusoidal components without an iterative calculation. In Sec. II, the theoretical background of this method is described. The basic idea and example of the parameter determination is presented in Sec. III. In Sec. IV, the effectiveness of the method is examined in parameter determination for the first harmonic component of synthesized and naturally uttered speech signals. Some conventional methods are also examined for the comparison. Finally, the discussion is presented in Sec. V.

II. NONSTATIONARY SINUSOID

A voiced speech signal $x(t)$ is expressed as a sum of sinusoidal components:

$$x(t) = \sum_{k=1}^K \exp A_k(t) \cos P_k(t). \quad (1)$$

Here, $A_k(t)$ and $P_k(t)$ represent instantaneous logarithmic amplitude and phase for the k th component, respectively. Instantaneous frequency $F_k(t)$ corresponds to the time derivative of $P_k(t)/2\pi$. In sinusoidal modeling, the problem to solve is the determination of $A_k(t)$ and $F_k(t)$ [or $P_k(t)$] from the input signal $x(t)$. Apparently, the determination is not difficult when both $A_k(t)$ and $F_k(t)$ are stationary. However, these parameters are time-varying in most parts of natural utterances as a result of continuous movements of articulators. In the following, the parameter determination for nonstationary sinusoid will be presented analytically.

A. Stylized nonstationary sinusoid

First, we define a sinusoidal signal that models the nonstationarity of voiced speech components. For each component of the utterance, $P_k(t)$ can be approximated by Taylor's polynomials around $t=t_c$ to

$$P(t) \approx \sum_{n=0}^{\infty} p_n(t-t_c)^n/n!, \quad (2)$$

$$p_n = \frac{d^{(n)}P(t_c)}{dt^{(n)}}. \quad (3)$$

Here, we dropped k for simplicity. If the frequency $F(t)$ is time varying, then there must be at least one nonzero p_n for $n \geq 2$. Thus, $P(t)$ is approximated to a quadratic function by assuming that the higher terms might be negligible. The amplitude $A(t)$ is also expressed as another quadratic function in a similar manner. Based on the approximations, a quadratic-parameter sinusoid (QPS) is defined as

$$q(t) = \exp(a_0 + a_1 \cdot t + a_2 \cdot t^2/2) \cos(p_0 + p_1 \cdot t + p_2 \cdot t^2/2), \quad (4)$$

where the analysis center time t_c is set to zero for simplicity. In the signal, instantaneous frequency $F(t)$ is $p_1/2\pi$ at $t=0$ and linearly varies with time. By setting six parameters in Eq. (4) appropriately, the QPS simulates any time-varying component of the voiced speech in a local time period close to t_c .

The QPS is effective for investigating the nonstationarity because its spectrum is analytically predictable in a certain condition. A short-time Fourier transform (STFT) of the QPS at the center time is expressed as

$$Q_0(\omega) = \int_{-T}^T w(t)q(t)\exp(-j\omega t)dt, \quad (5)$$

$$w(t) = \exp(-\gamma \cdot t^2)/\sqrt{2\pi}. \quad (6)$$

Here, $w(t)$ corresponds to a Gaussian window function in which γ is a positive real constant. Let us consider the case in which the amplitude envelope of $w(t)\exp[A(t)]$ is convex upward and can be ignored in the integral for $t < -T$ and $t > T$. This condition is written as

$$|a_1| < (\gamma - a_2) \cdot T - \sqrt{(\gamma - a_2) \cdot M \cdot \log(10)/10}, \quad (7)$$

where the peak envelope is M dB or greater than the edge. Under the condition, the spectrum $Q_0(\omega)$ can be approximated to a simple formula at $\omega > 0$ as

$$Q_0(\omega) \approx \frac{1}{2\sqrt{\gamma - a_2}} \exp \left\{ \frac{(\omega + j \cdot \alpha_1)^2}{2(\alpha_2 - \gamma)} + \alpha_0 \right\}, \quad (8)$$

$$\alpha_n = a_n + j \cdot p_n. \quad (9)$$

This approximated spectrum is essentially equivalent to one presented in Marques and Almeida (1989). Most of the QPS, which simulates voiced speech components, satisfies the condition of Eq. (7) if appropriate window constant γ is selected.

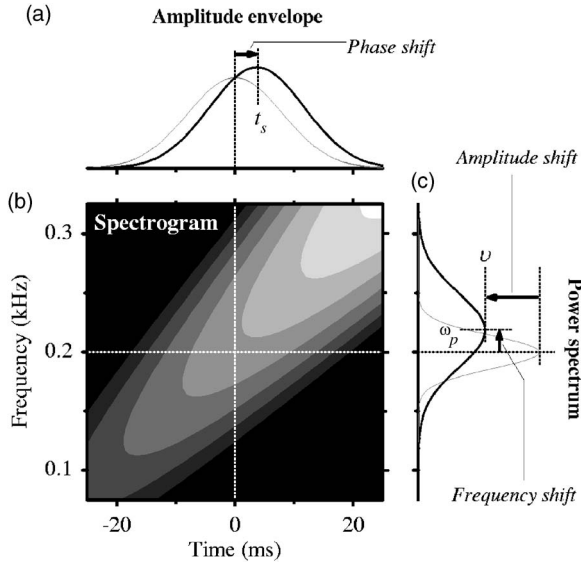


FIG. 1. LSA-based analysis for a time-varying sinusoidal signal whose instantaneous logarithmic amplitude $A(t)$ and frequency $F(t)$ linearly increase with time. (a) Thick line represents an amplitude envelope of $w(t) \cdot \exp[A(t)]$, where $w(t)$ corresponds to a Gaussian window function. Thin line represents another envelope for a signal where $A(t)$ is stationary. (b) Spectrogram of the time-varying sinusoid, where brighter color indicates greater power. (c) Thick line represents power spectrum at $t=0$ for the time-varying sinusoid. Thin line represents the spectrum for the stationary signal.

B. Local stationarity assumption

Next, we examine the parameter determination for QPS based on a local stationarity assumption. One of the widely accepted algorithms is a spectral peak-picking, in which instantaneous frequency of the signal component is determined from a peak frequency in the power spectrum. In addition, the instantaneous amplitude is determined from a peak amplitude. This algorithm assumes that the amplitude and frequency may be stationary within a local analysis period. As indicated in Eq. (8), logarithmic power spectrum of QPS is approximated to a quadratic function of angular frequency ω . Since this function is convex upward under the condition of Eq. (7), its peak frequency ω_p , peak bandwidth ω_B (−6 dB), and logarithmic peak amplitude ν can be estimated from Eq. (8) as

$$\omega_p = \arg \max_{\omega} |Q_0(\omega)| = p_1 - a_1 \cdot p_2 / a_G, \quad (10)$$

$$\omega_B = \sqrt{-2 \log(2)(a_G^2 + p_2^2) / a_G}, \quad (11)$$

$$\nu = \log |Q_0(\omega_p)| = a_0 - a_1^2 / 2a_G - \log(a_G^2 + p_2^2) / 4 - \log(2), \quad (12)$$

$$a_G = a_2 - \gamma. \quad (13)$$

The peak frequency ω_p agrees with p_1 , the instantaneous angular frequency at the center time, only when the second term of the Eq. (10) is zero. The peak amplitude ν is also affected by nonstationary parameters a_1 , a_2 , and p_2 [Eq. (12)]. Thus, the peak-picking algorithm produces inevitable errors in the frequency and amplitude determination. This is illustrated in Fig. 1, where the input signal is a QPS and $A(t)$

and $F(t)$ linearly increase with time [Fig. 1(b)]. A peak time of the amplitude envelope is shifted from the center to $t_s = a_1 / \gamma$ since the $A(t)$ is increasing [Fig. 1(a)]. This envelope can be regarded as another Gaussian window. Then, the spectrum at $t=0$ is equivalent to that of the constant-amplitude frequency-modulated (FM) signal at $t=t_s$. Since $F(t_s)$ is greater than $F(0)$, the peak frequency ω_p in the power spectrum is shifted to a higher frequency [Fig. 1(c)]. The peak bandwidth ω_p becomes greater as energy for the FM signal is spread over frequency, which causes decrease of the peak amplitude ν .

These errors are inevitable for the other LSA-based algorithms such as autocorrelation and cepstrum in F0 determination. For example, the cepstrum determines F0 based on the STFT of the logarithmic power spectrum. As indicated in Eq. (10), peak frequency for each signal component is individually shifted according to their nonstationary parameters. This shift distorts a harmonic structure in the spectrum, which induces the determination error. In addition, the peak bandwidth may be another error source. The bandwidth ω_B for the higher component greatly increases because p_2 for the k th component is k times greater than the fundamental in a harmonic signal. When the bandwidth exceeds a frequency interval between the neighboring peaks, their interference becomes substantial. This often reduces the accuracy of the cepstrum.

C. Time corrected instantaneous frequency

Next, we examine another conventional method, time-corrected instantaneous frequency originally proposed by Kodera and his colleagues (Kodera *et al.*, 1976). This method can accurately estimate the instantaneous frequency for a nonstationary amplitude and/or frequency-modulated line component because it does not assume the local stationarity of the signal. The algorithm is summarized as follows. First, a spectrogram is calculated using STFT from the input signal. Then, channelized instantaneous frequency (CIF) and local group delay (LGD) are estimated at every time-frequency point (ω, t_c) in the spectrogram as

$$\text{CIF}(\omega, t_c) = \frac{\partial}{\partial t_c} \arg(\text{STFT}(\omega, t_c)), \quad (14)$$

$$\text{LGD}(\omega, t_c) = -\frac{\partial}{\partial \omega} \arg(\text{STFT}(\omega, t_c)). \quad (15)$$

Using these variables, the point (ω, t_c) is remapped in a new representation as

$$(\omega, t_c) \rightarrow [\text{CIF}(\omega, t_c), t_c + \text{LGD}(\omega, t_c)]. \quad (16)$$

The remapped representation shows highly resolved instantaneous frequency tracks. For example, the spectrogram in Fig. 1(b) is remapped into a straight line corresponding to the $F(t)$. Thus, the shifts of center time t_s and peak frequency ω_p , observed in the LSA-based analysis, can be cancelled using the TCIF. This is a remarkable advantage of the method.

However, the problem of determining the instantaneous amplitude remains. In the TCIF, the amplitude at (ω, t_c) in

the original spectrogram is used as is in the remapped representation. Since the peak amplitude in the power spectrum is affected by the nonstationarity [Eq. (12)], the remapped amplitude also shows a dependency on it. In other words, the peak amplitude shift in Fig. 1(c) is not cancelled by the remapping process. Thus, the TCIF cannot accurately determine $A(t)$ for the time-varying component, similar to the LSA-based algorithm. The method that can accurately determine both the instantaneous amplitude and frequency for a time-varying sinusoid has not yet been proposed.

III. LOCAL VECTOR TRANSFORM

A voiced speech consists of nonstationary sinusoidal components, in which both the instantaneous amplitude and frequency are time varying. As mentioned in the previous section, it is difficult to analyze these parameters accurately by using conventional algorithms. To overcome this difficulty, we propose a new method termed a local vector transform (LVT) as follows.

A. Basic idea

The QPS is a model for signal component of voiced speech within a local time period. For this signal, it is sufficient to determine six parameters (a_0 , a_1 , a_2 , p_0 , p_1 , and p_2) for analyzing its instantaneous amplitude and frequency accurately. Since the QPS spectrum $Q_0(\omega)$ is approximated to a simple form [Eq. (8)], its first and second derivatives with respect to angular frequency are uniquely calculated as

$$Q_1(\omega) = \frac{\partial Q_0}{\partial \omega} = \left(\frac{\omega + j \cdot \alpha_1}{\alpha_2 - \gamma} \right) \cdot Q_0, \quad (17)$$

$$Q_2(\omega) = \frac{\partial^2 Q_0}{\partial \omega^2} = \left[\frac{1}{\alpha_2 - \gamma} + \left(\frac{\omega + j \cdot \alpha_1}{\alpha_2 - \gamma} \right)^2 \right] \cdot Q_0. \quad (18)$$

These derivatives can be obtained from the signal as simple differentials of the spectrum or by calculating additional two STFTs for signals of $-j \cdot t \cdot q(t)$ and $t^2 \cdot q(t)$. Then, an inverse transformation from the spectra Q_0 , Q_1 , and Q_2 to the signal parameters is expressed as

$$\alpha_0(\omega) = \log \left(\frac{2Q_0^2}{\sqrt{Q_1^2 - Q_0Q_2}} \right) + \frac{Q_1^2}{2(Q_1^2 - Q_0Q_2)}, \quad (19)$$

$$\alpha_1(\omega) = j \cdot \left(\omega - \frac{Q_0Q_1}{Q_0Q_2 - Q_1^2} \right), \quad (20)$$

$$\alpha_2(\omega) = \frac{Q_0^2}{Q_0Q_2 - Q_1^2} + \gamma. \quad (21)$$

Thus, the parameters can be uniquely determined from the input signal without iterative optimization. Since the determined amplitude and phase are not only the instantaneous values (α_0) but also the first and second time derivatives (α_1 and α_2), the determined instantaneous frequency forms a vector comprising its instantaneous value and the ratio of the change within a local time period. Therefore, we call this determination process a local vector transform (LVT).

For a voiced speech signal consisting of many sinusoidal components, LVT can determine the parameters for each component if the following two conditions are satisfied: First, $A(t)$ and $P(t)$ must be approximated to quadratic functions of time within a local analysis segment close to $t=t_c$. Determination errors may occur if these approximations are not valid (track-complexity errors). Second, the power for a component must be negligible at peak frequencies for neighboring components in the power spectrum. Determination errors may occur if there is non-negligible interference between the components (interference error).

B. Example of parameter determination

Next, the algorithm is explained in detail using a voiced speech signal as an example. Figure 2(a) shows a spectrogram of the signal $x(t)$ uttered by a female, where the time-varying feature is observed in both the amplitude and frequency for each component. Figure 2(b) shows amplitude and phase spectra for the signal at $t_c=100$ ms, which are calculated from a STFT of $X_0(\omega)$. Three prominent peaks, corresponding to the lower three harmonic components, are observed in the amplitude spectrum. Since the peaks are significantly apart from each other in the frequency, their interferences can be negligible in this case. For natural utterances, the interferences are often observed in the higher harmonic components because the peak bandwidth ω_B becomes larger according to $|p_2|$ [Eq. (11)], where p_2 for the k th component is k times that for the first component in a harmonic signal.

Based on the spectrum and its first and second derivatives, the six signal parameters are determined using Eqs. (19)–(21). Figures 2(c) and 2(d) show determined signal parameters of a_0 , a_1 , a_2 , p_0 , p_1 , and p_2 . Each determined parameter is almost constant in the frequency close to the peak in the power spectrum. In contrast, the parameters are fluctuated at the frequency between the peaks since the interferences cannot be negligible in the region. To reduce misdetections caused by the fluctuations, only the parameters are selected at the frequencies which fulfill the following condition:

$$\omega_B = \text{Im}[\alpha_1(\omega_B)] = p_1(\omega_B). \quad (22)$$

In Figs. 2(c) and 2(d), open circles represent the selected frequencies.

Based on each of the selected parameter sets, the instantaneous amplitude can be approximated to a quadratic function within a local analysis period. This is illustrated on the right side of Fig. 2(c). Since the instantaneous phase is also approximated to another quadratic function, the instantaneous frequency can be expressed as a linear function [(Fig. 2(d)].

Figure 2(e) shows determined instantaneous amplitudes for the components of the input speech. Each curve represents a quadratic function calculated from the LVT parameters. Figure 2(f) shows determined instantaneous frequency, where each segment represents a linear function. As shown in the figure, the amplitude and frequency are smoothly aligned at least for the lower three harmonic components.

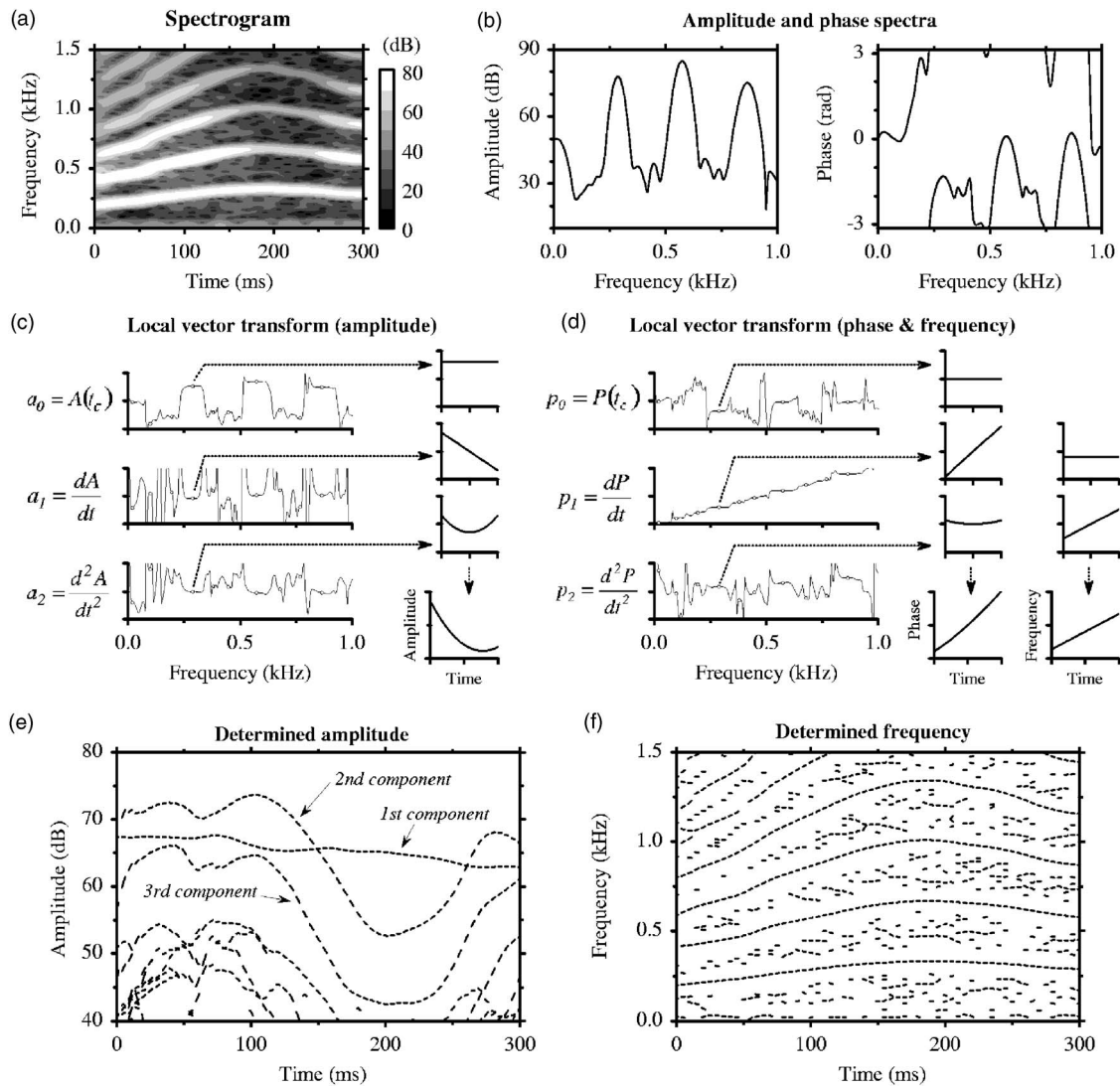


FIG. 2. (a) Spectrogram of a voiced speech signal uttered by a female. (b) Amplitude and phase spectra for the speech at $t=100$ ms. Left panel represents the amplitude spectrum in which three prominent peaks corresponds to the first three harmonic components. Right panel represents the phase spectrum. (c) Solid lines represent the determined a_0 , a_1 , and a_2 from the spectra using LVT. Open circles indicate the selected frequencies under a criterion of Eq. (22). At every selected frequency, the instantaneous logarithmic amplitude is approximated to a quadratic function corresponding to the determined parameters as shown on the right side of the figure. (d) Solid lines represent the determined p_0 , p_1 , and p_2 from the spectra. The instantaneous phase at each selected frequency is approximated to a quadratic function of the determined parameters. The instantaneous frequency is approximated to a linear function of time. (e) Solid lines represent the determined quadratic functions of the instantaneous amplitude for the speech at every 5 ms. Tracks corresponding to the first three harmonic components are indicated by arrows. (f) Solid lines represent the determined linear functions of the instantaneous frequency.

This implies that the track-complexity errors may not be crucial for these components. The actual determination accuracy will be evaluated quantitatively in the next section.

IV. EXPERIMENT

As mentioned in the previous section, LVT can accurately determine the signal parameters for a simple time-varying sinusoid (QPS). However, the determination accuracy for actual voiced speech signals is not clear because their amplitude and frequency often vary with more complicated time-varying functions than those of the QPS. In this section, the effectiveness of the LVT is examined for synthesized and naturally uttered speech signals.

A. Non-stationarity in voiced speech

For the quantitative evaluation, synthesized signals for which the parameters were clearly defined should be used as the input with appropriate ranges of the nonstationarity for natural utterances. Thus, the nonstationarity in naturally uttered voiced speech was investigated first. The instantaneous frequency and amplitude of voiced speech components can be uniquely determined by F0 and spectral envelope. For the envelope, some models have been proposed [e.g., Klatt (1980)], for which time-varying characteristics are realized by smooth transition of the formant frequencies. However, there have been few studies that model the time-varying F0 in natural utterances. Therefore, the nonstationarity in F0 was investigated.

For this purpose, voiced speech signals were acousti-

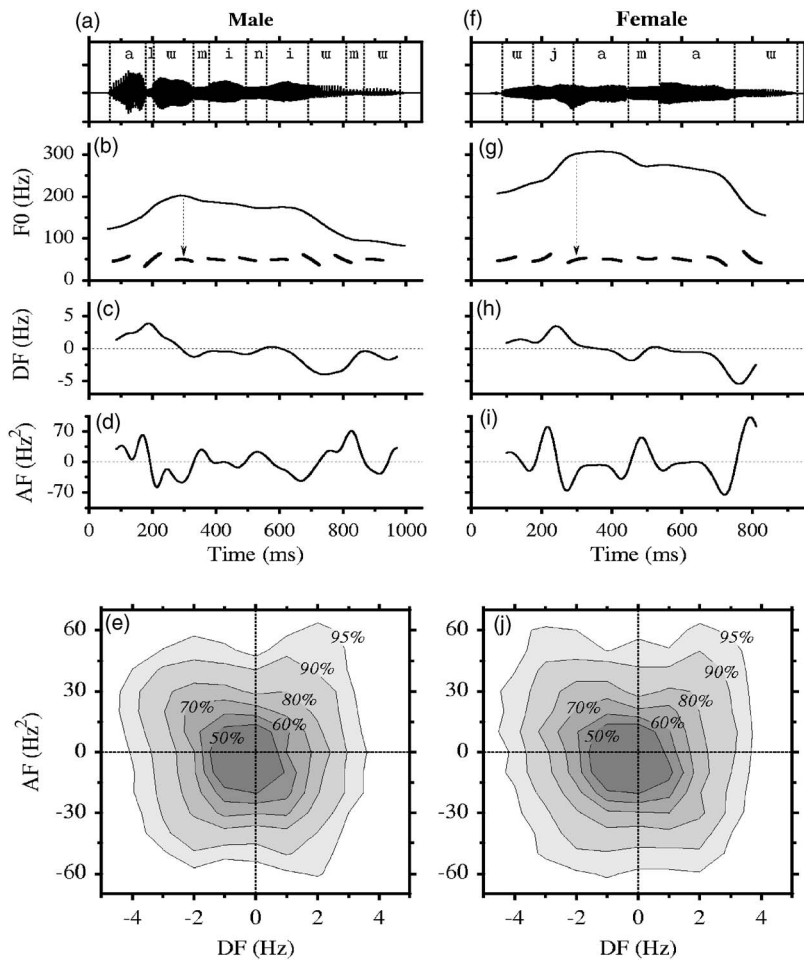


FIG. 3. F0 nonstationarity in voiced speech signals. (a) A waveform for a speech signal uttered by a male speaker. Dotted lines represent phoneme boundaries. (b) Thin line represents the determined F0 for the speech. Thick lines represent the appropriate quadratic functions. (c) Normalized velocity (DF) of the F0. (d) Normalized acceleration (AF) of the F0. (e) The probabilistic distribution of the F0 nonstationarity for every utterance of male speakers. The probability is indicated by color. Each point represents the probability at which corresponding DF and AF may appear in an utterance. (f)–(i) Same process for a speech signal uttered by a female speaker. (j) The probabilistic distribution for female speakers.

cally analyzed. The speech signals were Japanese words extracted from the Advanced Telecommunications Research Institute International (ATR) Japanese Speech Database (Kuwabara *et al.*, 1989), in which every speech was well articulated and carefully recorded in quiet rooms. To evaluate the time-varying feature effectively, twelve words consisting of only voiced phonemes (vowels /a/, /i/, /u/, /e/, and /o/; semivowel /j/; flap /l/; nasals /m/ and /n/) were selected from the database. The speakers were 16 male and 16 female adults. Thus, 384 speech signals were analyzed.

In the voiced period of each word, F0 was determined using a TCIF at every 1 ms with a 50 ms local analysis window. The half-pitch and double-pitch errors were manually corrected by an operator. Then, each F0 track was smoothed by a low-pass filter to reduce rapid changes that might be the artifacts. Figure 3(b) shows an example of a calculated F0 track for a word uttered by a male speaker.

LVT approximates the instantaneous frequency for each component to a linear function within a local analysis period. To examine the effect of this simple approximation for voiced speech, the number of segments contained in voiced speech should be quantified where this linear approximation is invalid. Thus, each F0 track was nonlinearly approximated within a local analysis period (50 ms) to

$$F_0(t) = CF \cdot [1 + DF \cdot (t - t_c) + AF \cdot (t - t_c)^2/2]. \quad (23)$$

Here, CF , DF , and AF were uniquely determined from the track using a least mean square measure. The estimated qua-

dratic functions are indicated at the bottom of Fig. 3(b). DF and AF are shown in Figs. 3(c) and 3(d), where the parameters continuously varied with time in certain ranges. For every speech sample uttered by the male speakers, the mean probabilistic distribution was calculated, as shown in Fig. 3(e), in which each point represents a probability that the nonstationarity corresponding to the position might appear in an utterance. Figures 3(f)–3(j) show the same process for the female speakers. The mean and standard deviation of CF , DF , and AF are listed in Table I. Although there was significant gender dependence in CF , the nonstationarity parameters DF and AF distributed in a similar range regardless of gender [Figs. 3(e) and 3(j)]. In the following, these distributions were used as the typical F0 nonstationarity in natural utterances.

TABLE I. F0 nonstationarity in voiced speech signals (384 utterances). Each F0 track is approximated to a quadratic function of $F_0(t) = CF \cdot (1 + DF \cdot t + AF \cdot t^2/2)$ within a local analysis period (50 ms). The mean and standard deviation of the coefficients are listed in the first three columns (mean \pm SD).

	CF (Hz)	DF (Hz)	AF (Hz ²)	Period (ms)
Male	140.3 \pm 27.1	-0.33 \pm 1.40	-0.80 \pm 22.42	107,561
Female	258.8 \pm 46.6	-0.25 \pm 1.48	-0.89 \pm 23.90	114,619
Total	201.5 \pm 70.6	-0.29 \pm 1.44	-0.85 \pm 23.19	222,180

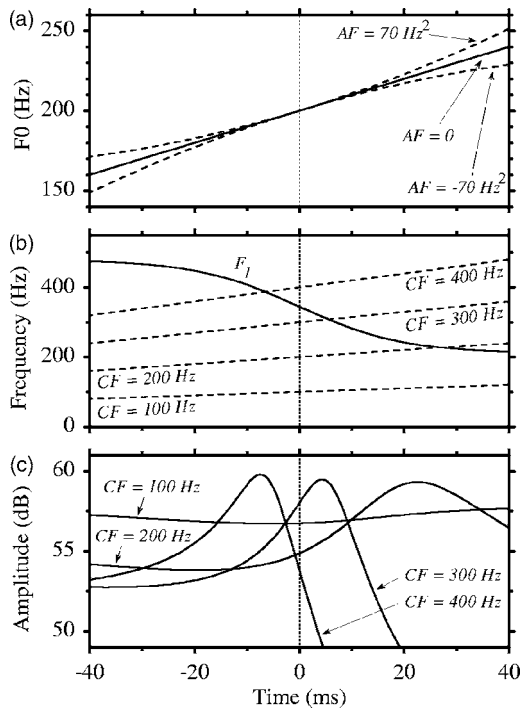


FIG. 4. (a) Three F0 tracks of synthesized speech signals. Solid line represents the track of $DF=5$ Hz and $AF=0$. Dashed lines represent the track of $DF=5$ Hz and $AF=\pm 70$ Hz. (b) Solid line represents the first formant frequency for the transition from vowel /e/ to /i/ of the male speaker. Dashed lines represent F0 tracks of $CF=100, 200, 300,$ and 400 Hz. For every track, DF and AF are 5 Hz and zero, respectively. (c) Solid lines represent the instantaneous amplitudes of the first harmonic components whose F0 tracks are shown in (b). Each track has the same spectral envelope of /e/ to /i/ transition of male speaker.

B. Input signals

Based on the acoustic analysis, a speech-like harmonic signal $h(t)$ was synthesized as follows. The signal consisted of K harmonic components as

$$h(t) = \sum_{k=1}^K E[t, k \cdot F_0(t)] \cos\left(2\pi k \cdot \int_0^t F_0(\tau) d\tau + \varphi_k\right). \quad (24)$$

Here, $E(t, f)$ and $F_0(t)$ represent the time-varying spectral envelope and fundamental frequency, respectively. The number of the components, K , was set to 50 for every signal. $F_0(t)$ was defined by Eq. (23) with $t_c=0$. CF was varied over 100, 200, 300, and 400 Hz. DF ranged from -5 to 5 Hz with intervals of 0.5 Hz. AF ranged ± 70 Hz² with a step of 7 Hz². Then, 1 764 tracks were prepared, which covered

almost the entire range of F0 nonstationarity in the analyzed speech signals [Figs. 3(e) and 3(j); Table I]. Three representative F0 tracks are shown in Fig. 4(a). It should be noted that LVT may produce inevitable errors for the signal of greater AF because its instantaneous frequency cannot be approximated to a linear function accurately.

The spectral envelope $E(t, f)$ was determined from the frequency response of the cascade-Klatt synthesizer (Klatt, 1980). The synthesizer had five formants, and each formant frequency $F_n(t)$ smoothly varied with time as

$$F_n(t) = [1 - R(t)] \cdot F_n^P + R(t) \cdot F_n^S, \quad (25)$$

$$R(t) = \exp(\lambda t) / [1 + \exp(\lambda t)]. \quad (26)$$

Here, F_n^P and F_n^S were selected from n th formant frequencies for the five Japanese vowels shown in Table II. Thus, the spectral envelope simulated the transition from a vowel to another. Two sets of the formant frequencies, extracted from isolated vowels uttered by a male and female speaker, were prepared to reduce the speaker-specific effect. Although the number of possible vowel combinations was 25 for each set, 10 redundant combinations were removed (e.g., /ai/ and /ia/). Then, 30 spectral envelopes were used in the experiments. For every envelope, λ was set to 100 Hz. The initial phase $\varphi_k(t)$ was determined from phase response of the Klatt synthesizer at $t=0$.

For all the combinations of the F0 tracks and spectral envelopes, 52 920 signals were synthesized. In most components, the instantaneous amplitudes varied with more complicated function of time than the quadratic one as a reflection of time-varying features for $F_0(t)$ and $E(t, f)$. Figure 4(c) shows four amplitude tracks of the first harmonic components. The first formant frequency $F_1(t)$ and the tracks of the first four harmonics are indicated in Fig. 4(b). As shown in the figure, the amplitude function became complicated when $F_1(t)$ was close to $F_0(t)$. This amplitude complexity may be another error source of the LVT.

C. F0 determination

Next, the accuracy of F0 determination of LVT was examined for the synthesized speech-like signals. The F0 was determined only at $t=0$ for simplicity where the ideal value was CF for every signal [Eq. (23)]. To avoid half- and double-pitch errors, a simple comb filter was utilized in the LVT to select the best ω_B . The possible F0 ranged from 50 to 1 000 Hz. The Gaussian window constant γ was set to $1/64$ kHz². In addition, four conventional algorithms were

TABLE II. The formant frequencies for five Japanese vowels (Hz). The bandwidth for each formant is identical for male and female speakers: $B_1=90$ Hz; $B_2=110$ Hz; $B_3=130$ Hz; $B_4=250$ Hz; and $B_5=300$ Hz.

	Male					Female				
	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/
F1	700	210	230	480	350	1 050	250	430	720	680
F2	1 140	2 520	1 050	2 130	640	1 730	3,140	1 760	2 570	1 030
F3	2 700	3 520	2 360	2 610	2 720	3 220	3 610	3 080	3 350	3 030
F4	3 660	3 940	3 560	3 730	3 250	4 190	4 580	4 190	4 130	4 030
F5	4 500	4 500	4 500	4 500	4 500	5 400	5 400	5 400	5 400	5 400

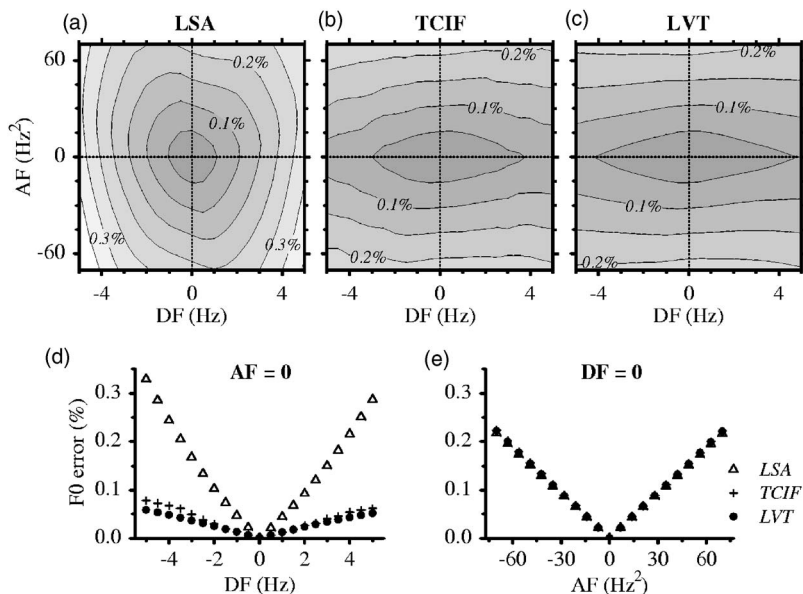


FIG. 5. Mean F0 determination errors (%) for the synthesized speech signals. (a) Errors of the spectra of the peak-picking algorithm. Brighter color represents greater error. (b) Errors of the TCIF. (c) Errors of the LVT. (d) Errors of the peak picking (triangles), TCIF (crosses), and LVT (circles) for the signals where $AF=0$. (e) Errors for the signals of $DF=0$.

examined for the comparison. The spectral peak-picking and TCIF were realized with the same window function as the LVT. The autocorrelation and cepstrum, which are widely accepted LSA-based F0 determination algorithms, were also tested. In these conventional algorithms, a range of possible values of F0 was limited from $0.8 \cdot CF$ to $1.2 \cdot CF$ to avoid half- and double-pitch errors. The determination accuracy was evaluated by a relative error originally proposed by Rabiner and his colleagues (Rabiner *et al.*, 1976) as

$$\epsilon = \frac{|F0_{EST} - CF|}{CF} \times 100 (\%). \quad (27)$$

Here, $F0_{EST}$ corresponds to the F0 determined by the algorithms.

Figures 5(a)–5(c) show the mean error ϵ of the peak picking, TCIF, and LVT, respectively. The error of the LVT increased according to the absolute value of AF , which was the track-complexity error (Sec. II A). Despite this AF dependence, the LVT marked the smallest errors in the examined algorithms (Table III). This result suggested that the effect of the track-complexity error was not crucial in the F0 determination, at least for the examined range of nonstationarity.

The TCIF determined F0 with almost the same accuracy as the LVT [Fig. 5(b)]. There were few differences between the errors of the algorithms. This is reasonable because neither of these algorithms assumes the local stationarity in the instantaneous frequency. In contrast, the LSA-based peak-

picking algorithms could not determine accurate instantaneous frequency for the time-varying signal, in which the error increased according to the absolute values of AF and DF [Fig. 5(a)]. Similar results were obtained in the other LSA-based algorithms. There were often extremely large errors in the determination of the autocorrelation and cepstrum, in which the interferences of the neighboring peaks in the higher components became additional error sources.

D. A0 determination

Next, the determination accuracy for the instantaneous amplitude was examined. For simplicity, the amplitude of the first component (A0) was determined at $t=0$, where the ideal A0 was $E(0, CF)$. Three algorithms, the spectral peak picking, TCIF, and LVT were examined. For the peak picking and TCIF, A0 was determined from the spectral peak amplitude ν with constant addends of $[\log(4\gamma)/2]$ to minimize the error for the stationary signal [Eq. (11)]. The determination accuracy was evaluated by the absolute value of the difference between the input and estimated amplitudes in logarithmic scale (dB).

Figures 6(a)–6(c) show the mean A0 errors of the peak picking, TCIF, and LVT, respectively. The errors of the conventional algorithms were similar to each other, in which the error increased according to the absolute value of DF [Figs. 6(a) and 6(b)]. In addition, there were non-negligible mean errors even for the stationary F0 track ($DF=AF=0$). This was caused by the nonstationarity of the instantaneous amplitude since 20 of 30 spectral envelopes, corresponding to the successive vowels, were time-varying in the duration for each F0 track. Thus, we concluded that it was difficult to determine A0 accurately using these algorithms. In contrast, the mean errors of the LVT had little dependence on AF and DF . It marked the smallest errors in the tested algorithms (Table IV). This result suggested that the effect of the track-complexity errors was not crucial in determining A0.

TABLE III. F0 determination errors (%). The mean errors for 52 920 synthesized signals are listed in the first row while the maximum errors are listed in the second row.

	LSA			TCIF	LVT
	Autocorrelation	Cepstrum	(peak picking)		
Mean	1.244	0.697	0.197	0.131	0.123
Maximum	19.385	3.628	2.334	1.311	0.860

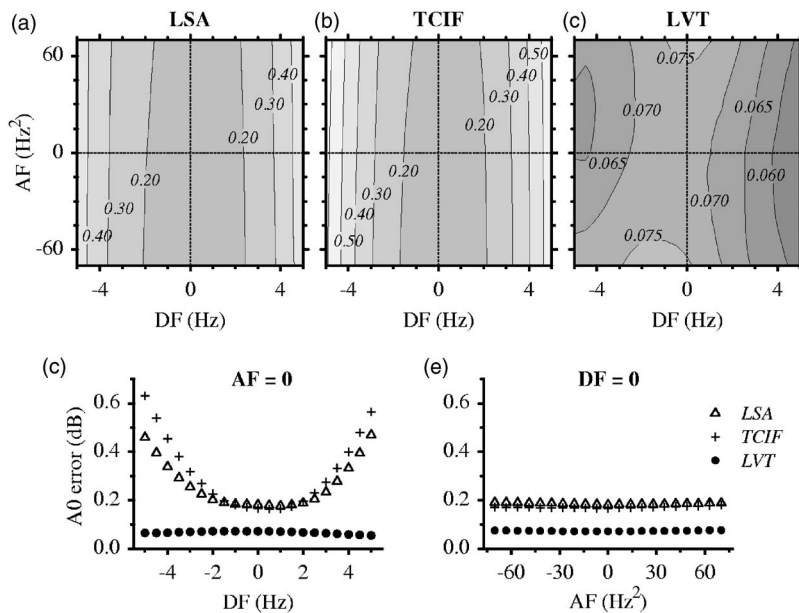


FIG. 6. Mean amplitude determination errors (dB) for the synthesized speech signals. (a) Errors of the spectra of the peak-picking algorithm. Brighter color represents greater error. (b) Errors of the TCIF. (c) Errors of the LVT. (d) Errors of the peak-picking (triangles), TCIF (crosses), and LVT (circles) for the signals where $AF=0$. (e) Errors for the signals of $DF=0$.

E. Signal reconstruction

For the synthesized input signals, the LVT could determine F0 and A0 with the highest accuracy in the tested algorithms. If the determined parameters are accurate, the signal reconstructed from them must agree with the corresponding component in the input. To examine this, the first harmonic component for voiced speech was reconstructed from the determined parameters.

The input was a voiced speech uttered by a female speaker [Fig. 7(a)]. In the voicing period, the instantaneous amplitude, frequency, and phase were determined using the LVT at every sampling point. For the comparison, the amplitude and frequency were also determined using the spectral peak-picking and TCIF. Figure 7(b) shows the determined frequencies, where the mean of the three methods is subtracted from each value. As shown in the figure, LVT and TCIF determined almost the same frequency. This was consistent with the result of the F0 determination experiment. Figure 7(c) shows the determined amplitudes, in which the value determined by LVT was greatly different from those of the conventional algorithms. This was also consistent with the result of the A0 determination experiment.

In LVT, the component signal could be uniquely reconstructed from the determined amplitude and phase. However, the conventional algorithms needed appropriate initial phases to reconstruct the signal component because they could not determine the instantaneous phase. Then, the initial phase for each algorithm was determined from an optimization process to minimize the energy of the residual signal, which corre-

sponded to the difference between the input and estimated signals. Figures 7(d) and 7(e) show the spectrograms of the residual signals determined by the peak-picking and TCIF with the optimized initial phases, respectively. As shown in the figures, the energy for the first harmonic component in the residual spectrogram was reduced but remained for each algorithm. This indicates that at least one of the determined amplitudes or frequencies was not accurate in the algorithms. In contrast, the energy for the first harmonic component was almost completely suppressed in the residual signal of LVT [Fig. 7(c)], which suggested that both the amplitude and frequency (and phase) were accurately determined by this method.

The residual energy of the peak picking and TCIF was extremely small at the period close to 250 ms, where their determined amplitudes and frequencies almost agreed with those determined by LVT [Figs. 7(b) and 7(c)]. This was consistent with the results of the experiments using the synthesized signals, that the differences between the algorithms were small for the signal where F0 was stationary. The F0 was almost constant at the period, as shown in Fig. 7(a).

V. DISCUSSION

We proposed a new method LVT to analyze instantaneous frequency and amplitude for time-varying sinusoidal components. In this method, the frequency and logarithmic amplitude were approximated to linear and quadratic functions of time, respectively, within a local analysis period (Fig. 2). These time-varying models distinguish this method from conventional algorithms which assume the local stationarity of the parameters such as spectral peak-picking (McAulay and Quatieri, 1986). The effectiveness of the LVT was quantitatively evaluated in parameter determination experiments for the first component of speech-like harmonic signals.

In the frequency determination, LVT marked the highest mean accuracy among the examined methods (Table III). The accuracy of TCIF was almost equal to the LVT because nei-

TABLE IV. A0 determination errors (dB). The mean and maximum errors are listed in the first and second row, respectively.

	LSA (peak-picking)	TCIF	LVT
Mean	0.264	0.312	0.068
Maximum	2.656	3.316	1.105

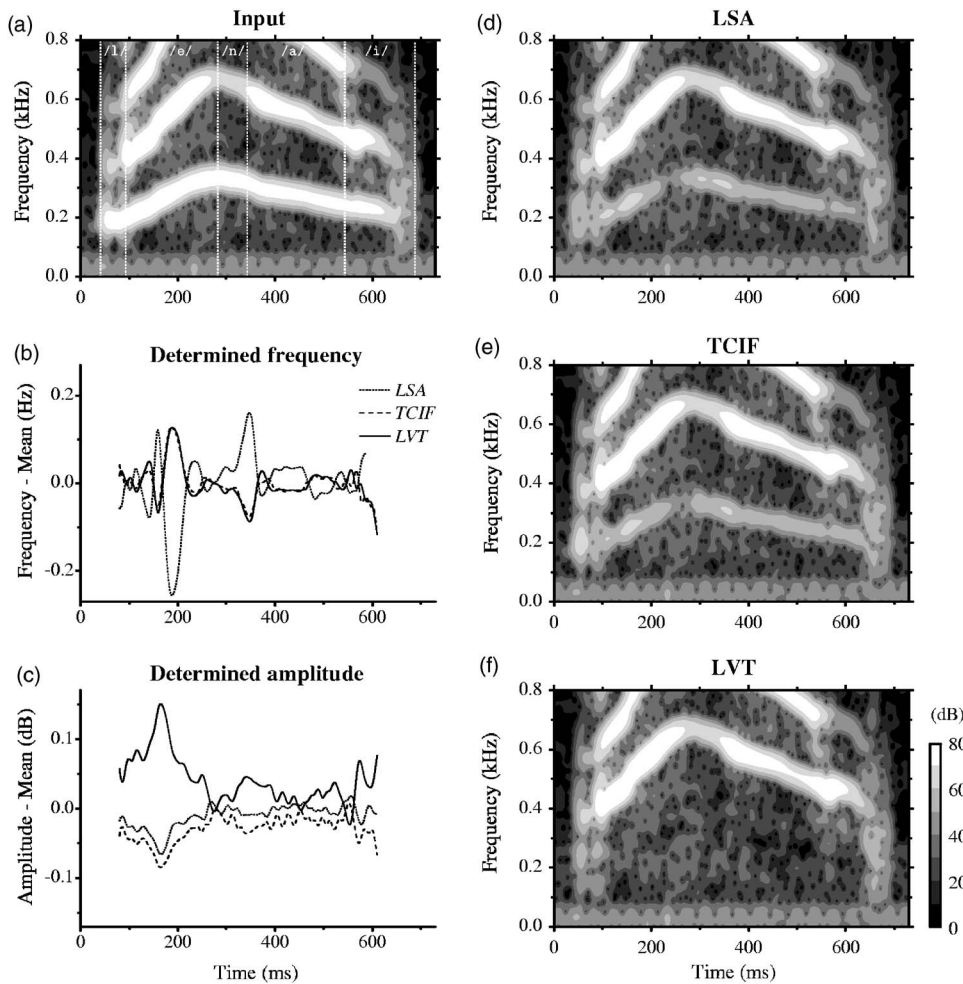


FIG. 7. Reconstruction of the component for a voiced speech signal. (a) Spectrogram of the input speech uttered by a female. Brighter color represents greater power. Dotted lines correspond to phoneme boundaries. (b) Determined instantaneous frequencies for the first harmonic component using the spectral peak picking (dotted line), TCIF (dashed line), and LVT (solid line). Each frequency is normalized by the mean value of the three algorithms at each time point. (c) Determined instantaneous amplitudes for the first harmonic component using the peak picking, TCIF, and LVT. (d) Spectrogram of the residual signal. The residual signal corresponds to the input subtracted from the signal, which is reconstructed from the determined parameters of the spectral peak picking. (e) Spectrogram of the residual signal using TCIF. (f) Spectrogram of the residual signal using LVT.

ther of these methods assume the local stationarity of the instantaneous frequency. In contrast, the mean accuracy for spectral peak-picking, autocorrelation, and cepstrum, all of which are based on the LSA, did not reach the level of the LVT and TCIF. The accuracy difference between the two groups became greater for the signal where the absolute value of DF was large [Fig. 5(d)]. In other words, the effectiveness of the LVT and TCIF is more prominent when the frequency is increasing or decreasing in a local analysis period. Since the LVT approximates the frequency to a linear function, inevitable errors occur in the determination where the frequency track forms a local peak or valley [Fig. 5(e)]. Despite this drawback, LVT is the best method for accurate determination of the instantaneous frequency in the examined algorithms because these errors are also observed in the other conventional algorithms.

The fundamental frequency is one of the important characteristics in voiced speech. The F0 determination algorithms have been studied extensively (Noll, 1967; Rabiner *et al.*, 1976; Rabiner, 1977; Hess, 1983; Krubsack and Niederjohn, 1991; Rouat *et al.*, 1997; Kawahara *et al.*, 1999; Wu *et al.*, 2003), in which several aspects such as voicing detection, noise robustness, half- or double-pitch errors, and multiple pitch were focused. Nonstationarity is one of the important research topics. As shown in the acoustic analysis (Sec. IV A), F0 is time varying in most parts of the utterance that causes determination errors for the LSA-based algorithms.

The high-accuracy frequency determination for a time-varying sinusoid makes LVT noteworthy in F0 determination algorithms.

For the analysis of instantaneous amplitude, LVT also marked the highest accuracy in the examined algorithms (Table IV). The determination accuracy of spectral peak picking and TCIF did not reach the level of the LVT. These algorithms incorrectly determine the smaller amplitude than the actual value for time-varying signals as a reflection of the peak amplitude shift in the power spectrum (Fig. 1). Although the peak amplitude is affected by several nonstationary parameters [Eq. (12)], the determination error was mainly dominated by DF in the tested range of the parameters. In other words, the amplitude determination errors of these algorithms become greater for the signal whose instantaneous frequency is increasing or decreasing. In contrast, the accuracy of LVT had little dependence on the F0 nonstationarity [Figs. 6(d) and 6(e)]. This indicates that a quadratic function is sufficient to approximate the local period of the instantaneous amplitude in the speech-like signals.

These results were obtained from the parameter determination for the synthesized harmonic signals. The range of the parameters was set to simulate voiced speech based on the acoustic analysis of natural utterances (Sec. IV A). The estimated probability distribution [Figs. 3(e) and 3(j)] was not free from the intrinsic characteristics of the input speech, such as context, speed of the utterance, and emotional state

of the speaker. In addition, only the small sets of formant frequencies (Table II) and their time-varying pattern, which determine the spectral envelope of the signal, were examined in the experiments. Thus, the parameters of the synthetic signals did not cover the entire range of natural utterances while they were observed in actual speech. Several aspects such as creaky voice and noise addition might affect the performance of LVT for analyzing voiced speech in daily conversational situation. This remains as a matter to be discussed further. However, it can be concluded at least that LVT accurately determines the instantaneous frequency and amplitude for the first harmonic component of the voiced speech signals which are well articulated and uttered in low-noise environment.

The reversibility between the determined parameters and signal waveform is an important characteristic in the sinusoidal modeling. It enables evaluation of the accuracy of the signal for which actual parameters are not clear, such as natural utterances. For example, the component signal reconstructed from the parameters determined by LVT agreed well with that of the input voiced speech [Fig. 7(f)]. This strongly supported the fact that the determined frequency and amplitude were accurate in the entire range of the period. In contrast, the non-negligible energy observed in the residual signals of the peak-picking and TCIF indicated that at least one of the parameters was incorrectly determined [Figs. 7(d) and 7(e)]. For the reversibility-based accuracy evaluation, LVT is effective for two reasons. First, it can accurately determine not only the instantaneous frequency but also amplitude of signal component. Second, it can also determine the instantaneous phase, which is indispensable for the signal reconstruction.

Simplicity is another important characteristic of the LVT. It can uniquely determine the signal parameters from the input spectra without any iterative calculations. This characterizes LVT from conventional methods, which need iterative optimization processes to determine signal parameters (Marques and Almeida, 1989; George and Smith, 1997). Compared with these methods, LVT is advantageous not only for the computational efficiency but also for the stability in the determination. The iterative optimization often falls in a local minimum when appropriate initial parameters are not selected.

One of the limitations of the LVT is that it is applicable only for the lower harmonic components of voiced speech. As shown in Fig. 2, this method can determine amplitude and frequency not only for the first harmonic component but also for some lower components. However, the determination errors may be great for the higher harmonic components because of the interference between the neighboring components. As mentioned in Sec. III B, the peak bandwidth corresponding to each component in the power spectrum increases according to the harmonic number when F0 is not stationary. The TCIF has the same limitation. Further research is needed to overcome this limitation.

In summary, LVT can accurately determine time-varying instantaneous amplitude and frequency for the lower harmonic components in voiced speech. Although its model of the time-varying nature is simple, it is concluded to be suf-

ficient to realize the reversibility between the parameters and waveform. The method is effective for F0 determination of voiced speech, in particular where the determination accuracy can be evaluated by comparing the input and reconstructed signals. We believe this method will shed new light on the acoustic analysis of speech.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the Sound Technology Promotion Foundation of Japan and Grant-in-Aid No. 17075003 from the Ministry of Education, Culture, Sports and Technology of Japan. We would like to thank Dr. Y. Suzuki and Dr. S. Makino of Tohoku University, Japan, for their fruitful discussions.

- Atal, B. S., and Hanauer, S. L. (1971). "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.* **50**, 637–655.
- Auger, F., and Flandrin, P. (1995). "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.* **43**, 1068–1089.
- Fulop, S. A., and Fitz, K. (2006). "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *J. Acoust. Soc. Am.* **119**, 360–371.
- George, E. B., and Smith, M. J. T. (1997). "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Trans. Speech Audio Process.* **5**, 389–406.
- Hedelin, P. (1981). "A tone-oriented voice-excited vocoder," in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, pp. 205–208.
- Hermus, K., Verhelst, W., Lemmerling, P., Wambaq, P., and Huffel, S. V. (2005). "Perceptual audio modeling with exponentially damped sinusoids," *Signal Process.* **85**, 163–176.
- Hess, W. (1983). *Pitch Determination of Speech Signals: Algorithms and Devices* (Springer, Berlin).
- Kawahara, H., Katase, H., de Cheveigne, A., and Patterson, R. D. (1999). "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity," *Proc. Eurospeech99*, pp. 2781–2784.
- Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.
- Kodera, K., de Villedary, C., and Gendrin, R. (1976). "A new method for the numerical analysis of non-stationary signals," *Phys. Earth Planet. Inter.* **12**, 142–150.
- Krubsack, D. A., and Niederjohn, R. J. (1991). "An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech," *IEEE Trans. Signal Process.* **39**, 319–329.
- Kuwabara, H., Sagisaka, Y., Takeda, K., and Abe, M. (1989). "Construction of ATR Japanese Speech Database as a Research Tool," ATR Technical Report, TR-I-0086.
- Laroche, J. (2000). "Synthesis of sinusoids via non-overlapping inverse Fourier transform," *IEEE Trans. Speech Audio Process.* **8**, 471–477.
- Marques, J. S., and Almeida, L. B. (1989). "Frequency-varying sinusoidal modeling of speech," *IEEE Trans. Acoust., Speech, Signal Process.* **35**, 763–765.
- McAulay, R. J., and Quatieri, T. F. (1986). "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.* **34**, 744–754.
- Nelson, D. (2001). "Cross-spectral methods for processing speech," *J. Acoust. Soc. Am.* **110**, 2575–2592.
- Noll, A. M. (1967). "Cepstrum pitch determination," *J. Acoust. Soc. Am.* **41**, 293–309.
- Rabiner, L. R., Cheng, M. J., Rosenberg, A. E., and McGonegal, C. A. (1976). "A comparative performance study of several pitch detection algorithms," *IEEE Trans. Acoust., Speech, Signal Process.* **24**, 399–418.
- Rabiner, L. R. (1977). "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. Acoust., Speech, Signal Process.* **25**, 24–33.
- Rouat, J., Chun, Y., and Morissette, D. (1997). "A pitch determination and voiced/unvoiced decision algorithm for noisy speech," *Speech Commun.* **21**, 191–207.
- Wu, M., Wang, D., and Brown, G. J. (2003). "A multipitch tracking algorithm for noisy speech," *IEEE Trans. Speech Audio Process.* **11**, 229–241.

Sound pressure level and spectral balance linearity and symmetry in the *messa di voce* of female classical singers^{a)}

Sally Collyer^{b)} and Pamela J. Davis^{c)}

National Voice Centre, The University of Sydney, Sydney NSW 2006, Australia

C. William Thorpe^{d)}

School of Communication Sciences and Disorders (C42), The University of Sydney, Sydney NSW 2006, Australia

Jean Callaghan

School of Contemporary Arts, The University of Western Sydney, Penrith NSW 2750, Australia

(Received 15 July 2006; revised 22 December 2006; accepted 2 January 2007)

The *messa di voce*, in its pure form a crescendo and decrescendo on one note, has been revered for centuries in classical singing, but the pedagogical assumptions of linearity and symmetry have received little critical assessment, especially across a wide fundamental frequency (F_0) range. Five trained female classical singers performed a total of 318 *messa di voce* across their musical F_0 range to identify its acoustical characteristics and the influence of F_0 . Sound pressure level (SPL) range was generally greater during crescendo at higher F_0 's and during decrescendo at lower F_0 's. Change in SPL during the *messa di voce* was predominantly nonlinear, and the shape of the SPL traces differed greatly between crescendo and decrescendo. Nonlinearity in SPL change was not related to SPL range but did show a F_0 influence in decrescendo. Change in spectral balance (0–2 vs 2–4 kHz) with respect to SPL change showed markedly more symmetry than linearity, so that changes in the mode of phonation during the *messa di voce* were dependent upon SPL regardless of whether the singer was in crescendo or decrescendo. Perceptual and physiological implications are discussed. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436639]

PACS number(s): 43.75.Rs [BHS]

Pages: 1728–1736

I. INTRODUCTION

The *messa di voce* in Western classical singing can be plainly described as a crescendo then decrescendo (an increase then decrease in intensity) on one note. This simple description, however, belies the profound expressivity and the technical complexity with which classical singing pedagogy has associated this exercise for centuries (Fleming, 2004; Tosi, 1987). In particular, pedagogy desires that change during the ideal *messa di voce* be both linear and symmetrical, but these assumptions have been questioned (Titze *et al.*, 1999). This study examined to what extent linearity and symmetry are characteristic of *messa di voce* performed by well trained female classical singers.

There are two major challenges in the *messa di voce*. First, the singer must demonstrate a smooth transition in all aspects associated with the crescendo (Miller, 1996; Tosi, 1987; Vennard, 1967). Laryngeal tension must be reduced as intensity rises (and raises vocal fold tension) or fundamental

frequency (F_0) will also tend to rise (Titze, 1989, 1992). Second, the change in dynamic across the crescendo and decrescendo must be symmetrical (Garcia, 1982; Reid, 1965). Singers are warned against dynamic variation which relies on breathiness for the piano and pressed phonation for the forte (Miller, 2000; Vennard, 1967). The decrescendo of the *messa di voce* has long been considered especially difficult, and many have noted the challenge of maintaining timbre so as not to descend into breathiness (Burgin, 1973; Miller, 1996; Vennard, 1967).

It is unclear whether the perception of loudness change in the *messa di voce* relates directly to sound pressure level (SPL) change, although Bretos and Sundberg observed that “a crescendo is generally realized by a linear increase of sound level with time” (2003, p. 351). Titze *et al.* (1999) found that SPL change in *messa di voce* with small SPL ranges tended to be linear and symmetrical, whereas those with large SPL ranges evidenced nonlinearity and asymmetry in SPL change, “usually with a linear or s-shaped SPL rise and an exponential-like SPL fall. The expected (model) shape was triangular, i.e., an increasing ramp followed by a mirrored decreasing ramp....Symmetry did not appear to be of primary importance when large SPL ranges were attempted by these singers” (p. 2937). Their published figures suggested linearity and symmetry of the SPL contour might be affected by F_0 , but this is uncertain because each subject

^{a)}Portions of this work were presented at the Sixth Australian Voice Symposium held in October 2002 in Adelaide, Australia

^{b)}Current address: P. O. Box 156, Box Hill, Victoria 3128, Australia; Electronic mail: sallycollyer@yahoo.com.au

^{c)}Current address: School of Communication Sciences, La Trobe University, Victoria 3086, Australia.

^{d)}Current address: Bioengineering Institute, University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand.

TABLE I. Details of participants' age, self-reported voice type, years of singing training, and professional taxonomy in accordance with Bunch and Chapman (2000).

	Age (years)	Self-reported voice type	Years of singing training	Taxonomy [refer to Bunch and Chapman (2000)]
Singer A	24	Lyric soprano	12	4.1b Regional/Touring: Minor Principal
Singer B	29	Lyric coloratura soprano	11	3.1c National/Big City: Chorus
Singer C	30	Lyric soprano	6	4.1b Regional/Touring: Minor Principal
Singer D	30	Lyric soprano	6	3.1c National/Big City: Chorus
Singer E	27	Lyric mezzo	11	2.1 International: Opera principal

performed at only three F_0 's. Given that professional opportunities for classical singers lie mostly in opera which requires a substantial dynamic range, and given that linearity and symmetry are considered fundamental to the well-executed *mesa di voce*, the suggestion that a wider dynamic range may be inimical to a smooth *mesa di voce* has major implications.

Their figures also suggested that SPL range in the decrescendo tended to exceed that in the crescendo, i.e., that SPL at the end was less than at the start of the *mesa di voce*. Schmidt *et al.* (1990) compared SPL ranges of crescendo and decrescendo while counting with trained and untrained subjects. SPL range averaged 2.1 dB greater in decrescendo, although this was a speaking, not a singing, task and they did not report their criteria for acceptable phonational quality. It is unclear whether a similar result would occur with well trained female singers in a singing task across a wide F_0 range.

The demand for smoothness and symmetry in timbral change has further implications in dynamic variation. An increase in SPL leads to a decrease in spectral balance (i.e., a greater rise in energy in higher than in lower partials) (Stout, 1938; Sundberg, 2001), due to the effects of rising subglottal pressure on the shape of the vocal-fold vibratory (source) wave form (Gauffin and Sundberg, 1989; Sundberg *et al.*, 1993) and to formant enhancement of this greater higher-partial energy (Titze and Sundberg, 1992). Sjölander and Sundberg (2004) found a substantially linear relationship between spectral balance (SB; there defined as the difference between the level of the first formant and the level of the singer's formant) and the natural log of subglottal pressure ($\text{Ln}(P_s)$), and in turn between $\text{Ln}(P_s)$ and SPL, suggesting a substantially linear inverse relationship between SPL and SB. Pedagogical demands regarding timbral change suggest that the change in the SB during the *mesa di voce* should be smooth and symmetrical, but this seems inconsistent with the lack of linearity and symmetry in SPL change observed by Titze *et al.* (1999). Sjölander and Sundberg also observed a F_0 influence on the relationship between $\text{Ln}(P_s)$ and SPL, suggesting that the relationship between SPL and SB may likewise be influenced by F_0 . Furthermore, Ternström, *et al.* (2006) found that the decrease of SB in speech saturated at around 6 dB below the speaker's maximum SPL. It is unclear what the complexity of the relationship between SPL

and SB and its potential dependence upon F_0 and SPL imply for SB change in the *mesa di voce*.

This study sought to determine to what extent linearity and symmetry are characteristic of the *mesa di voce* of well trained singers by addressing three questions. First, does the decrescendo show a markedly greater SPL range than the crescendo? Second, is the change in SPL during the *mesa di voce* linear and symmetrical and, if so, is this influenced by SPL range and F_0 ? Third, is the relationship between changing SB (the ratio of energy in higher to lower partials) and changing SPL also linear and symmetrical across the *mesa di voce*?

II. METHOD

A. Subjects

Five female classically trained singers, ranging in age from 24 to 30 years, participated in this study, and their profiles are set out in Table I. Three identified their voice type as lyric soprano, one as lyric coloratura soprano, and one as lyric mezzo. All held undergraduate degrees majoring in singing, and two held postgraduate diplomas in singing performance. Their years of singing tuition ranged from 6 to 12, and their professional experience ranged from 4.1b (Regional/Touring: Minor Principal) to 2.1 (International: Opera principal), using the classification system proposed by Bunch and Chapman (2000). This system aims to provide a basis for comparison of subject pools across different studies by classifying singers according to their professional status rather than by assessment of singers' vocal or performance capabilities, which are highly subjective.

B. Equipment and environment

All recordings were made in a sound-treated room 3.7 m wide \times 3.5 m long \times 2.7 m high at the Faculty of Health Sciences, The University of Sydney. The singers were recorded using a cardioid microphone (Pearl Mikrofonlaboratorium, model TL6C) mounted on a stand placed 30 cm from the singer's mouth. The singer was asked to position herself comfortably for singing. The front-most foot position was marked by tape on the floor and a polystyrene block mounted on a boom behind her was adjusted to touch the back of her head. The singer was asked to keep her feet

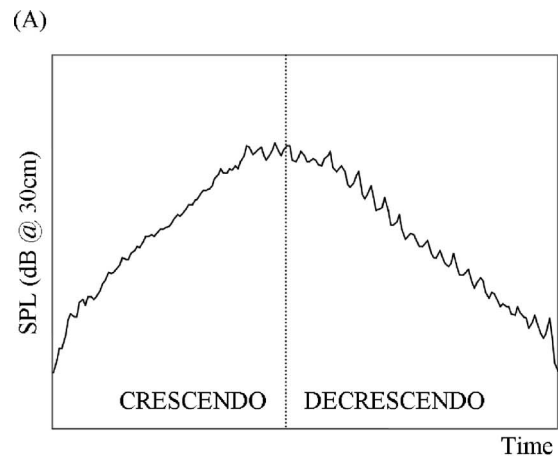
behind the tape mark and maintain contact with the polystyrene block, in order to maintain microphone distance. Periodic checks throughout the protocol confirmed that the microphone distance was well maintained.

For Singers A to D, the audio signal was recorded via a preamplifier (Rane MS1) onto a Tascam portable DAT recorder (model DA P1) at 44.1 kHz. The audio signal was monitored by an oscilloscope (Hewlett Packard 1741A) to confirm there was no clipping of the audio signal. Singer E's audio signal was recorded via a preamplifier (Behringer Ultragain Pro model MIC2200) onto a Marantz CD Recorder (model CDR 631) at 44.1 kHz. The DAT-recorded audio signal was transferred to computer through a LynXOne Mixer (Lynx Studio Technology) digital soundcard using COOLEEDIT 96 (Syntrillium) and loaded onto an IBM Aptiva 833 MHz. Acoustic data were analyzed using MATLAB (MathWorks), COOLEEDIT 96, and EXCEL (Microsoft). The signal was calibrated using a 1 kHz tone from a signal generator (Power Acoustik CP 500C) and a Rion NL 06 SPL meter set to slow damping with no weighting.

C. Protocol

The study was approved by the Human Ethics Committee of The University of Sydney. The entire protocol, including explanations of procedures, lasted between 2 and 3 h per singer. All singers had warmed up their voices before arriving and were given as much time as they desired to familiarize themselves with the acoustic of the recording room.

Messa di voce were obtained on /a/ throughout the singer's musically acceptable phonational range, beginning at A4 and descending by semitone (ST), then repeating A4 and rising by ST. Target pitches were provided by S.C. using a Yamaha Clavinova CLP-811. Two tokens per F_0 were required, acceptable to both the singer and the S.C. A high standard was set for acceptable phonational quality in order to ensure the results would be truly representative of the demands of the *messa di voce*. To maximize comparability of tokens, both inter- and intrasinger, it was important to standardize the duration of the *messa di voce*, while ensuring that such standardization was achieved in a manner natural to a musician and was achieved silently so as not to compromise the audio recording (Titze *et al.*, 1999). Accordingly, the S.C. conducted each *messa di voce* with a target duration of 8 s, being a 4-s crescendo and a 4-s decrescendo according to a digital clock not visible to the singer. Although great care was taken to ensure that the extent of left hand movement (indicating crescendo or decrescendo) was steady, consistent and maximal for all tokens, some variability in conducting may have occurred both during and between sessions. However, it was felt to be more important to elicit conditions as similar as possible to the musical training and performance situation. Using a mechanically based conducting mechanism might have alienated the singer by being impersonal and inflexible and failed to elicit the singer's full performance commitment and involvement. The singers also sang a vocalize which was based on the *messa di voce* (Concone No. 5). Those results are not analyzed here but are mentioned during discussion of SPL change.



(B)

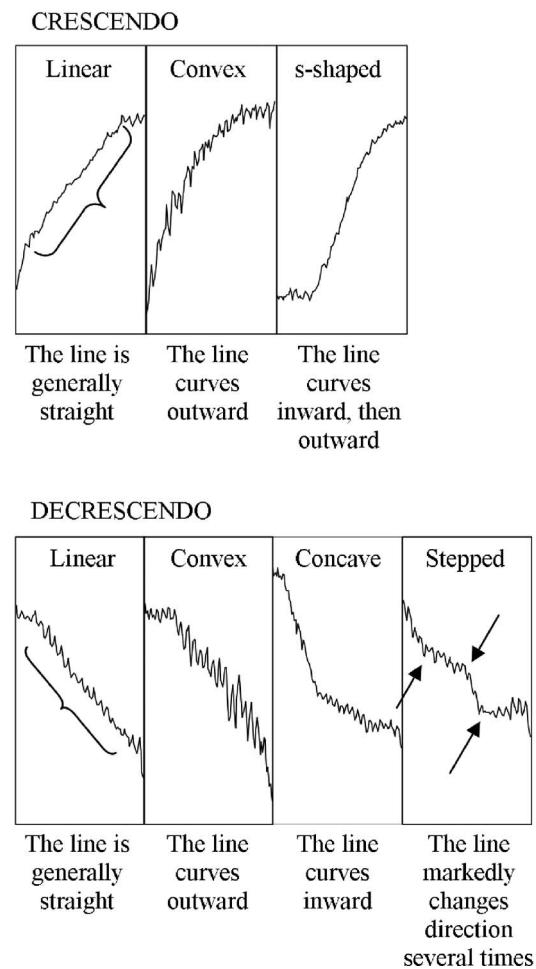


FIG. 1. Descriptions and examples of the shapes of the SPL (dB at 30 cm) trace during the crescendo and during the decrescendo. (a) An example of a SPL trace with a linear crescendo and linear decrescendo (Singer E's second C5 token). (b) Examples of the three categories of shape in crescendo and four in decrescendo. Percentage of tokens (from a total of 318 crescendi and 311 decrescendi) for each category were as follows: in crescendo, linear 15.1%, convex 61.0%, and s-shaped 23.9%; in decrescendo, linear 29.9%, convex 12.5%, concave 45.7%, and stepped 11.9%.

D. Data analysis

Data analysis began by isolating each *messa di voce* token from the audio signal. Using COOLEEDIT 96, the start and

end sample points of each token were manually identified from the audio signal. F_0 and calibrated SPL (rms amplitude) contours for each token were derived (at a sampling rate of 100 per s) using custom-written MATLAB programs.

To compare SPL range in crescendo and decrescendo, given that these are sequential within the one *messa di voce* and so share the same maximum SPL, the question was restated as comparing minimum SPL at the start of each token with that at the end of each token. In order to counteract the effects of amplitude modulation, which was apparent in many tokens, SPL was averaged over twenty samples (200 ms). Minimum SPL at the start and end of each token was calculated by averaging SPL over the first 20 samples and the last 20 samples, respectively. Maximum SPL was calculated as the average of 20 samples surrounding the point of maximum SPL.

Assessment of the linearity of SPL change during each token using quantitative methods proved unsatisfactory because the disparate shapes resulted in an unsatisfactorily large number of characteristic traces. Instead, categorization was constructed as a shape-recognition task by a panel of three judges. Preliminary examination of all traces suggested characteristics which differentiated three typical shapes in the crescendo (linear, convex, and s-shaped) and four in the decrescendo (linear, convex, concave, and stepped). Descriptions and examples were prepared, as set out in Fig. 1. Three independent judges assigned each crescendo (and each decrescendo) to one of the three (four) typical shapes based upon these exemplars. To identify F_0 and SPL influence on SPL trace shape, the tokens were divided approximately evenly into three F_0 ranges (Low-, Mid- and High- F_0) across SPL trace classification and analyzed using a chi-squared analysis at $p < 0.025$ ($p < 0.05$ adjusted to test crescendo and decrescendo).

To analyze SB change, the sound signal of each token was filtered using the COOLEEDIT 96 “FFT Filter” facility into separate low-pass (0–2 kHz) and high-pass (2–4 kHz) bands (settings: FFT size 4096, Blackman window). These two wave forms were then processed by the above-mentioned MATLAB program to obtain low-band and high-band SPL vectors, representing the SPL in the frequency bands 0–2 and 2–4 kHz, respectively (refer to Fig. 2). The low-band vector was subtracted from the high-band vector to obtain a vector tracing SB over the course of the *messa di voce*. Each SB trace was divided into crescendo and decrescendo sections at the point of maximum SPL, and a least-squares regression line for each section was plotted in EXCEL. Lines returning $R^2 < 0.7$ for either the crescendo or the decrescendo were deemed unrepresentative of the underlying tracing and were removed from further analysis. For each token, the SPL and SB traces across time and an x - y plot of SB change against SPL change were drawn and examined. Visual examination of SB change over time found these traces showed markedly less linearity and symmetry than did SPL change over time, and categorical analysis as described earlier for SPL change was not possible, although the traces did inform analysis of SPL change for Singers B and D.

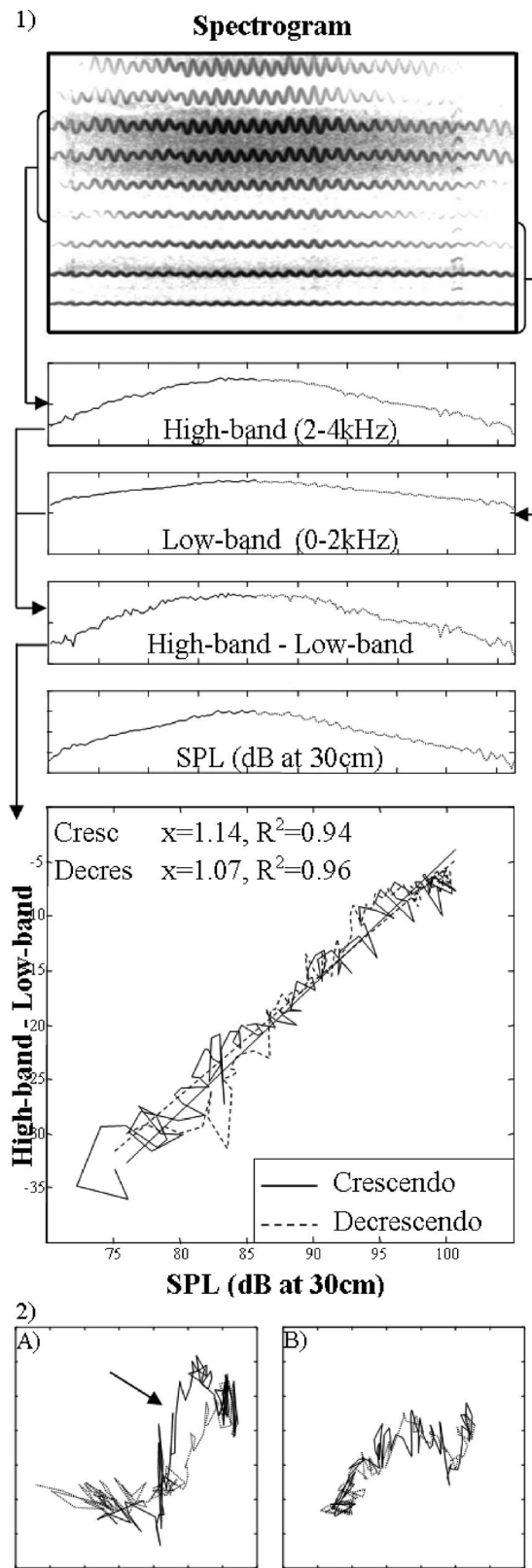


FIG. 2. (1) Demonstration of method for analyzing linearity and symmetry of spectral balance (SB) change with changing SPL (Singer E’s second C5 token) by plotting the difference between the high-band (2–4 kHz) and low-band (0–2 kHz) vectors as SPL changes during the *messa di voce*. The method is in Sec. II D. (2) Examples of nonlinearity in the relationship between SB change and changing SPL (Singer D). (a) The relationship during crescendo (solid line) and decrescendo (dotted line) is nonlinear, and the crescendo and decrescendo are asymmetrical, as indicated by the arrow. (B) The relationship during crescendo is nonlinear but is mirrored in the decrescendo and thus is symmetrical.

TABLE II. SPL was different at the start and at the end of each *messa di voce* token, but the percentage of tokens where the SPL was lower at the start (i.e., where SPL range was greater in crescendo than in decrescendo) varied between singers, as did the effect of $F0$ on the SPL difference at the start and at the end.

	Absolute difference in SPL at the start and at the end of the <i>messa di voce</i>		SPL lower at the start than at the end (i.e., greater SPL range in crescendo than in decrescendo)		
	Average (s.d.)	Correlation of difference with rising $F0$ (r)	% of tokens A4 and below	% of tokens A#4 and above	% of all tokens
Singer A	4.4 (2.9)	0.13	59	68	63
Singer B	7.6 (6.3)	0.69	46	100	75
Singer C	4.3 (3.0)	-0.11	21	70	44
Singer D	6.4 (5.1)	0.48	15	77	42
Singer E	6.3 (4.0)	-0.54	0	61	30

III. RESULTS

A total of 318 *messa di voce* tokens were recorded between E3 (165 Hz) and E6 (1319 Hz), with $F0$ range per singer averaging 30.8 ST [standard deviation (s.d.) 3.8 ST], from 27 ST for Singer B to 37 ST for Singer E.

A. Comparing SPL range in crescendo and in decrescendo

There were large differences in the SPL ranges of the crescendo and decrescendo in the same *messa di voce* (refer to Table II), with the average difference ranging from 4.3 dB (Singer C) to 7.6 dB (Singer B). The size of the difference showed no strong $F0$ influence. Nor was SPL range consistently larger in crescendo or in decrescendo: SPL range was greater in crescendo for 75% of tokens by Singer B but in only 30% of tokens by Singer E, with the remaining singers' results falling between these extremes. However, $F0$ did appear to influence whether SPL range was greater in crescendo or decrescendo. For all singers except Singer A, greater SPL range occurred markedly more often in crescendo at higher $F0$'s (A#4 and above) and in decrescendo at lower $F0$'s (A4 and below); Singer A's results were more evenly distributed but still followed this trend.

B. Linearity and symmetry of SPL change

The traces of SPL change were categorized by three independent judges according to the descriptions in Fig. 1. The three judges categorized 60.4% of crescendi unanimously,

and two of the three agreed on the remaining 39.6%. For the decrescendo, 53.5% were categorized unanimously and 44.3% were agreed by two judges, with no agreement on the remaining 2.2% (seven tokens). These seven tokens were excluded from subsequent analyses. Two factors accounted for the differing assessments in crescendo: the degree to which some variability in SPL increase was acceptable before the token could no longer be classified as linear, and whether a short delay in SPL rise at the start always constituted an s-shaped trace. Differing assessments in the decrescendo centered around the judge's distinction between human variability and the clear changes in direction specified by that categorization. In order to minimize influence, no further instruction than that provided by Fig. 1 was given to the judges, although it was recognized beforehand, and confirmed in unsolicited comments by each of the judges, that some traces were very easily classified whilst others were borderline.

Table III sets out the assessed shapes of the SPL traces of the crescendo and decrescendo, as described in Fig. 1. SPL traces during crescendo were assessed as linear in only 15.1% of all tokens, whilst convex traces predominated (61.0%). Individually, this was the case for all singers except Singer B, for whom more than half the crescendi (58.9%) were s-shaped. Whereas the salient characteristic of linearity was a steady increase in SPL, the SPL increase in the convex shape was greater at the start of the crescendo and gradually tapered off, whilst in the s-shaped trace the increase in SPL did not commence immediately but was delayed.

TABLE III. Summary of shapes of SPL traces, expressed as percentage of tokens by singer and of the total 318 tokens. The judges disagreed on seven decrescendi. Descriptions and examples of the shapes are illustrated in Fig. 1.

	Crescendo (% of tokens)			Decrescendo (% of tokens)				
	Linear	Convex	S-shaped	Linear	Convex	Concave	Stepped	Undecided
Singer A	17.7	53.2	29.0	39.7	5.2	36.2	19.0	6.5
Singer B	23.2	17.9	58.9	8.9	0.0	83.9	7.1	0.0
Singer C	6.3	81.3	12.5	36.5	55.6	3.2	4.8	1.6
Singer D	21.7	71.7	6.7	37.9	1.7	50.0	10.3	3.3
Singer E	9.2	73.7	17.1	26.3	0.0	56.6	17.1	0.0
Total	15.1	61.0	23.9	29.2	12.3	44.7	11.6	2.2

TABLE IV. Crescendo and decrescendo SPL trace shapes were divided into three F_0 groups: Low, Mid, and High. Chi-squared analysis showed a significant difference between groups in decrescendo ($p < 0.25$, adjusted for two tests).

	Percentage of tokens			Chi-squared analysis
	Low- F_0 (E4 and below)	Mid- F_0 (F4 to C#5)	High- F_0 (D5 to E6)	
Crescendo	[106 tokens]	[100 tokens]	[112 tokens]	
Linear	16	20	10	$\chi^2=9.74$
Convex	55	65	63	$p=0.0449$
S-shaped	29	15	27	
Decrescendo	[101 tokens]	[99 tokens]	[111 tokens]	
Linear	29	34	27	$\chi^2=22.38$
Convex	15	10	12	$p=0.0010$
Concave	34	45	57	
Stepped	22	11	4	

There was a greater amount of linearity in the decrescendo (29.2%), but nearly half the total decrescendi (44.7%) were described as concave, where the decrease in SPL was faster at the start and reduced toward the end. Individually, Singers B and E showed markedly more concave than linear shapes, whilst Singers A and D were more evenly balanced between the two shapes. Singer C also showed a marked degree of linearity (36.5%) but, alone of the five singers, her most common decrescendo shape was convex (55.6%), where the fall in SPL was slower at the beginning and became greater as the decrescendo progressed. Unsteadiness in the decrease in SPL was a notable feature in the decrescendo, with 11.6% of decrescendi showing a stepped pattern, where the decrease in SPL sped then slowed several times.

Assessment of symmetry in SPL change across the *mesa di voce* was affected by the difference between the SPL trace shapes apparent in crescendo and in decrescendo. For example, the concave and stepped decrescendo patterns had no counterparts in crescendo; all tokens with either shape in decrescendo were thus automatically asymmetrical. It was noted earlier that linear SPL change occurred in 15.1% of crescendi and 29.2% of decrescendi, but further analysis found that only 3.5% of tokens were linear in both crescendo and decrescendo and that these were shared across all singers. SPL traces were convex in both crescendo and decrescendo in 33 tokens (10.4%), of which 30 were by Singer C.

The SPL trace shapes were divided into three categories of Low- F_0 (E4 and below: 106 crescendi, 101 decrescendi), Mid- F_0 (F4 to C#5: 100, 99), and High- F_0 (D5 to E6: 112, 111), as set out in Table IV. Chi-squared analysis (adjusted to $p < 0.025$ for two tests) identified a significant difference between categories in decrescendo ($p=0.0010$) but not in crescendo ($p=0.0449$). Rising F_0 was associated with a marked decrease in the occurrence of stepped decrescendo and increase in concave decrescendo. Although not reaching significance, crescendo SPL traces also showed trends associated with F_0 , with s-shaped crescendi markedly more prevalent at Low- and High- F_0 than Mid- F_0 , and linearity in crescendo halving from Mid- F_0 to High- F_0 . However, comparison of average SPL range with SPL trace shape did not

TABLE V. Average SPL range for the SPL trace shapes in crescendo and decrescendo were comparable and showed no tendency for linear traces to be associated with smaller SPL ranges. (N/A=not applicable).

	Average (s.d.) SPL range (dB)	
	Crescendo Average (s.d.)	Decrescendo Average (s.d.)
Linear	23.0 (7.8)	22.8 (6.9)
Convex	21.8 (6.3)	19.6 (5.2)
S-shaped	22.3 (5.7)	N/A
Concave	N/A	20.9 (6.9)
Stepped	N/A	24.8 (5.5)

find that linear traces were associated with smaller SPL ranges in either crescendo or decrescendo (refer to Table V).

As noted in Sec. II D, change in SB over time was markedly less linear and symmetrical than was SPL change over time (see Fig. 3 for examples). Correlation of the SPL trace with the low-band trace showed that for all tokens except one, $R^2 \geq 0.98$, so that variation between the SPL and SB traces was due to variability in the high-pass band (2–4 kHz). However, visual comparison of the SB and SPL traces highlighted one interesting exception. For Singer B (B4 to A#5 inclusive) and Singer D (D#5 to A5 inclusive), the decrescendo SPL trace was concave and SPL was markedly higher at the end than at the start of the *mesa di voce* but the SB trace was substantially symmetrical; two examples are shown in Fig. 3 (D5 and F5 for Singer B).

C. Linearity and symmetry of the relationship between spectral balance and sound pressure level change

Markedly more crescendi (67.0%) than decrescendi (53.5%) showed a substantially linear relationship ($R^2 \geq 0.7$) between changing SB and changing SPL (refer to Table VI). Of the 318 tokens, 47.5% were substantially linear in both the crescendo and the decrescendo. Examination of SB plots for the remaining tokens, especially those for Singer D whose linearity was the least of all the singers, showed two forms of nonlinearity [refer to Figs. 2, 2(a), and

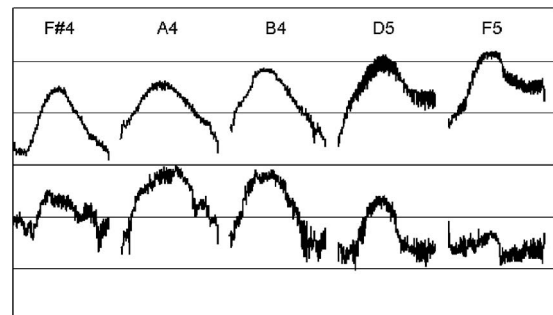


FIG. 3. Examples of SPL (upper trace) and SB (lower trace) change across individual *mesa di voce*. SB change across time showed markedly less linearity and symmetry than did SPL change, with the exception of certain higher F_0 's for Singers B and D, where SPL was markedly higher at the end than at the start but SB was effectively the same. Examples shown are Singer B F#4 (second token), Singer C A4 (descending, second token), Singer E B4 (second token), and Singer B D5 (first token) and F5 (first token).

TABLE VI. Linearity and symmetry of change in spectral balance (SB) with respect to changing SPL. Percentage of 318 tokens in which SB change was substantially linear ($R^2 \geq 0.7$) with respect to SPL change in crescendo, in decrescendo and in both, and percentage in which SB change with respect to SPL change in crescendo was mirrored in decrescendo (irrespective of linearity). Examples are given in Figs. 2(a) and 2(b).

	Linear change in spectral balance (% of 318 tokens)			Visual assessment of symmetry (% of 318 tokens)
	Crescendo	Decrescendo	Both	
Singer A	62.9	38.7	38.7	64.5
Singer B	55.4	39.3	33.9	82.1
Singer C	85.9	67.2	65.6	87.5
Singer D	43.3	33.3	13.3	68.3
Singer E	81.6	80.3	76.3	92.1
Total	67.0	53.5	47.5	79.6

2(b)]. In the first, the relationship between SB change and SPL change during the crescendo differed from that during the decrescendo. In the second, however, the relationship during the crescendo was mirrored in the decrescendo, so that the relationship was symmetrical but nonlinear. Visual assessment of all tokens found 79.6% were symmetrical.

There was some influence of F_0 , with linearity in the SB:SPL relationship far more likely at lower than higher F_0 's, although there was strong intersinger variability. Singers generally achieved linearity in crescendo at higher F_0 's than in decrescendo, leading to the markedly greater number of linear crescendi than decrescendi noted earlier. Singer E's results were unique, with consistent linearity in crescendo and decrescendo (ceasing abruptly above A#5).

IV. DISCUSSION

This study posed three questions regarding linearity and symmetry in the *mesa di voce* as performed by female classical singers. The first question was, does the decrescendo show a markedly greater SPL range than the crescendo. Although there were clear differences in SPL range by token, neither crescendo nor decrescendo consistently prevailed in this study. All but Singer A showed a marked influence of F_0 , where greater SPL range occurred more often in crescendo at higher F_0 's and in decrescendo at lower F_0 's. It might be that these differences reflect a shortcoming in the singer's skill, or it might be that the perception of musicality means that a balanced *mesa di voce* requires an asymmetry which changes with rising F_0 . Although no perceptual studies of the *mesa di voce* could be located, increases in SPL have been shown to be more salient than decreases, generally (Neuhoff, 1998) and in music (Huron, 1990). The degree to which such auditory asymmetry influences both the singer and the listener in the *mesa di voce* awaits perceptual study.

The second question was, is the change in SPL during the *mesa di voce* linear and symmetrical and, if so, is this influenced by SPL range and F_0 . This question assumed that a perceptually smooth transition in the crescendo and decrescendo equated with a linear change in SPL. Only 15.1% of crescendi were linear, compared with 61.0% convex; only 29.2% of decrescendi were linear, whereas 44.7% were con-

cave. In order to standardize token duration, the singers were conducted, and this may have induced nonlinearity. However, nonlinearity also occurred when the singers performed a vocalize (not reported here), where the singers were not conducted and were free to pace SPL change within the tempo constraints of piano accompaniment. Nonlinear shapes also predominated in the *mesa di voce* reported by Titze *et al.* (1999), whose subjects sang uncondacted. Nonlinearity of SPL change might reflect a shortcoming in the singer's skill and the inherent difficulty of the exercise, or linearity might be too simplistic an acoustical correlate of a perceptually smooth SPL change in an emotionally engaging *mesa di voce*. Perhaps listeners perceive an initially faster SPL rise and fall, such as in the convex crescendo and concave decrescendo, as musically symmetrical. Mathews (1979) varied the average intensity of crescendi with equal dB change and found that listeners judged crescendi with greater average intensity to be stronger. From analysis of musical scores, Huron (1991) has suggested a "ramp archetype" in which musical dynamic builds gradually but subsides relatively quickly. Perhaps singers perceive these patterns, aurally and kinesthetically, as musical or linear or both. These issues await investigation of the relationship between perceptually smooth SPL change and linearity in the SPL trace.

Whereas these results confirm the findings of Titze *et al.* (1999) that linearity in SPL change is not a predominant feature of the *mesa di voce*, there are certain differences. Titze *et al.* found their subjects' SPL traces to have "a linear or s-shaped SPL rise and an exponential like SPL fall" (p. 2937), the fall equating to the concave decrescendo shape in this study. Whilst linear and s-shaped crescendi in this study accounted for 39.0% of tokens, the convex shape which was determined in 61.0% of crescendi in this study was not a shape reported by Titze *et al.* There are several possible reasons for this. The much larger number of tokens in this study might have highlighted a distinction between linear and convex crescendi not so apparent in the smaller sample, particularly as they did note a number of "flat top" traces, where SPL peaked and remained constant for some time. The distinction might be influenced by the duration of the *mesa di voce* and its components, which was not reported by Titze *et al.* It is possible their subjects simply evidenced more linearity in crescendo, just as the singers in this study differed in their shape preferences.

It is interesting to note that nearly one-eighth of the SPL traces in decrescendo showed marked unevenness (stepped shape) not evident in the crescendo. This seems to support the general view of singing pedagogy that the decrescendo is more difficult than the crescendo. On the other hand, the SPL traces of around twice as many decrescendi as crescendi were classified as linear, and the renowned greater difficulty of the decrescendo did not lead to a consistently smaller SPL range than in the crescendo.

Combining the results relating to SPL range and to SPL trace shape with respect to F_0 raises interesting points for crescendo and for decrescendo. Although it did not achieve statistical significance, the s-shaped trace in the crescendo

occurred almost twice as often at Low- F_0 and High- F_0 than at Mid- F_0 , suggesting that the singers needed time to establish a quality phonation at *pianissimo* (*pp*) when outside the middle region of the musical F_0 range. Such care does not appear to be directly associated with the level of phonation threshold pressure or with SPL at *pp*, since these generally rise with F_0 rather than being greater in the middle range. In the decrescendo, higher F_0 was associated with smaller SPL range than in the crescendo and with increasing occurrence of concave SPL traces. This occurred consistently in the range B4 to A#5 for Singer B and D#5 to A5 for Singer D. It is possible that the singers reached a minimum SPL in decrescendo and stayed at that SPL for the remainder of the *mesa di voce*, unable to reduce phonation threshold pressure any further. However, it is interesting that these tokens also showed the only symmetry observed in SB change across time, and it may be that they were guided by a timbral aesthetic in this F_0 region. Subglottal pressure was not estimated in this study. However, if we assume that greater SPL range in decrescendo than in crescendo reflected a lower phonation threshold pressure at offset than onset, then the results at lower F_0 's were generally consistent with the model of vocal-fold behavior described by Lucero (1995, 1999, 2005) using the body-cover model of vocal-fold vibration described by Titze (1988). However, this was not the case at higher F_0 's. The increased occurrence of the concave SPL trace indicates that singers may have targeted a lower *pp* SPL at the end of the *mesa di voce* but were unable to achieve it. It is unclear why phonation threshold pressure at higher F_0 's should appear to be generally lower at onset than at offset. Plant *et al.* (2004) did find marked variability in threshold pressure hysteresis, and the tendency for F_0 to influence SPL range in crescendo or decrescendo was not as strong with Singer A as it was with the other singers.

The third question was, is the relationship between changing SB (the ratio of energy in higher to lower partials) and changing SPL also linear and symmetrical across the *mesa di voce*. Only 47.5% of tokens were linear in both crescendo and decrescendo, yet 79.6% of tokens were symmetrical. Symmetrical but nonlinear change suggests not only that the mode of phonation changed but that this change may have been an integral part either of technique or of habit in that it occurred across the same SPL range in both the crescendo and decrescendo without having any observable effect on the SPL trace itself. One scenario for this example might be that phonation tended toward being pressed at the start, shifted toward breathy phonation within a particular SPL range, and shifted back toward pressed above this range, although such a scenario would seem to be antithetical to the pedagogical ideal of consistency of timbre in the *mesa di voce*. The acoustical record cannot support more than speculation, but irregularities in the changes in both subglottal pressure change and maximum flow declination rate during the *mesa di voce* were observed by Titze *et al.* (1999). Subglottal pressure data in particular would clarify this situation, although, as noted by Titze *et al.*, the */p/*-occlusion estimation method is itself inconsistent with the essence of the *mesa di voce*. Perceptual studies are needed to clarify whether symmetry is more important than linearity in the

relationship between SB and SPL and to clarify the apparent targeting of SB rather than SPL symmetry by Singers B and D in higher F_0 's.

Although the relationship between SB and SPL may have been influenced by room resonances affecting the perceived loudness and SB at particular pitches, plots of SB at *pp* and *ff* across each singer's full F_0 range identified no consistent patterns which indicated room resonance effects. Whereas an anechoic recording environment would have been preferable, a sound-treated room was considered an acceptable compromise to a musically familiar environment, and any effects were made consistent for each singer by standardizing the location of the microphone and the singer in the room. Although it is also possible that the order of tasks (beginning at A4 and descending, then returning to A4 and ascending) influenced the results, singers are more comfortable building up to higher notes (Elliot *et al.*, 1995; Motel *et al.*, 2003) and it is generally accepted in classical singing pedagogy that exercises should begin with the middle and lower registers (Miller, 1996; Vennard, 1967). Given that we were recording the full musical F_0 range of the singers, we did not envisage any advantage in randomizing F_0 .

V. CONCLUSION

Measuring the *mesa di voce* on every note across each singer's F_0 range provided a comprehensive record of each singer's capability and permitted a detailed investigation of the acoustical characteristics of her performance and the influence upon them of F_0 and of SPL range. The findings generally support the conclusions of Titze *et al.* (1999) with respect to nonlinearity of SPL change, although the shapes observed differed somewhat, which might be due to the much larger sample size of tokens and greater similarity of the participants in this study. The suggested relationship between linearity of SPL change and SPL range, however, was not observed here. Differences in SPL range in crescendo and decrescendo were influenced by F_0 , as was the SPL trace shape in decrescendo. Analysis of the relationship between SB change and SPL change found a surprisingly high proportion of symmetry, markedly greater than the proportion of tokens in which the relationship was linear. However, symmetry in SPL change was complicated by differing patterns of SPL change in crescendo and in decrescendo. The findings also identified areas in which perceptual and physiological studies would clarify the nature and desirability of nonlinearity, particularly in the shape of the SPL trace and in the relationship between SB and SPL change. Singing pedagogy would benefit greatly from clarification of these areas.

Bretos, J., and Sundberg, J. (2003). "Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos," *J. Voice* **17**, 343–352.

Bunch, M., and Chapman, J. (2000). "Taxonomy of singers used as subjects in scientific research," *J. Voice* **14**, 363–369.

Burgin, J. C. (1973). *Teaching Singing* (Scarecrow, Metuchen, NJ).

Elliot, N., Sundberg, J., and Gramming, P. (1995). "What happens during vocal warm-up," *J. Voice* **9**, 37–44.

Fleming, R. (2004). *The Inner Voice: The Making of a Singer* (Viking, New York).

Garcia, M. II. (1982). *A Complete Treatise on the Art of Singing, in Two*

- Parts*, translated and edited by D. V. Paschke (Da Capo, New York) (original work published 1841 and 1872).
- Gauffin, J., and Sundberg, J. (1989). "Spectral correlates of glottal voice source waveform characteristics," *J. Speech Hear. Res.* **32**, 556–565.
- Huron, D. (1990). "Crescendo/diminuendo asymmetries in Beethoven's piano sonatas," *Music Percept.* **7**, 395–402.
- Huron, D. (1991). "The ramp archetype: A score-based study of musical dynamics in 14 piano composers," *Psych. Mus.* **19**, 33–45.
- Lucero, J. C. (1995). "The minimum lung pressure to sustain vocal fold oscillation," *J. Acoust. Soc. Am.* **98**, 779–784.
- Lucero, J. C. (1999). "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," *J. Acoust. Soc. Am.* **105**, 423–431.
- Lucero, J. C. (2005). "Oscillation hysteresis in a two-mass model of the vocal folds," *J. Sound Vib.* **282**, 1247–1254.
- Mathews, M. V. (1979). "Perception of crescendos as a function of duration," *J. Acoust. Soc. Am.* **65**, S123.
- Miller, R. (1996). *The Structure of Singing: System and Art in Vocal Technique* (Schirmer, New York).
- Miller, R. (2000). *Training Soprano Voices* (Oxford University Press, Oxford).
- Motel, T., Fisher, K. V., and Leydon, C. (2003). "Vocal warm-up increases phonation threshold pressure in soprano singers at high pitch," *J. Voice* **17**, 160–167.
- Neuhoff, J. G. (1998). "Perceptual bias for rising tones," *Nature (London)* **395**, 123.
- Plant, R., Freed, G. L., and Plant, R. E. (2004). "Direct measurement of onset and offset phonation threshold pressure in normal subjects," *J. Acoust. Soc. Am.* **116**, 3640–3646.
- Reid, C. L. (1965). *The Free Voice: A Guide to Natural Singing* (Patelson, New York).
- Schmidt, C. P., Gelfer, M. P., and Andrews, M. L. (1990). "Intensity range as a function of task and training," *J. Voice* **4**, 30–36.
- Sjölander P. and Sundberg J. (2004). "Spectrum effects of subglottal pressure variation in professional baritone singers," *J. Acoust. Soc. Am.* **115**, 1270–1273.
- Stout, B. (1938). "The harmonic structure of vowels in singing in relation to pitch and intensity," *J. Acoust. Soc. Am.* **10**, 137–146.
- Sundberg, J. (2001). "Level and center frequency of the singer's formant," *J. Voice* **15**, 176–186.
- Sundberg, J., Titze, I., and Scherer, R. (1993). "Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source," *J. Voice* **7**, 15–29.
- Ternström, S., Pohman, M., and Södersten, M. (2005). "Loud speech over noise: Some spectral attributes with gender differences," *J. Acoust. Soc. Am.* **120**, 1648–1665.
- Titze, I. R. (1988). "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.* **83**, 1536–1552.
- Titze, I. R. (1989). "On the relation between subglottal pressure and fundamental frequency in phonation," *J. Acoust. Soc. Am.* **85**, 901–906.
- Titze, I. R. (1992). "Messa di voce," *NATS J.* **48**(3), 24.
- Titze, I. R., Long, R., Shirley, G. I., Stathopoulos, E., Ramig, L. O., Carroll, L. M., and Riley, W. D. (1999). "Messa di voce: An investigation of the symmetry of crescendo and decrescendo in a singing exercise," *J. Acoust. Soc. Am.* **105**, 2933–2940.
- Titze, I. R., and Sundberg, J. (1992). "Vocal intensity in speakers and singers," *J. Acoust. Soc. Am.* **91**, 2936–2946.
- Tosi, P. F. (1987). *Observations on the Florid Song*, translated by J. E. Galliard, edited with additional notes by M. Pilkington (Stainer and Bell, London) (original work undated).
- Vennard, W. (1967). *Singing: The Mechanism and the Technic* (Fisher, New York).

Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California

Marie A. Roch^{a)}

Department of Computer Science, San Diego State University, 5500 Campanile Drive,
San Diego, California 92182-7720

Melissa S. Soldevilla, Jessica C. Burtenshaw,

E. Elizabeth Henderson, and John A. Hildebrand

Scripps Institution of Oceanography, The University of California at San Diego,
La Jolla, California 92093-0205

(Received 12 May 2006; revised 16 October 2006; accepted 24 October 2006)

A method for the automatic classification of free-ranging delphinid vocalizations is presented. The vocalizations of short-beaked and long-beaked common (*Delphinus delphis* and *Delphinus capensis*), Pacific white-sided (*Lagenorhynchus obliquidens*), and bottlenose (*Tursiops truncatus*) dolphins were recorded in a pelagic environment of the Southern California Bight and the Gulf of California over a period of 4 years. Cepstral feature vectors are extracted from call data which contain simultaneous overlapping whistles, burst-pulses, and clicks from a single species. These features are grouped into multisecond segments. A portion of the data is used to train Gaussian mixture models of varying orders for each species. The remaining call data are used to test the performance of the models. Species are predicted based upon probabilistic measures of model similarity with test segment groups having durations between 1 and 25 s. For this data set, 256 mixture Gaussian mixture models and segments of at least 10 s of call data resulted in the best classification results. The classifier predicts the species of groups with 67%–75% accuracy depending upon the partitioning of the training and test data. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2400663]

PACS number(s): 43.80.Ev, 43.60.Uv [WWA]

Pages: 1737–1748

I. INTRODUCTION

Long-term acoustic monitoring is an established technique for assessing cetacean relative abundance and seasonality (Thompson and Friedl, 1982). Key steps in processing acoustic monitoring data are acoustic call detection and species classification. Identification of stereotyped mystecete calls has been accomplished using automatic detectors (e.g., Sirovic *et al.*, 2004), but odontocete call identification is more difficult owing to their calls' greater complexity. Species identification for odontocete calls has been accomplished using trained analysts, as well as automated classification based on similarity to calls collected in the presence of known species (Oswald *et al.*, 2003).

Recent advances in acoustic recording capabilities allow remote autonomous recordings with terabyte data storage (Wiggins, 2003). Manual analyses of these large data sets are prohibitively expensive. Reliable automated methods are needed for detection and classification of odontocete calls to allow rapid analysis of these large acoustic data sets.

Unlike many mammals (Fenton and Bell, 1981; Goold and Jones, 1995; Thompson *et al.*, 1992, 1996) and birds (Marler, 1957) which exhibit stereotyped calls that are readily distinguishable by species, delphinids have a wide and varied vocal repertoire that makes species identification more complex (Oswald *et al.*, 2003; Thompson and Richard-

son, 1995). Dolphin calls can be broken down into three general categories: echolocation clicks, burst-pulsed calls, and whistles (Popper, 1980). Each of these call types exhibits complex time- and frequency-varying features that differ across species and individuals. Echolocation clicks are broadband, impulsive sounds which typically range between 10 and 150 kHz in many dolphin species (Au, 1993). Echolocation clicks are used for prey-finding and navigation. Burst-pulsed calls are rapid series of broadband clicks which are not individually distinguishable to humans, resulting in calls with a scream-like, tonal quality (Murray *et al.*, 1988). These calls can range from 5 to 150 kHz and are thought to function for communicative purposes. Whistles are frequency modulated narrowband tonal calls which occur between 2 and 35 kHz (Thompson and Richardson, 1995). Whistles are thought to have communicative functions and it has been suggested that they may carry individual-specific information in some species (Caldwell *et al.*, 1990). While all dolphin species recorded to date produce click type calls, some species may not produce whistles (Herman and Tavolga, 1980).

Automatic classification of marine mammal calls involves at least three steps: signal detection, feature extraction, and classification. During signal detection, calls of interest are located within the larger time series. Feature extraction transforms each call to a feature vector or set of feature vectors which represents the salient characteristics of the call. Finally, the feature vectors are classified as belong-

^{a)}Electronic mail: marie.roch@sdsu.edu

ing to one of the target classes, or possibly as an unknown class. Target classes may be specific call types, individual animals, or species. The level of automation of each of these steps varies greatly in previously described cetacean classification studies.

Until recently, the standard approach has been to manually locate the end points of tonal calls and measure features such as fundamental frequency, harmonics, slope, and inflection points (e.g., Rendel *et al.*, 1999). More recent work, such as that of Oswald *et al.* (2005) can extract an expanded list of similar information automatically when given the start and end of a call. Datta and Sturtivant (2002) used edge detection techniques from computer vision to locate the whistle segments in a spectrogram.

Cepstral processing is a useful feature extraction technique used in human speech analysis. The real cepstrum is the discrete cosine transform (DCT) of the log of the short time spectral magnitude (Picone, 1993). When the source-filter model (Harrington and Cassidy, 1999) is assumed for the production of calls, this transformation results in the source information being typically contained in higher orders of the cepstrum. These can be discarded, resulting in a feature vector which captures information about the filter. While this will result in the loss of information about the source, it typically leads to reductions in both the amount of data needed to train effective models and in the computational time needed to train and use the classifier. Fitch (2000) notes that the source-filter model appears to be applicable for all mammals whose sound production has been studied. While odontocetes have different sound production systems from other mammals, the source-filter model is still a relevant paradigm (Cranford, 2000).

To further reduce the size of the feature vector, it is common to apply a set of filter banks that are spaced linearly at low frequencies and logarithmically at higher frequencies before computing the DCT. In bioacoustic studies of elephant calls and bird song (Clemins *et al.*, 2005; Kogan and Margoliash, 1998), a spacing based upon human psychophysics studies (the Mel scale, Sundberg, 1991) has been proposed. Due to differences in hearing, we believe that a more neutral approach or an approach specific to individual species such as the extensions to Hermansky's perceptual linear prediction (1990) proposed by Clemins and Johnson (2006) are appropriate. However, when working with multiple species, care must be taken to either design an aggregate filterbank or to perform the feature extraction for each species.

The classification of extracted feature vectors is accomplished using machine learning techniques. The majority of researchers have used algorithms that require supervised classification. These algorithms learn a partitioning of the feature space based upon training vectors. Once the classifier has been trained, feature vectors are assigned to classes based upon tests which determine to which partition the feature vectors belong.

One of the simplest supervised classifiers is linear discriminant analysis (Duda *et al.*, 2001), a technique which finds the hyperplane that best separates pairs of classes in a

set of labeled training data. Steiner (1981) used linear discriminant analysis to differentiate the whistles of five dolphin species.

Other techniques use combinations of hyperplanes. Classification and regression trees (Duda *et al.*, 2001) is a related technique where multiple hyperplanes hierarchically partition the feature space into hypercubes. This technique has been used by Oswald *et al.* (2004) to determine the species of wild dolphin calls. Neural networks have been used by numerous groups. They are capable of separating the feature space into complex subregions associated with specific classes. A common method of integrating temporal data is to take a set of N evenly spaced samples from the feature data and to assemble them into a higher-dimensional feature vector. This approach has been used with backpropagation neural networks for the tasks of differentiating killer whale dialects, bottlenose dolphin clicks, and the detection of bowhead whale song notes (Deecke *et al.*, 1999; Houser *et al.*, 1999; Potter *et al.*, 1994). Alternative strategies to concatenating vectors are possible, such as the spectral averaging used by Potter *et al.* (1994).

When the goal is to recognize a specific call, the situation is complicated in that different repetitions of the call may be produced at different rates (Buck and Tyack, 1993). The rates of different portions of the call can vary considerably, and linear scaling is unlikely to capture the variation appropriately. One technique to cope with this is the use of dynamic time warping (DTW) (Rabiner and Juang, 1993), a dynamic programming technique that aligns the call to be classified to a reference call. This has been used to recognize whistles from a small set of captive dolphins, calls in captive bird song, and free ranging bowhead whales' calls (Buck and Tyack, 1993; Kogan and Margoliash, 1998; Mellinger and Clark, 2000). An alternate strategy is the use of hidden Markov models (HMMs) (Rabiner and Juang, 1993). HMMs are also capable of nonlinear time alignment, and have been shown both in the speech and bioacoustic communities to be effective classifiers. They have been used for distinguishing individual African elephants and their call types, dolphin group identities, bird individuals, and bowhead whale calls (Clemins *et al.*, 2005; Datta and Sturtivant, 2002; Kogan and Margoliash, 1998; Mellinger and Clark, 2000). In general, these provide a more robust performance than DTW, but HMMs require more data to estimate the model parameters.

A few research groups have studied unsupervised classifiers. Unsupervised classifiers attempt to learn classes from unlabeled data sets. Murray *et al.* (1988) used Kohonen's self-organizing maps and competitive learning to discern classes from a false killer whale call data set. With either technique, the goal is to have the network learn the similarities and differences between the feature vectors. Both methods were successful in learning a number of statistically homogeneous classes from calls produced by two individuals, and the authors were able to make links between the automatically discovered categories and those commonly given by humans such as clicks and whistles. Recently, Deecke and Janik (2006) combined adaptive resonance theory (ART) with DTW. In ART networks, a new pattern is compared to models for existing ones. If the new pattern differs suffi-

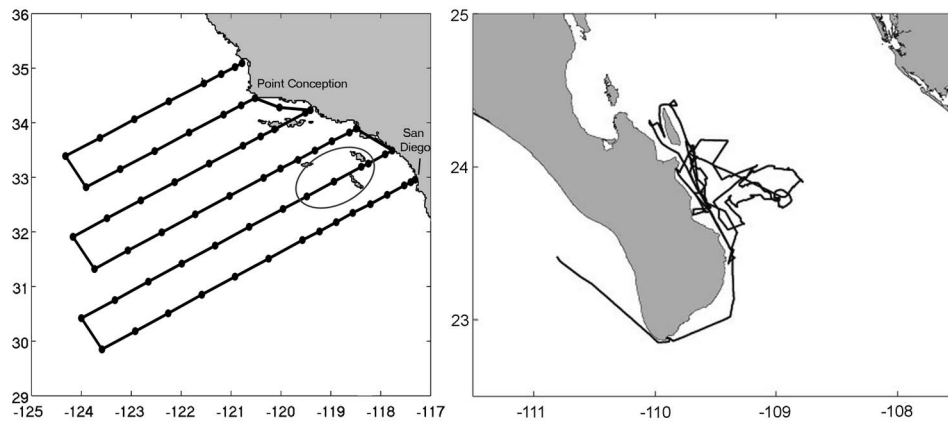


FIG. 1. Acoustic recordings were made along the transect and track lines in (a) the Southern California Bight and (b) the Gulf of California. The southern California data were collected on two series of cruises, one of which concentrated on the area contained in the square region.

ciently as determined by an experimenter controlled threshold, a new pattern is formed. DTW is used to provide the difference measure between existing patterns in the ART network and new ones.

Our study examines the performance of Gaussian mixture model (GMM) classifiers (Huang *et al.*, 2001) for determining the species of groups of free ranging dolphins recorded in the offshore Southern California Bight and the Gulf of California. GMMs are well known for their ability to model arbitrarily complex distributions with multiple modes and are effective classifiers for many tasks. They are functionally equivalent to ergodic hidden Markov models, and are appropriate for the species identification task where there is no expectation as to what component of a call will follow the current one and multiple calls may occur simultaneously. The classifier uses cepstral feature vectors and is able to distinguish the studied species accurately in 67%–75% of the test cases.

II. METHODS

The methods of this study are organized by task. We first describe the collection of the call data and its characteristics. The processing of the data is separated into call detection, feature extraction, and classification.

A. Data collection and species call descriptions

Acoustic recordings were collected on multiple cruises offshore of Southern California and within the Gulf of California between 2001 and 2005 (Fig. 1). Standard line-transect surveys were conducted to visually identify cetaceans in the study area. When a single-species school was encountered, an SSQ-57B sonobuoy was deployed, and the ship was positioned 1 to 2 km away from the school. SSQ-57B sonobuoys have a flat frequency response from 20 Hz to 20 kHz. The sonobuoy signal was transmitted to a multi-channel receiver located on the ship. Acoustic data sampled at 48 kHz were recorded either directly to hard drive or to one of the following Sony DAT recorders: PCM-M1, TCD-D7, or TCD-D8. To improve the likelihood that recordings contain only single species call-types, we only analyzed recordings obtained when no other species' schools were

sighted within 5 km of the sonobuoy location, and only included calls with high signal-to-noise ratio (SNR).

Four dolphin species that are commonly found and recorded in this region include short-beaked common (*Delphinus delphis*), long-beaked common (*Delphinus capensis*), Pacific white-sided (*Lagenorhynchus obliquidens*), and bottlenose (*Tursiops truncatus*) dolphins. As sighting logs for our early recordings did not distinguish the two species of common dolphins, the automatic classification system uses the *genus* for these animals. Bottlenose dolphins are the least abundant of the delphinids in our study area. When they are sighted, they are frequently in mixed groups with Risso's dolphins (*Grampus griseus*). Much of our bottlenose dolphin data are from the Gulf of California, where the same collection procedures were used with the exception of line-transect surveys. Other species known to inhabit the Southern California Bight, Risso's and northern right whale (*Lissodelphis borealis*) dolphins, were not encountered frequently enough to be included in the analysis.

Common dolphins produce whistles, burst pulses, and echolocation click trains (Au, 1993; Caldwell and Caldwell, 1968; Moore and Ridgway, 1995). Their whistles have a mean duration of 0.8 s, a mean minimum frequency of 7.4 kHz, a mean maximum frequency of 13.6 kHz, and a mean of 1.2 inflection points (Oswald *et al.*, 2003). Common dolphin clicks have source levels of 160–170 dB re 1 μ Pa at 1 m, pulse durations between 50 and 250 μ s, and peak frequencies between 23 and 67 kHz (Au, 1993; Evans, 1973; Fish and Turl, 1975). Whistles made up the majority of calls we recorded (99%), with many of them overlapping, whereas burst pulses (1%) and the lower portion of their clicks were present at low numbers (<1%). For all the species we recorded, percentages of click trains may be low as overlapping click trains were not distinguished.

Free-ranging bottlenose dolphins produce all three call types, with individual whistle characteristics including durations between 0.6 and 1.4 s, minimum frequencies between 5.4 and 8.5 kHz, maximum frequencies between 11.32 and 17.2 kHz, and 1.86 and 3.7 inflection points for a variety of populations (Acevedo-Gutiérrez and Stienessen, 2004; Oswald *et al.*, 2003; Steiner, 1981; Wang *et al.*, 1995). Bottlenose echolocation clicks have source levels of 228 dB re

TABLE I. Number of seconds of usable call data obtained for each dolphin species by date.

Recording session	Common		Pacific white-sided		Bottlenose	
	Date	s	Date	s	Date	s
1	30 April 2001	526	30 April 2001	398	15 April 2002	242
2	1 May 2001	159	20 June 2001	59	6 March 2004	350
3	4 November 2003	330	20 August 2003	401	8 March 2004	240
4					10 March 2004	409
5					15 May 2005	363
6					17 May 2005	488
7					18 May 2005	1133
8					19 May 2005	264
9					21 May 2005	349
Total		1015		858		3838

1 μPa at 1 m, pulse durations between 50 and 80 μs and peak frequencies between 110 and 130 kHz, though these may vary with location (Au, 1993). Our bottlenose dolphin recordings contained 81% whistles, 8% click trains, and 11% burst pulses.

Pacific white-sided dolphin echolocation clicks have been recorded with source levels of 170 dB re 1 μPa , pulse durations between 25 and 1000 μs , and peak frequencies between 50 and 80 kHz and 100 and 120 kHz (Evans, 1973; Fahner *et al.*, 2004; Nakamura and Akamatsu, 2004). While whistles have been recorded from Pacific white-sided dolphins (Caldwell and Caldwell, 1971; Whitten and Thomas, 2001), few were recorded during our sessions (21%) when compared to other species. The majority of Pacific white-sided calls we recorded were burst pulses (70%) and the lower frequency portion of click trains (10%).

B. Call detection

The detection of calls was accomplished manually. The start and end point of sets of whistles, burst pulses, and clicks were identified using spectrograms and audition when possible. Only calls which were deemed to be of sufficient quality as judged by SNR across the call bandwidth were used. Typically, these calls had SNRs of greater than 18 dB,

and comprised approximately 65% of all detected calls. No effort was made to denote the start or end of individual calls, or to segregate individual calls from those that occurred with other conspecifics. Table I summarizes the call data used in this study by recording date.

Sessions 2–9 of the bottlenose call data were recorded in the Gulf of California; all other recordings were made in the Southern California Bight.

C. Feature extraction

Cepstral feature vectors were used to represent the short time spectrum of odontocete call data. No attempt was made to isolate individual calls, and the classifier learned the collection of sounds produced by groups of dolphins.

As shown in Fig. 2, the process consisted of computing the squared magnitude frequency response of a 21 ms frame which had been windowed with a Hamming window. A filter bank consisting of 64 linearly spaced overlapping triangular filters was applied between 5 and 23.5 kHz. The lower edge of 5 kHz was selected as the SNR for calls tended to be poor at frequencies beneath this threshold. The discrete cosine transform of the log filter bank outputs was computed, resulting in a 64 dimensional cepstral feature vector. These frames were computed every 11 ms, resulting in a 52% overlap be-

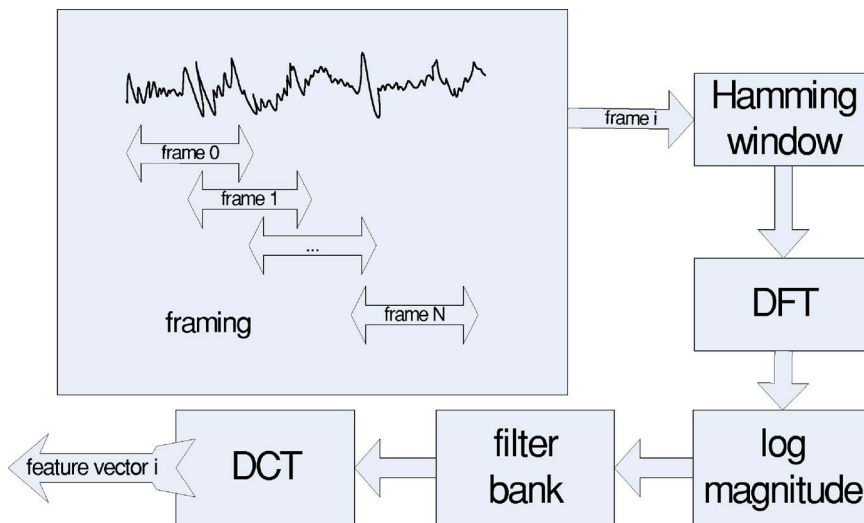


FIG. 2. Flow diagram for feature extraction. Overlapping frames of 21 ms are taken from the signal and transformed to the cepstral domain.

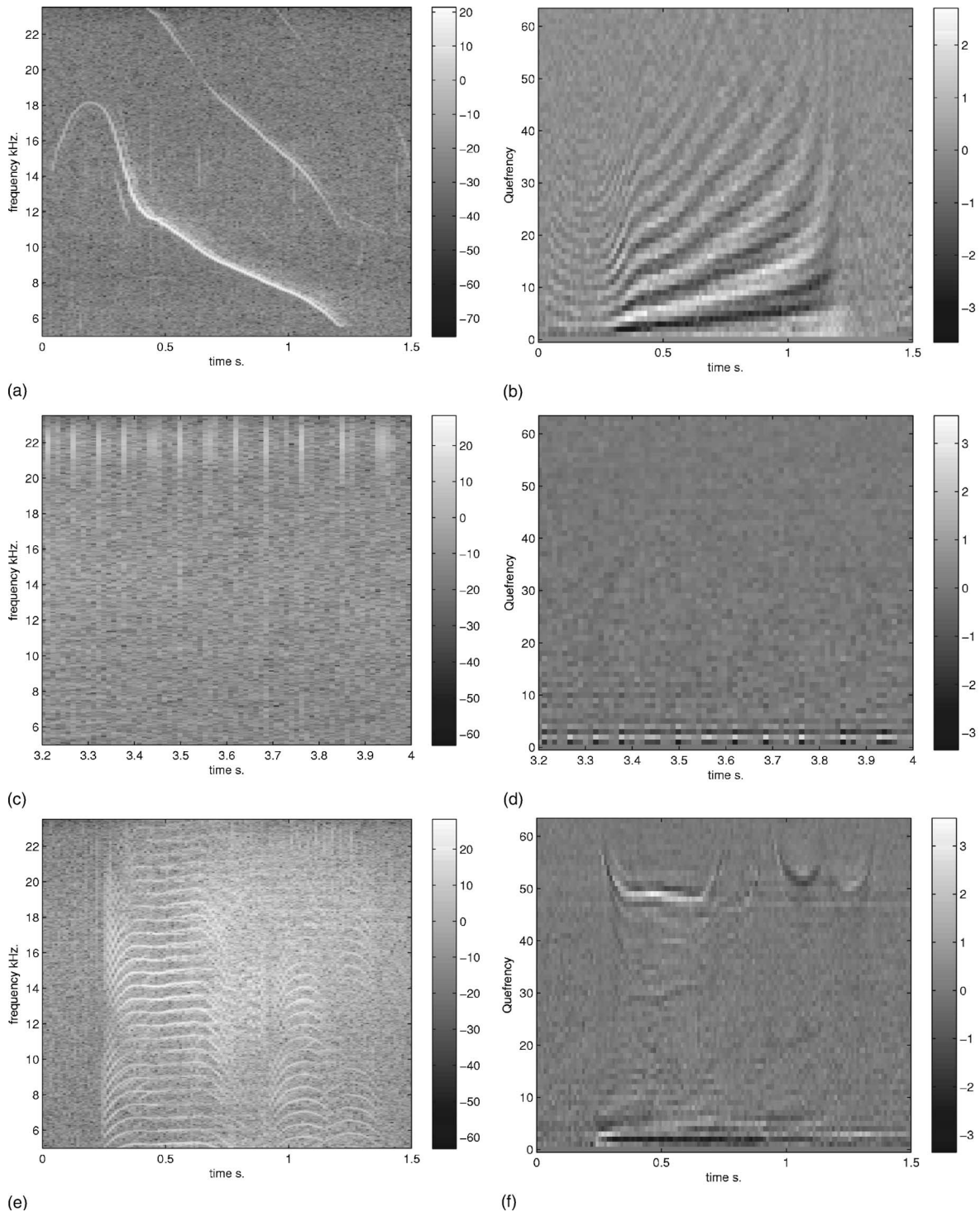


FIG. 3. Spectrograms and cepstral coefficients for (a) and (b) whistles, (c) and (d) clicks, and (e) and (f) burst pulses. The left column shows a spectrogram, and the right column shows the corresponding cepstrogram of the same signal after the application of a 64 point DFT based filter bank. The whistles are produced by bottlenose dolphins, the clicks and burst pulses by Pacific white-sided dolphins.

tween successive frames. Further details of this process can be found in Clemins *et al.* (2005) or Picone (1993).

Figure 3 shows spectrograms and their corresponding cepstrograms for each of the three call types. Cepstrograms are similar to spectrograms and display time series of cepstral vectors. In the whistle cepstrogram, one can see “harmonic” like structure in the cepstral domain related to the frequency modulated (FM) sweep. The harmonics move farther apart from one another as the frequency falls. Figures

3(c) and 3(d) show a spectrogram and cepstrogram for a click train. The location of the clicks is readily apparent in the cepstral domain and the majority of the information is concentrated in the lower queffrequencies (cepstral coefficients). The burst pulse of Fig. 3(e) is also easily seen in Fig. 3(f).

Figure 3 shows information from single calls. In practice, many of the calls in the data set contain overlapping data. Figures 4(a) and 4(b) show the spectrogram and corresponding cepstrogram from a short segment of overlapping

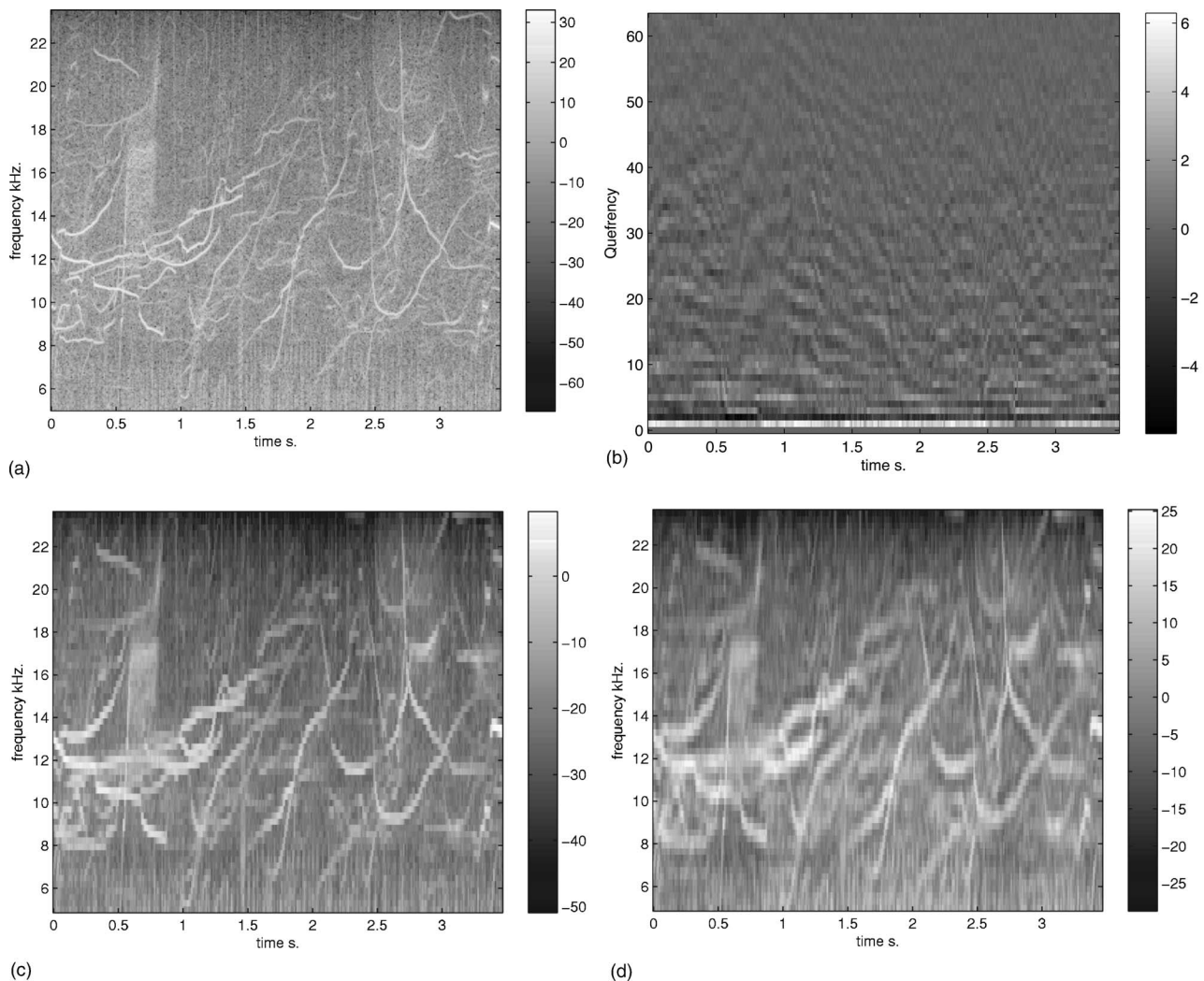


FIG. 4. Illustration of (a) spectrogram and (b) ceprogram for multiple overlapping calls from a group of common dolphins. The classifier seeks patterns associated with multiple calls and does not attempt to separate out individual calls. The effects of a linearly spaced 64 band filterbank are shown in spectrogram (c). The cepstral series can be truncated while still retaining much of the original information as seen in the reconstruction of (c) using the first 32 cepstral coefficients of (b).

calls from common dolphins. To successfully classify this type of data, our strategy is to have the classifier learn the cepstral patterns of overlapping calls rather than attempting to isolate the individual calls.

The effect of the bandpass filtering operation on the same set of common dolphin calls can be seen in Fig. 4(c). As described in Sec. I, discarding part of the cepstrum can result in significant reductions of the feature space. This is advantageous as lower order models typically require less data to train. Higher queffencies correspond to the fine detail in the log spectrum, and provide an opportunity to reduce the dimensionality of the feature space. To determine the effectiveness of the truncated cepstrum for representing delphinid calls, we reconstructed spectrograms by inverting the operations used to form the cepstrum. It was determined that retaining the first 32 frequencies resulted in spectrograms where major features of the call were still clearly evident as illustrated in Fig. 4(d).

An additional step of the feature extraction is to apply cepstral means subtraction which detrends the cepstrum by subtracting the mean vector. This operation has the dual effects of removing the mean from the log spectrum (Herman-

sky, 1995) and removing any constant contribution to the cepstrum caused by the convolution of the signal and the hydrophone. The removal of the hydrophone-specific contribution is critical when there are mismatches between the frequency responses of hydrophones used in the training and test sets. As all hydrophones used in this study were from SSQ-57B sonobuoys, it is assumed that the primary benefit of using cepstral means subtraction is to detrend the log spectrum.

The cepstral coefficients yield a static representation of the short term spectrum. Many audio classification tasks benefit from adding information about how the spectrum is changing, and this can be done by taking the first and second derivatives of the cepstrum. These were appended to the feature vector and improved the accuracy by approximately 20%.

D. Classification

The feature vectors were classified using GMMs. A GMM M consists of N normal distributions with mean μ_i and covariance matrix Σ_i where $1 \leq i \leq N$. Each of these dis-

tributions is scaled by c_i ($1 \leq i \leq N$) and the sum of the c_i 's must be one to ensure that the GMM represents a probability distribution.

The likelihood of each d dimensional observation x can be found by the following (Huang *et al.*, 2001):

$$\Pr(x|M) = \sum_{i=1}^N c_i \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \times \exp\left(-\frac{1}{2}(x - \mu_i)' \Sigma_i^{-1} (x - \mu_i)\right), \quad (1)$$

where the prime ($'$) denotes the transpose operator and $|\cdot|$ the determinant.

The parameters for GMMs cannot be estimated using closed form equations. As the contribution of each training vector to specific mixtures is unknown, standard estimation techniques such as maximum likelihood estimation are not possible. However, if one assumes the existence of a model which approximates the distribution, it is possible to use an iterative algorithm and the training data to find a new model which better approximates the training data. One way of forming the initial model (Young *et al.*, 2002) is to start with a Gaussian classifier (equivalent to a single mixture GMM with $c_1=1$) that can be estimated by using the sample mean and covariance. The single mixture is split into two mixtures with identical covariances and means offset by 0.2 s.d. Each of the new mixtures is assigned half of the old mixture's weight. The iterative algorithm is typically executed for a few iterations and then the mixtures are split again. This is repeated until the desired number of mixtures is formed.

The iterative algorithm is an application of the expectation maximization (EM) algorithm (Moon, 1996). The central idea is that while it is not known which of the N Gaussian distributions is responsible for generating the t^{th} observation x_t , we can use the expected value as an estimator. This is represented by the notation γ_m^t , which is the expected contribution of the m^{th} mixture to the total likelihood associated with observation x_t in the t^{th} version of the model:

$$\gamma_m^t = \frac{c_m \Pr(x_t | \mu_m^{(i)}, \Sigma_m^{(i)})}{\Pr(x_t | M^{(i)})}. \quad (2)$$

With the expectation known, maximum likelihood techniques can be used to produce the next iteration of the model $M^{(i+1)}$:

$$c_m^{(i+1)} = \frac{\sum_{t=1}^T \gamma_m^t}{N}, \quad (3)$$

$$\mu_k^{(i+1)} = \frac{\sum_{t=1}^T \gamma_m^t x_t}{\sum_{t=1}^T \gamma_m^t}, \quad (4)$$

$$\Sigma_k^{(i+1)} = \frac{\sum_{t=1}^T \gamma_m^t (x_t - \mu_k^{(i)})(x_t - \mu_k^{(i)})'}{\sum_{t=1}^T \gamma_m^t}. \quad (5)$$

At each iteration, the likelihood of the training data with respect to the new model is guaranteed to be greater than or equal to the likelihood with respect to the old model. Thus

the EM algorithm will converge to a local maximum. While there are no known proofs of the rate of convergence, convergence is typically fast with anywhere from 5 to 15 iterations. The derivation of these equations can be found in Huang *et al.* (2001).

A number of assumptions were made with respect to our use of GMMs. Iteration was stopped when the likelihood of the new model was no more than 2% greater than the previous one. Based upon the asymptotic independence of the components of cepstral feature vectors (Merhav and Lee, 1993), it was also assumed that the components of the feature vectors were independent. This resulted in diagonal covariance matrices which significantly reduced the computational cost.

Once the models were trained, the posterior probability of each species was computed with respect to a set of test vectors:

$$\Pr(\text{species}|\text{test}) = \frac{\Pr(\text{test}|\text{species})\Pr(\text{species})}{\Pr(\text{test})}. \quad (6)$$

The right-hand side of Eq. (6) was obtained from Bayes rule. The class label for the test segment was decided using Bayes decision rule (Duda *et al.*, 2001), which selected the class that produces the maximum probability.

In this work, a uniform prior distribution was assumed, resulting in $\Pr(\text{species})$ having a constant contribution to each posterior probability. As $\Pr(\text{test})$ was also constant across species, the maximum posterior probability relied solely on the class conditional likelihood, $\Pr(\text{test}|\text{species})$, which was evaluated by Eq. (1). As observations were assumed to be independent from one another, the log likelihoods from each observation were summed to produce the joint posterior likelihood as shown in Fig. 5.

The GMMs were implemented using the Hidden Markov Model Toolkit (HTK) by Young *et al.* (2002), an open source suite of programs for speech recognition. Customizations were made to support the linear filter bank. A series of control programs were written in PYTHON, a general purpose object oriented scripting language.

Experiments were conducted to examine the effect of model order, length of training and test data, and variations of the choice of training data. Table II indicates the partitioning of call data from Table I into training and test data. With the exception of the experiments which examined the choice of training data, all experiments used partition 1. Two sessions were selected for the bottleneck data simply due to the abundance of data available. The common dolphin data from session 1 were not used as training data as they represented the longest session. Using this session would have reduced the number of test cases resulting in an increase of the 95% confidence interval (CI).

When designing the partitions, the authors attempted to minimize the risk that calls from the same animals in the same behavior state were contained in both the training and test sets. On some of the cruises, ship track patterns may have allowed resampling the same group of dolphins on the same day, so sessions from a single day are never split into training and test data. Given the fission-fusion nature of dolphin school groupings (Connor *et al.*, 2000; Neumann, 2001)

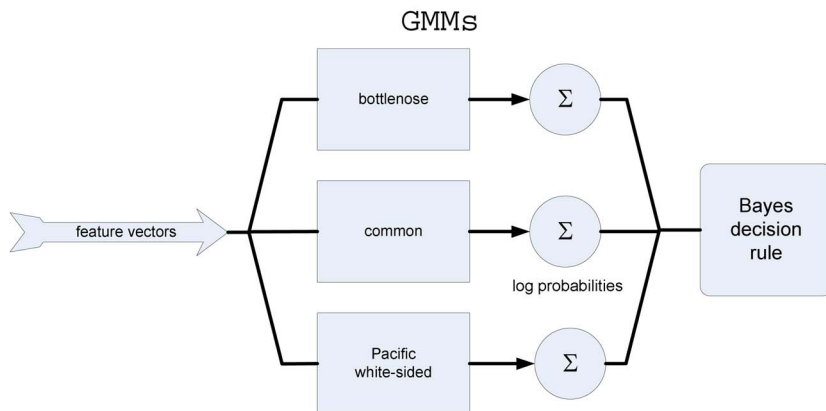


FIG. 5. Classification of sequences of feature vectors by Gaussian mixture models.

along with changes in behavior state, the authors believe that using recordings from different days is a conservative approach to reducing the likelihood of encountering the same group of animals in the same behavior state. Due to the limited number of recording days, a separate evaluation set was not established.

The first set of experiments tested model order. GMMs with 64, 128, 256, and 512 mixtures were trained. The mixture splitting algorithm made powers of two a logical choice for the model order. Test data were partitioned into 20 s segments that were classified as a single unit regardless of the number of calls contained therein.

Test segment length was examined by using 1, 3, 5, 10, 15, 20, 25, and 30 s segments for the best model order associated with this data set. The effect of varying the quantity of training data was also explored. The common dolphin data from session 2 was the shortest data set with 159 s of training data, and all training length tests were designed to be shorter, with tests at 30, 60, 90, and 120 s of training data. The remaining training data were not added to the test set as this would have violated the constraint that calls from the same day should not serve as both training and test data. The bottlenose dolphin data were split between sessions 1 and 5 evenly, and tests were performed with 20 s segments of call data.

To ensure that the results were not overly dependent upon the specific calls used in the training set, three additional experiments were performed with 256 mixture GMMs and 10 s test segments. In each experiment, the training data for one of the groups was substituted with a different set of training data. When feasible, training data of a similar length as in the original set of experiments were used. The test sets were appropriately updated by deleting the new training data

TABLE II. Selection of train/test data. The listed sessions from Table I are used for training with the remaining sessions used as test. Partitions 2–4 were chosen such that the training data for a single species is replaced by one of the other sessions.

Partition	Common	Pacific white-sided	Bottlenose
1	2	1	1, 5
2	3	1	1, 5
3	2	3	1, 5
4	2	1	8, 9

and inserting the old. Table II provides a listing of the sessions used to produce these three new partitions (2–4) of the data set.

Overall accuracy was defined as the percentage of test segments that were correctly identified. If the correctness of the outcome of each test segment is considered as a binomial trial, it is possible to construct a 95% CI for the mean (Huang *et al.*, 2001). The confidence interval is defined as follows:

$$CI(\alpha, p, N) = \pm F_n^{-1} \left(1 - \frac{\alpha}{2} \middle| 0, \sqrt{\frac{p(1-p)}{N}} \right), \quad (7)$$

where $F_n^{-1}(\cdot | \mu, \sigma)$ denotes the inverse cumulative distribution function of a normal distribution with mean μ and standard deviation σ . The confidence interval is controlled by α , which is set to one minus the desired confidence interval (0.05 for the 95% CI). The variable p denotes the accuracy, and N is the number of trials.

A second statistic was defined to prevent the large quantity of bottlenose dolphin calls from biasing the results toward classifiers that favor that species. The average of the per species accuracies was calculated where per species accuracy was defined as the number of correct classifications of each species divided by the number of species-specific classification attempts. Large deviations in this statistic from the overall accuracy are indicative of a classifier bias toward a specific species.

III. RESULTS

The accuracy with respect to the number of mixtures is reported in Fig. 6 where the number of mixtures varies between 64 and 512 by powers of 2. The circles represent the percentage of correctly classified segments from all species, and the error bars show the 95% CI for each test. The classifier accuracy increases as the number of mixtures climbs, with a maximum accuracy of 78.1% with 512 mixtures. The average per species accuracies are plotted with triangles, and reported by species in Fig. 7. With respect to species-specific performance, varying the number of mixtures resulted in trade-offs between common and bottlenose dolphin performance. Pacific white-sided dolphins were always well classified.

The 256 mixture models provided the best balance of overall and per species accuracy, and all subsequent experi-

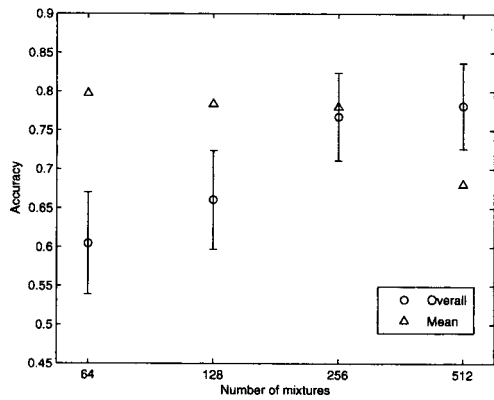


FIG. 6. Classification accuracy using Gaussian mixture models with differing numbers of mixtures. The circles represent the percentage of segments correctly classified and their 95% confidence intervals. The triangles show the means of the classification rates across species. Considering only overall accuracy would bias the classifier toward the species with the greatest number of test utterances.

ments were conducted with 256 mixture GMMs. The overall and species-specific results of varying the test length between 1 and 30 s are summarized in Figs. 8 and 9. Accuracy increased as a function of test segment length up to 10 s. When test segments were longer than 10 s, any further increases generally fell within the 95% confidence interval for means.

The effect of training data length was also investigated with test segments of 20 s. As shown in Table III, reducing the amount of training data impacted both the overall and mean species accuracies. It was of note that the recognition rate for common dolphins was actually higher with shorter amounts of training data and this is discussed in the next section.

The final set of experiments examined the effect of varying the training and test partitions and is reported in Table IV. The experiments on partitions 2 and 3 had overall accuracies that were within the 95% confidence interval of partition 1. The partition 3 experiment showed a marked decrease in the accuracy of identifying Pacific white-sided dolphins, with 10 of the 41 tests incorrectly identified as common dolphins. The partition 4 experiment had a lower accuracy (67.1%).

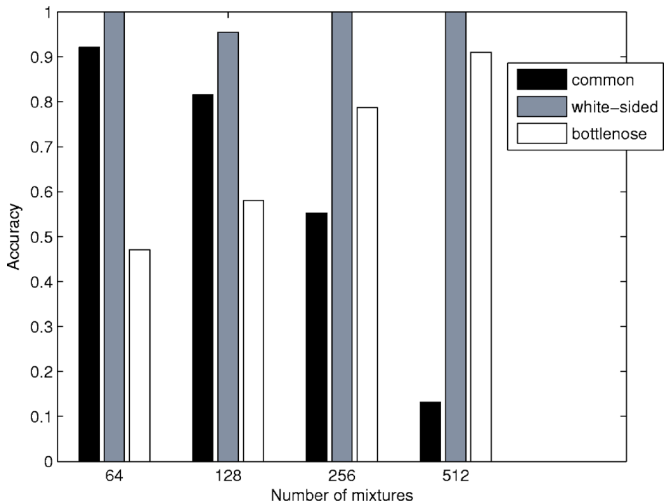


FIG. 7. Accuracy by species as the number of mixtures per GMM varies.

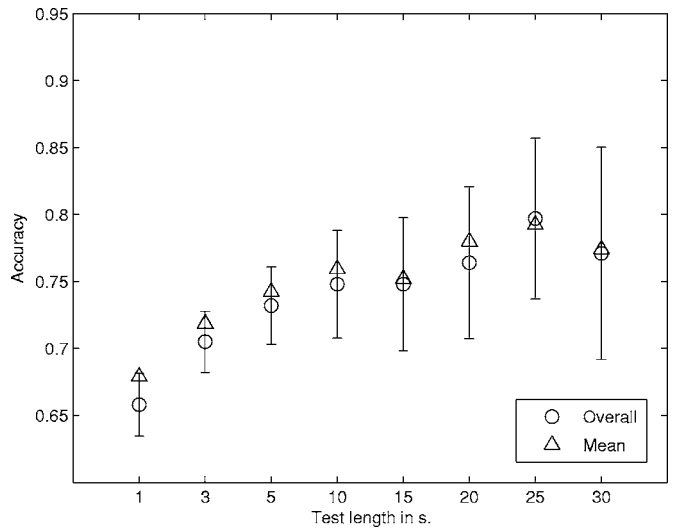


FIG. 8. Classification accuracy when varying the length of test segments classified by 256 mixture GMMs. The circles represent the percentage of segments correctly classified and their 95% confidence intervals. The triangles show the means of the classification rates across species.

IV. DISCUSSION

The experiments which varied the number of mixtures (Fig. 6) suggest that the optimal number of mixtures may vary from one species to another. Pacific white-sided dolphins were always recognized well. This may be related to their calls having less in common with the other species, including a substantially lower number of whistles. The primary trade off appeared to be between common and bottlenose dolphins with bottlenose dolphins better recognized by high order models and common dolphins by low order models. This may be due to the greater complexity of the bottlenose call repertoire, but could also be attributed to differences in the quantity of training data.

When the length of the test segment was varied (Figs. 8 and 9), increases in accuracy were correlated with the test length. However, with the exception of the experiment with a 25 s test length (accuracy: overall 79.7% \pm 6.0%, species mean 79.2%), overall accuracy was contained within the 95% confidence interval of the 10 s test segments. Test segments longer than 10 s did not generally contribute significantly to increased accuracy. As the length of each test segment was increased, the total number of tests decreased. This accounts for the increased spread of the 95% confidence intervals seen in Fig. 8. Unlike the mixture experiment, the per species accuracies generally followed the same trend as the overall accuracy.

Varying the quantity of training data showed that accuracies of above 70% could be seen with as little as 90 s of training data per species. However, this should be taken with a caveat as an equally important question is related to the number of behavior states and environments from which the researcher collects data. One anomalous result within this set was the reduction of accuracy in the common dolphin data with 60 s of training data. This result is difficult to interpret, but may represent similarities between behavior states in different species. If the behavior state occurs in the training data of one of the other species but not in the common dolphin

training data, it may classify to the other species. Another possible explanation is that with reduced training data, there is the danger of learning the environment as opposed to the species. Cepstral means subtraction will reduce the possibility of this occurring for stationary events, but not for other characteristics of the auditory scene.

Our examination of the training data variation experiments revealed that the Pacific white-sided dolphin data in session 3 contained brief but periodic signal drop outs due to ship board radio interference with the sonobuoy signals. This may contribute to the decreased accuracy when these data are used as training material (during test, the drop outs are small portions of any 10 s segment and are less likely to play a large role in the overall likelihood). Alternatively, it may be that the Pacific white-sided call set in partition 1 has more variation in the call data than that of partition 3.

The experiment with partition 4 was the only experiment that did not fall within the 95% confidence interval of the first partition's overall accuracy. The lower accuracy of $67.1\% \pm 4.4\%$ was due entirely to bottlenose dolphins which were misclassified as common dolphins. As with the decrease in Pacific white-sided dolphins, it is possible that the training set was less diverse. In addition, this experiment is the only experiment that used bottlenose training data exclusively from the Gulf of California and tested using data from both the Gulf of California and the Southern California Bight. An examination of the error rate of calls recorded in the Southern California Bight showed a below average accuracy of 33.3%, but the number of test segments ($N=24$) from a single day's recordings is too small to draw any conclusions about possible dialectal differences between the two groups. It should also be noted that other segments were misclassified. The 16 misclassifications of the Southern California Bight population were in a larger context of 105 bottlenose misclassifications. Recalling that the bottlenose data set is much larger than that of the other species, it should be noted that the mean accuracy of the three species is similar to the other partition tests.

Fluctuations in accuracy are to be expected with different partitions of the call data. Nonetheless, classification accuracy for three of the four tests fell within each other's 95% confidence interval for means. In all cases, classification accuracy was well above chance levels of guessing using a uniform prior (33.3%).

Finally, it should be noted that some of the choices made for feature selection were the result of compromises and will be the subject of future work. The current analysis window of 21 ms is too long for clicks which can be as short as 40 μ s in bottlenose dolphins (Au, 1993). Similarly, the trun-

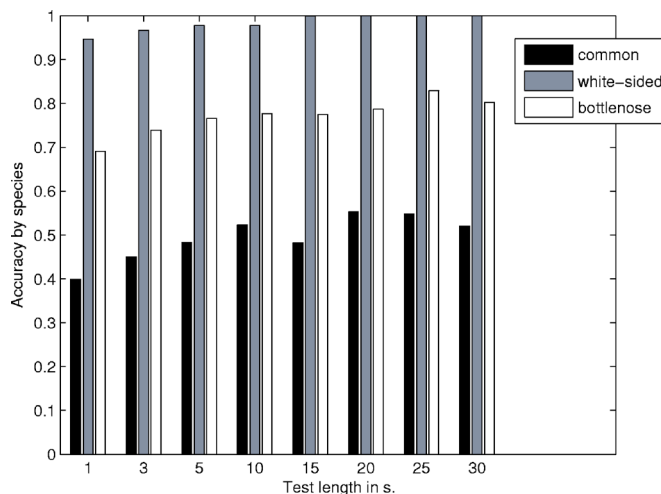


FIG. 9. Accuracy by species as the length of the test segment classified by a 256 mixture GMM is varied.

cation of the cepstral series to 32 coefficients results in the loss of the inclick interval for burst pulses, which may be of relevance in species identification. As inclusion of the interclick intervals would have resulted in a significant increase of the dimensionality of the feature vector, a better strategy to include this may be to estimate the interclick interval separately and include it as one additional component of the feature vector. These issues are the subject of future work, but the current system still proves to be an effective classifier.

V. CONCLUSION

A system has been presented for the classification of free ranging dolphin calls that functions with an accuracy which generally ranges from 67% to 75% with a 95% CI of approximately $\pm 4\%$. The system does not rely on specific call types and there is no requirement to separate out individual calls from the aggregate, which can be difficult with large groups of highly social and vocal animals. Using a cepstral feature space allows the system to capture the timbre of the calls, which is lost in systems that extract simple parameters from whistle curves. The use of supervised statistical learning permits the system to be used for other species and tasks with little modification.

We have not discussed comparisons between our system's performance and that of other call recognition systems. In addition to differences between the detection, feature extraction, and classification approach that each method uses, there are a number of parameters that make direct comparisons between methods tenuous at best. The vast differences

TABLE III. Performance of 256 mixture GMMs with 20 s test segments and varying amounts of training data per species.

Train s	Overall accuracy	$\pm 95\%$ CI	Species mean	Species		
				Common	Pacific white-sided	Bottlenose
30	0.512	0.067	0.682	0.842	0.818	0.387
60	0.735	0.059	0.666	0.158	1.000	0.839
90	0.767	0.057	0.714	0.447	0.864	0.832
120	0.725	0.060	0.741	0.474	1.000	0.748

TABLE IV. Results of 256 mixture GMM tests with 10 s test segments on the partitions resulting from Table II. Column CI is the 95% confidence interval on the overall accuracy, and species mean represents the mean of the individual species' accuracies.

Partition	Test count	Overall accuracy	$\pm 95\%$ CI	Species mean	Common	Pacific white-sided	Bottlenose
1	449	0.748	0.040	0.759	0.523	0.978	0.777
2	433	0.707	0.043	0.739	0.557	0.956	0.704
3	447	0.729	0.041	0.690	0.523	0.767	0.780
4	447	0.671	0.044	0.723	0.523	0.978	0.668

in collection methods and difficulty of specific corpora have led human speech classification researchers to offer yearly competitions with common tasks and data sets (e.g. Przybocki and Martin, 2001).

While the great variety of bioacoustic classification tasks make the establishment of standards difficult, the authors believe that members of the bioacoustics community should continue to take steps to establish common data (Gaunt *et al.*, 2005) and software repositories which will permit direct comparison of classification algorithms.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their comments and suggestions as well as Ted Cranford for his helpful discussions on dolphin vocalizations. This work was supported by US Navy Chief of Naval Operations N45 – Frank Stone and Ernie Young, the Strategic Environmental Research and Development Project–Robert Holst, the Office of Naval Research–Bob Gisiner and Ellen Livingston, and US Government Award No. 0016.

Acevedo-Gutiérrez, A., and Stienessen, S. C. (2004). "Bottlenose dolphins (*Tursiops truncatus*) increase number of whistles when feeding," *Aquat. Mamm.* **30**, 357–362.

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

Buck, J. R., and Tyack, P. L. (1993). "A quantitative measure of similarity for *Tursiops truncatus* signature whistles," *J. Acoust. Soc. Am.* **94**, 2497–2506.

Caldwell, M. C., and Caldwell, D. K. (1968). "Vocalization of naive captive dolphins in small groups," *Science*, **159**, 1121–1123.

Caldwell, M. C., and Caldwell, D. K. (1971). "Statistical evidence for individual signature whistles in Pacific whitesided dolphins, *Lagenorhynchus obliquidens*," *Cetology* **3**, 1–9.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (1990). "Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego).

Clemins, P. J., and Johnson, M. T. (2006). "Generalized perceptual linear prediction feature for animal vocalization analysis," *J. Acoust. Soc. Am.* **120**, 527–534.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (2005). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," *J. Acoust. Soc. Am.* **117**, 956–963.

Connor, R. C., Wells, R. S., Mann, J., Read, A. J., Tyack, P. L., and Whitehead, H. (2000). "The bottlenose dolphin: Social relationships in a fission-fusion society," in *Cetacean Societies: Field Studies of Dolphins and Whales*, edited by J. Mann *et al.* (University of Chicago Press, Chicago), pp. 91–126.

Cranford, T. W. (2000). "In search of impluse sound sources in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 109–155; private communication (2005).

Datta, S., and Sturtivant, C. (2002). "Dolphin whistle classification for determining group identities," *Signal Process.* **82**, 127–327.

Deecke, V. B., Ford, J. K. B., and Spong, P. (1999). "Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects," *J. Acoust. Soc. Am.*, **105**, 2499–2507, 2006.

Deecke, V. B., and Janik, V. M. (2006). "Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls," *J. Acoust. Soc. Am.* **199**, 645–653.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*, 2nd ed. (Wiley-Interscience, New York).

Evans, W. E. (1973). "Echolocation by marine delphinids and one species of freshwater dolphin," *J. Acoust. Soc. Am.* **54**, 191–199.

Fahner, M., Thomas, J., Ramirez, K., and Boehm, J. (2004). "Echolocation in bats and dolphins," in *Acoustic Properties of Echolocation Signals by Captive Pacific White-Sided Dolphins (*Lagenorhynchus obliquidens*)*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago).

Fenton, M. B., and Bell, G. P. (1981). "Recognition of species of insectivorous bats by their echolocation calls," *J. Mammal.* **62**, 233–243.

Fish, J. E., and Turl, C. W. (1975). "Source level of four species of small toothed whales," in *Conference on the Biology and Conservation of Marine Mammals*, Santa Cruz.

Fitch, W. T. (2000). "The evolution of speech: A comparative review," *Trends Cogn. Sci.* **4**, 258–267.

Gaunt, S. L. L., Nelson, D. A., Dantzer, M. S., Budney, G. F., and Bradbury, J. W. (2005). "New directions for bioacoustics collections," *Auk* **122**, 984–988.

Goold, J. C., and Jones, S. E. (1995). "Time and frequency-domain characteristics of sperm whale clicks," *J. Acoust. Soc. Am.* **98**, 1279–1291.

Harrington, J., and Cassidy, S. (1999). *Techniques in Speech Acoustics* (Kluwer Academic, Dordrecht).

Herman, L. M., and Tavolga, W. N. (1980). "The communication system of cetaceans," in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Herman (Wiley Interscience, New York), pp. 149–209.

Hermansky, H. (1990). "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Am.* **87**, 1738–1752.

Hermansky, H. (1995). "Exploring temporal domain for robustness in speech recognition," in *The 15th International Congress on Acoustics*, Trondheim, Norway, Vol. 2, pp. 61–64.

Houser, D. S., Helweg, D. A., and Moore, P. W. (1999). "Classification of dolphin echolocation clicks by energy and frequency distributions," *J. Acoust. Soc. Am.* **106**, 1579–1585.

Huang, X., Acero, A., and Hon, H.-W. (2001). *Spoken Language Processing* (Prentice Hall PTR, Upper Saddle River, NJ).

Kogan, J. A., and Margoliash, D. (1998). "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," *J. Acoust. Soc. Am.* **103**, 2185–2196.

Marler, P. (1957). "Specific distinctiveness in the communication signals of birds," *Behaviour* **11**, 13–39.

Mellinger, D. K., and Clark, C. W. (2000). "Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am.* **107**, 3518–3529.

Merhav, N., and Lee, C.-H. (1993). "On the asymptotic statistical behavior of empirical cepstral coefficients," *IEEE Trans. Signal Process.* **41**, 1990–1993.

Moon, T. K. (1996). "The expectation-maximization algorithm," *IEEE Signal Process. Mag.* **13**, 47–60.

Moore, S. E., and Ridgway, S. H. (1995). "Whistles produced by common dolphins from the Southern California Bight," *Aquat. Mamm.* **21**, 55–63.

Murray, S. O., Mercado, E., and Roitblat, H. L. (1988). "The neural network

- classification of false killer whale (*Pseudorca crassidens*) vocalizations," *J. Acoust. Soc. Am.* **104**, 3626–3633.
- Nakamura, K., and Akamatsu, T. (2004). "Comparison of click characteristics among odontocete species," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago).
- Neumann, D. R. (2001). "The activity budget of free-ranging common dolphins (*Delphinus delphis*) in the northwestern Bay of Plenty, New Zealand," *Aquat. Mamm.* **27**, 121–136.
- Oswald, J., Rankin, S., and Barlow, J. (2004). "The effect of recording and analysis bandwidth on acoustic identification of delphinid species," *J. Acoust. Soc. Am.* **116**, 3178–3185.
- Oswald, J. N., Barlow, J., and Norris, T. F. (2003). "Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean," *Marine Mammal Sci.* **19**, 20–37.
- Oswald, J. N., Rankin, S., Barlow, J., and Lammers, M. O. (2005). "A new tool for real-time acoustic species identification of delphinid whistles," *J. Acoust. Soc. Am.* **118**, 1909 (abstract only).
- Picone, J. W. (1993). "Signal modeling techniques in speech recognition," *Proc. IEEE* **81**, 1215–1247.
- Popper, A. N. (1980). "Sound emission and detection by delphinids," in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Herman (Krieger, Malabar, FL).
- Potter, J. R., Mellinger, D. K., and Clark, C. W. (1994). "Marine mammal call discrimination using artificial neural networks," *J. Acoust. Soc. Am.* **96**, 1255–1262.
- Przybocki, M. A., and Martin, A. F. (2001). "The NIST speaker recognition evaluation: 1996-2001," in *Proceedings of the Odyssey*, Anogia, Crete, Greece.
- Rabiner, L. R., and Juang, B.-H. (1993). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
- Rendell, L. E., Matthews, J. N., Gill, A., Gordon, J. C. D., and Macdonald, D. W. (1999). "Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation," *J. Zool.* **249**, 403–410.
- Sirovic, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. (2004). "Seasonality of blue and fin whale calls and the influence of sea lee in the Western Antarctic Peninsula," *Deep-Sea Res., Part II*, **51**, 2327–2344.
- Steiner, W. W. (1981). "Species-specific differences in pure tonal whistle vocalizations of five western north Atlantic dolphin species," *Behav. Ecol. Sociobiol.* **9**, 241–246.
- Sundberg, J. (1991). *The Science of Musical Sound* (Academic, San Diego).
- Thompson, D. H., and Richardson, W. J. (1995). "Marine mammal sounds," in *Marine Mammals and Noise*, edited by W. J. Richardson, C. R. Greene, Jr., C. I. Malme, and D. H. Thomson (Academic, San Diego), pp. 325–386.
- Thompson, P., 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.
- Thompson, P. O., Findley, L. T., Vidal, O., and Cummings, W. C. (1996). "Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico," *Marine Mammal Sci.* **12**, 288–293.
- Thompson, P. O., and Friedl, W. A. (1982). "A long term study of low frequency sound from several species of whales off Oahu, Hawaii," *Cetology* **45**, 1–19.
- Wang, D., Wursig, B., and Evans, W. (1995). "Comparisons of whistles among seven odontocete species," in *Sensory Systems of Aquatic Mammals*, edited by J. A. Kastelein, R. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, NL).
- Whitten, J. L., and Thomas, J. A. (2001). "Whistle repertoire of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) at the John G. Shedd Aquarium," *J. Acoust. Soc. Am.* **109**, 2391 (abstract only).
- Wiggins, S. (2003). "Autonomous acoustic recording packages (ARP's) for long-term monitoring of whale sounds," *Mar. Technol. Soc. J.* **37**, 13–22.
- Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (2002). The HTK book, version 3.2, URL <http://htk.eng.cam.ac.uk>. Last viewed 2/8/07.

Echo-intensity compensation in echolocating bats (*Pipistrellus abramus*) during flight measured by a telemetry microphone

Shizuko Hiryu^{a)}

Research Center for Intelligent Information Science, Doshisha University, Kyotanabe 610-0321, Japan

Tomotaka Hagino

Department of Electrical Engineering, Doshisha University, Kyotanabe 610-0321, Japan

Hiroshi Riquimaroux

Department of Intelligent Information Engineering and Sciences, Doshisha University, Kyotanabe 610-0321, Japan and Bio-navigation Research Center, Doshisha University, Kyotanabe 610-0321, Japan

Yoshiaki Watanabe

Department of Electrical Engineering, Doshisha University, Kyotanabe 610-0321, Japan and Bio-navigation Research Center, Doshisha University, Kyotanabe 610-0321, Japan

(Received 3 October 2006; revised 20 November 2006; accepted 8 December 2006)

An onboard microphone (Telemike) was developed to examine changes in the basic characteristics of echolocation sounds of small frequency-modulated echolocating bats, *Pipistrellus abramus*. Using a dual high-speed video camera system, spatiotemporal observations of echolocation characteristics were conducted on bats during a landing flight task in the laboratory. The Telemike allowed us to observe emitted pulses and returning echoes to which the flying bats listened during flight, and the acoustic parameters could be precisely measured without traditional problems such as the directional properties of the recording microphone and the emitted pulse, or traveling loss of the sound in the air. Pulse intensity in bats intending to land exhibited a marked decrease by 30 dB within 2 m of the target wall, and the reduction rate was approximately 6.5 dB per halving of distance. The intensity of echoes returning from the target wall indicated a nearly constant intensity (-42.6 ± 5.5 dB weaker than the pulse emitted in search phase) within a target distance of 2 m. These findings provide direct evidence that bats adjust pulse intensity to compensate for changes in echo intensity to maintain a constant intensity of the echo returned from the approaching target at an optimal range. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431337]

PACS number(s): 43.80.Ka [JAS]

Pages: 1749–1757

I. INTRODUCTION

Bats (Microchiroptera) possess a highly developed sonar system. As biosonar animals, they are capable of echolocation that recognizes the physical attributes of their environment with great accuracy by comparing emitted pulses to returning echoes (Griffin, 1958). Echolocation pulses can differ markedly in structure, consisting of either only short downward frequency-modulated (FM) pulses or long constant-frequency (CF) pulses followed by a short FM component. Such variation in the structure of echolocation pulses is thought to reflect the adaptations of different species to the constraints imposed by their foraging behavior (Simmons and Stein, 1980; Neuweiler, 1984).

Rhinolophids, Hipposiderids, and a Mormoopid (the mustached bat, *Pteronotus parnellii*) all use a compound CF-FM pulse. As a bat approaches a target, the frequency of the returning echo is Doppler-shifted depending on the velocity of the bat in relation to the target. These bats compensate for the Doppler shifts by changing their pulse frequencies so that the echo frequencies remain constant and are precisely analyzed within the narrowly tuned frequency

range in which the bats can hear best. This behavior is termed “Doppler-shift compensation” (DSC) and is an important behavioral adaptation for echolocation in CF-FM bats (Schnitzler, 1968; Schuller *et al.*, 1974; Simmons, 1974; Gustafson and Schnitzler, 1979; Trappe and Schnitzler, 1982; Gaioni *et al.*, 1990; Lancaster *et al.*, 1992; Keating *et al.*, 1994).

During flight, bats decrease the intensity of their emitted pulses when they approach a prey item or an obstacle (e.g., Griffin, 1958; Jen and Kamada, 1982; Vogler and Neuweiler, 1983; Kick and Simmons, 1984; Hartley *et al.*, 1989; Tian and Schnitzler, 1997; Boonman and Jones, 2002). Kobler and colleagues (1985) have shown that a bat attached to a pendulum and swung toward a large fixed target decreases the intensity of its emitted pulse during the forward swing and increases the intensity during the backward swing. This behavior indicates the possibility of another compensation mechanism, in which pulse intensity is adjusted in relation to the distance to a target, resulting in maintenance of echo intensity within the optimal sensitivity range (echo-intensity compensation). Subsequently, reduction of pulse intensity has been quantitatively investigated in bats given a flight task (Hartley *et al.*, 1989; Waters and Jones, 1995; Tian and Schnitzler, 1997; Boonman and Jones, 2002) and using a

^{a)}Electronic mail: shiryu@mail.doshisha.ac.jp

pendulum device or a moving-target apparatus to elicit and measure changes in pulse intensity (Gaioni *et al.*, 1990; Hartley, 1992b). Adjustment of pulse intensity by bats has been discussed in a number of studies as being related to echo-intensity compensation (e.g., Kick and Simmons, 1984; Hartley *et al.*, 1989; Hartley, 1992a; b; Simmons *et al.*, 1992; Waters and Jones, 1995; Tian and Schnitzler, 1997; Boonman and Jones, 2002). Recently, Au and Benoit-Bird (2003) demonstrated that free-ranging dolphins in the wild changed the amplitude of emitted echolocation signals depending on the range of the target to compensate for propagating loss in sound. Therefore, echo-intensity compensation may be effective in the biosonar systems of both bats and dolphins. However, accurate measurement of the intensity of directional pulses emitted by biosonar animals is still considered difficult using the traditional microphone system. For quantitative analysis of the pulse intensity which the flying bats actually emitted, recorded pulses required several corrections for compensating the directional properties of the recording microphone and the emitted pulse, and also traveling loss of sound in the air depending on the distance between the bat and the microphone (e.g., Hartley *et al.*, 1989; Tian and Schnitzler, 1997; Boonman and Jones, 2002). Given these constraints, the pulse intensity measured by the traditional microphone potentially confounded interpretation of a bat's adjustment of pulse intensity in relation to its spatiotemporal behavior during echolocation. Furthermore, measuring the intensity of echoes as well as emitted pulses is necessary to reveal any echo-intensity compensation behavior employed by the animals. However, the relationship between the pulse and returning echo intensities has not been experimentally investigated in detail for bats during flight.

In this study, we examined echolocation behavior in Japanese house bats (*Pipistrellus abramus*, an FM bat) given a flight landing task in a laboratory. Echolocation sounds were recorded by a telemetry microphone (Telemike) mounted on the head of the bat (Henson *et al.*, 1987; Lancaster *et al.*, 1992; Riquimaroux and Watanabe, 2000; Hiryu *et al.*, 2005). Since well-studied FM bat species are small (e.g., *P. abramus* weighs 5–8 g) in relation to the weight of telemetry microphones, sound recording using telemetry microphones has never been used to investigate echolocation in flying FM bats. We developed a small Telemike that was light enough to be carried by the animals, and combined the Telemike with a high-speed video camera system in order to examine the relationship between the echolocation sounds and the spatiotemporal behavior of bats during flight.

Bats are assumed to adjust their call parameters by feedback control in response to the attended returning echoes. Therefore, it is particularly important to investigate the behavioral response (call parameters) of bats associated with their corresponding acoustic stimuli (echoes) under the spatiotemporal conditions in which the pulse-echo pairs were produced. Since *P. abramus* emit short downward FM pulses with a mean duration of approximately 1 ms, the echoes arriving from targets could be recorded by the microphone above the head without overlapping with the outgoing pulses. We analyzed the signal characteristics of pulse-echo pairs to which the flying bats actually listened, and investi-

gated adjustments of call parameters in the context of echo-intensity compensation.

II. MATERIALS AND METHODS

A. Subjects

Three adult Japanese house bats (*P. abramus*) were used in this study. The animals were captured from a large colony roosting in bridge girders near the campus of Doshisha University, Japan. *Pipistrellus abramus* is a member of the Vespertilionidae and is commonly found in Japan. Although closely related pipistrelle bats such as *P. pipistrellus* and *P. kuhli* have been studied extensively (e.g., Schnitzler *et al.*, 1987; Kalko, 1995; Waters and Jones, 1995; Holderied and Helversen, 2003), the echolocation characteristics of *P. abramus* are not well known. The body mass of *P. abramus* ranges from 5 to 8 g, and the wingspan measures approximately 10 cm. Echolocation pulses are emitted through the mouth. The bats were kept in a rearing cage of 0.9 m (L) \times 0.9 m (W) \times 0.6 m (H), and were allowed free access to food (mealworms) and water. On alternate days, the bats were allowed to fly freely for several hours in a large flight room. The experiments complied with the *Principles of Animal Care*, publication no. 86-23, revised 1985, of the National Institutes of Health, and with current Japanese laws.

B. General experimental procedure

All experiments were conducted in a flight chamber measuring 8 m (L) \times 3 m (W) \times 2 m (H) under long wavelength lighting with red filters (>650 nm) to avoid optical effects (Hope and Bhatnagar, 1979; Ghose and Moss, 2003). The chamber was made of steel plates to minimize interference from external electromagnetic waves. All internal walls were painted in black. The bats were released at one end of the flight chamber and allowed to fly freely to the opposite end where a landing mesh [1 m (W) \times 0.7 m (H)] was attached to the wall 1.8 m above the floor. This wall is referred to as the target wall during the landing flight task. Flight behavior was recorded as a flying bat approached the target wall for landing. Recording was conducted for three flight sessions of each bat (a total of nine flight sessions).

C. Telemike system

Echolocation sounds were recorded by a custom-made telemetry microphone (Telemike) mounted on the back of the bat (Fig. 1). The Telemike consists of a 1/8-inch omnidirectional condenser microphone (Knowles, FG-3329, Illinois, USA), an FM transmitter unit, a hearing aid battery of 1.5 V (Sony, SR421SW, Tokyo, Japan), and a transmitting antenna. The Telemike was attached to the back of a bat with a piece of double-sided glue tape, with the microphone positioned approximately 1 cm above the mouth. Because the Telemike weighed less than 0.6 g, including the battery, it was light enough to be carried by bats weighing about 5 g. The bats did not exhibit any fatigue during the experiments. Removal of the Telemike from the back of a bat after each experiment was facilitated by use of a parting agent to avoid skin irritation.

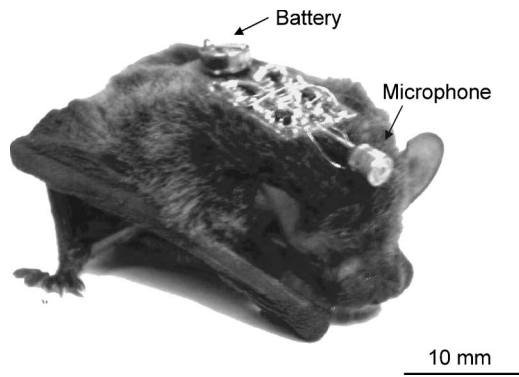


FIG. 1. Japanese house bat (*P. abramus*) with an onboard microphone (Telemike) mounted on its back. The Telemike consists of a 1/8-inch condenser microphone, FM transmitter, battery, and transmitting wire antenna (weight: <0.6 g, including the battery).

The Telemike transmitted signals with a carrier frequency between 100–105 MHz to a wire antenna attached to the ceiling of the flight chamber. Received signals were demodulated using a custom-made FM receiver. The signals were then high-pass filtered at 20 kHz (NF Corporation, model 3625, Yokohama, Japan), digitized by a 16-bit, 384 kHz DAT recorder (SONY, SIR-1000W, Tokyo, Japan), and stored on a hard disk of a personal computer. Prior to recording, the output of the Telemike system was calibrated using a loudspeaker (Pioneer, PT-R7, Tokyo, Japan) and Brüel and Kjaer 1/8-in. microphone. The total frequency response of the Telemike system was within ± 4 dB flat between 20 and 100 kHz. Since echolocation sounds were recorded directly by the microphone above the head of the flying bat, acoustic parameters such as amplitude could be measured precisely without interference from traveling loss of sound in the air or directional properties of the emitted pulse.

D. Three-dimensional reconstruction of spatiotemporal echolocation behavior

The flight behavior was recorded using a dual digital high-speed video camera system (NIPPON ROPER Co., Ltd., CR Imager model 2000s, Chiba, Japan). Cameras were placed at a corner of the flight chamber and did not interfere with the bats' flight path. The frame rate was 125 per second. Three-dimensional coordinates of the flying bats were reconstructed from these video images using a commercial motion analysis software (DITECT, Dipp-Motion 2D ver. 2.1). Prior to recording bat flights, a reference frame with known coordinates was positioned in the center of the flight chamber, then recorded by two video cameras. The analysis software calibrated the reconstruction system with the coordinate data of the reference frame. Based on a direct linear transformation technique, the position of the flying bat or other object was reconstructed from two-dimensional coordinate data in the video images. The signal triggering the video cameras was digitally stored using a DAT recorder so that flight coordinates could be synchronized with sound data. Using three-dimensional coordinate data, the flight trajectory of the bat and the distance from the target wall (target distance) or other objects could be determined. The echo delay t was

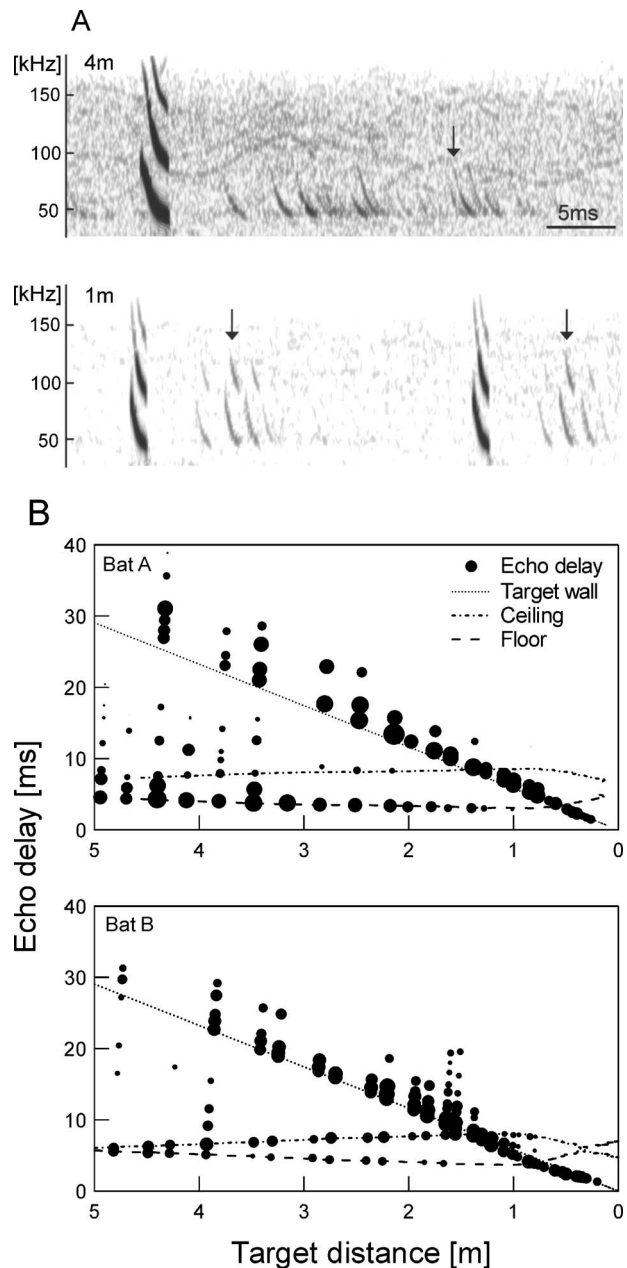


FIG. 2. Pulse-echo pairs recorded with the Telemike while a bat approached the target wall. (A) Sonograms at a target distance of 4 m (top) and 1 m (bottom). Arrows indicate the echoes returning from the target wall (see text). (B) Changes in echo delay determined from sound data as a function of target distance during the landing flight for two bats (solid circles). Circle size indicates relative variation in echo intensity where bigger circles designate sounds of greater intensity relative to smaller circles. The three lines represent echo delays between the flying bat and three targets—the target wall, floor, and ceiling—calculated using three-dimensional coordinate data of the bat.

calculated from the distance between the flying bat and an object d using the formula $t=2d/c$, where c is the sound velocity in the air.

E. Sound analysis

The acoustical characteristics of echolocation sounds were analyzed from the sonogram using a custom program of Matlab on a personal computer. Figure 2(A) shows sonograms of typical pulse-echo pairs recorded by the Telemike

TABLE I. Signal characteristics of the echolocation pulse from *P. abramus* at rest. iFM: initial FM; tFM: terminal FM; BW: bandwidth.

	N	Duration [ms]	iFM [kHz]	tFM [kHz]	BW [kHz]	Frequency at peak energy [kHz]	Min-Max SPL (dB peak-to-peak)	Mean SPL (dB peak-to-peak)
Bat A	100	0.90±0.14	79.3±4.71	42.7±2.73	36.6±5.72	54.5±5.91	97–120	112
Bat B	100	1.20±0.11	84.9±4.81	42.6±2.65	42.3±5.28	56.4±9.29	100–117	112
Bat C	100	1.57±0.30	83.6±4.60	46.0±2.86	37.6±5.87	57.4±8.57	99–117	109

when the flying bat approached the target wall. Near the center of the chamber (4 m from the target wall), a number of echoes (referred to as an echo train) were usually observed with different echo delays (the time difference between each pulse-echo pair) and intensities following an emitted pulse. The second and higher harmonic components of the returning echo were most often attenuated beyond 3–4 m from the target wall. On the other hand, when the bat was within a target distance of 1 and 2 m, the number of echoes reaching the flying bat decreased. The echoes were clearly separated from each other, and the second harmonic component appeared in the sonogram. In this study, only the fundamental component of the pulse-echo pair was analyzed. Each pulse or echo was extracted from the sonogram, and the echo delay was measured for all echoes observed by the Telemike. The sound pressure level of the pulse was calculated from the peak-to-peak amplitude voltage of the observed pulse in the time domain. Simultaneously, the sonogram exhibited a peak in energy at around 50–60 kHz in each sound of *P. abramus* during flight. Therefore, the spectral energy at the peak energy portion of each pulse and echo was measured in the displayed sonogram using the custom program of Matlab so that changes in intensity of a weak echo could be quantitatively evaluated in relation to the pulse intensity. The maximum magnitude of the spectral energy in each sound is referred to as the sound intensity in this study.

Typical changes in the echo delay determined from sound data are shown in Fig. 2(B) as a function of target distance during landing as the bat approached the target wall. The size of the solid circles indicates the relative variation in echo intensity where bigger circles designate sounds of greater intensity relative to smaller circles. In this flight experiment, the Telemike allowed us to observe echoes with an echo delay in the range of approximately 0.5–30 ms, corresponding to targets 0.1–5 m away from the bat. The three lines shown in Fig. 2(B) represent echo delays between the flying bat and three different objects—the target wall, floor, and ceiling of the chamber—calculated from the three-dimensional coordinate data of the flying bat. By comparing

the echo delays shown in sound data (solid circles), the echoes from each surrounding wall of the flight chamber could be identified in the observed sonogram. For example, the arrows in Fig. 2(A) indicate echoes from the target wall. Echoes from the floor and ceiling of the chamber consistently showed intense sound pressure, as well as echoes from the target wall directly in the flight path of the bat.

Prior to flight recording, pulses of the bat in a stationary position were recorded using the Telemike for quantitative analysis. The acoustical parameters of the echolocation pulses were analyzed using the custom program of Matlab described above. Signal bandwidth and duration were determined from the sonogram at –25 dB relative to the peak intensity of the pulse.

III. RESULTS

A. General echolocation behavior of *P. abramus*

Pipistrellus abramus emits a short downward FM pulse with maximum energy at the fundamental component. The mean duration at rest is 1.22±0.34 ms ($n=300$), and the fundamental frequency was modulated from approximately 83 to 44 kHz by the three bats (Table I). The sound pressure level (SPL) of an emitted pulse at rest ranged from approximately 100 to 120 dB peak-to-peak (re 20 μPa), with an average of 111 dB at the microphone above the bat’s head. The sonogram exhibited a peak in energy at 56.0±8.12 kHz, which was 11–14 kHz higher than the terminal frequency.

For the landing experiment, a total of nine flight sessions were recorded for three bats. In flight, the pulse duration was elongated to 3–4 ms, and then decreased to 0.5 ms before landing the target wall. Figure 3 shows the envelope and frequency structure of a typical echolocation pulse emitted by a *P. abramus* in flight, recorded by the Telemike at a target distance of 3 m. The frequency of the fundamental component was modulated exponentially from 100 to 40 kHz, and the bandwidth of the emitted pulse was extended when in flight. A prominent CF-like portion (a shallow sweep portion at the end of the pulse; see Fig. 3) was

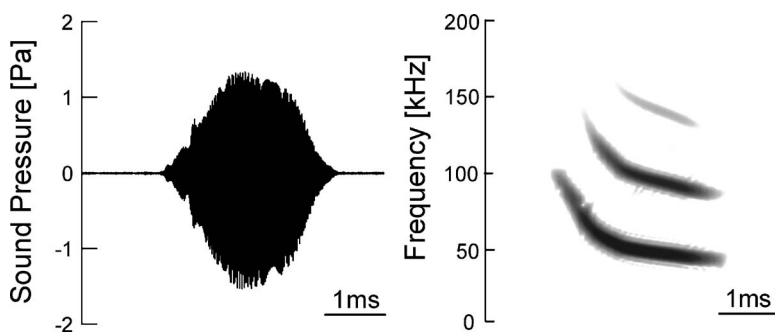


FIG. 3. Typical echolocation pulse emitted by *P. abramus* during the landing flight, recorded by the Telemike. Data were taken from sound recorded at a distance of 3 m from the target wall. Echolocation pulses usually contained several harmonics, with the first being dominant.

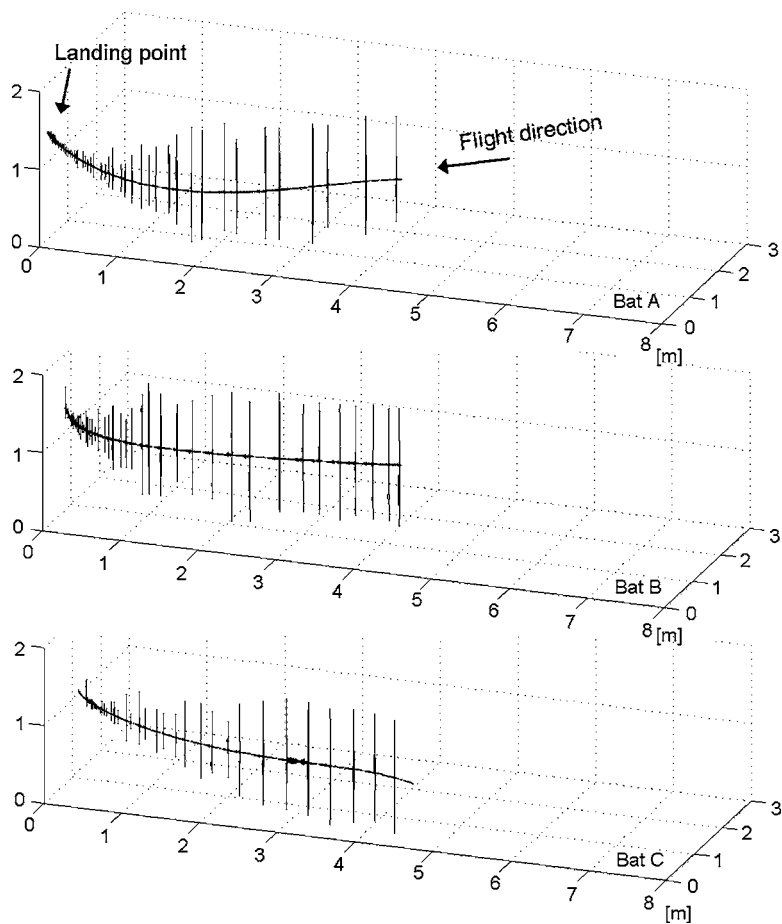


FIG. 4. Three-dimensional spatiotemporal reconstructions of echolocation behavior during landing flights for three bats. Coordinate grids (1 m^2) show the dimensions of the flight chamber ($8 \times 3 \times 2 \text{ m}$). Echolocation pulses recorded by the Telemike are placed alongside the flight trajectory. The arrow indicates the flight direction of the bat. The amplitude of emitted pulses decreased as the bat approached the target wall.

observed beyond a target distance of approximately 2 m.

Flight trajectories for the three bats during landing approaches are illustrated in three-dimensional images in Fig. 4. The echolocation pulses recorded with the Telemike are superimposed on the flight trajectory to indicate the spatiotemporal characteristics of echolocation behavior. The maximum flight speed for a direct approach to the target wall was approximately 4 m/s at 3–4 m from the target wall. Several common features were observed in the behavior of the three bats, including a marked decrease in interpulse interval and pulse amplitude as bats approached the target wall (approach phase), which usually started at a target distance of between 1 and 2 m. The pulse emission rate increased from 10–20 pulses per second at 3–4 m from the target wall to 130–140 pulses per second immediately prior to landing. The SPL of each pulse was calculated from the peak-to-peak amplitude voltage of the observed pulse in the time domain by the Telemike. The maximum SPL (peak-to-peak) at the microphone above the bat's head was approximately 130 dB during flight, which was almost 20 dB higher than when the bat was at rest. Before the bats started the approach phase, the sound pressure level of emitted pulse was almost constant at 130 dB peak-to-peak (search phase).

B. Changes in sound intensity of the pulse-echo pair

Figure 5(A) shows changes in intensity in the peak energy portions of the pulse and echo as a function of target distance for the three bats, normalized to the average of the

pulse intensity during the search phase. Data were taken from three flight sessions of each bat. Intensities of all echoes that could be extracted from the displayed sonograms are cumulatively plotted with open circles in Fig. 5(A).

The pulse intensity in bats intending to land started to decrease at a distance of between 1 and 2 m, and the reduction was approximately 30 dB before landing. On the other hand, the observed echo intensities were mainly between 40 to 50 dB weaker than the pulse but remained almost constant as the target distance decreased [open circles in Fig. 5(A)]. The three bats decreased the intensity of their emitted pulses logarithmically with target distance [Fig. 5(B)]. The reduction rate within 2 m of the target distance was 5.6–7.5 dB per halving of the target distance for the three bats. On average, the reduction rate of pulse intensity was 6.5 dB (21.6 dB decrease per decade of the target distance).

Figure 5(C) shows the distribution of all observed echo intensities relative to the average of the pulse intensity during the search phase for nine total flight sessions of three bats. The distribution showed a single peak, with a mean of $-42.6 \pm 5.5 \text{ dB}$, which corresponds to approximately 80–90 dB SPL (peak-to-peak). Variation in observed echo intensity may have been due to variation in the distance or direction of the object from which the echo returned, and/or the direction of the emitted pulse.

The change in intensity of a pulse-echo pair for one flight session by a bat is shown in Fig. 6. The intensity of the echo returning from the target wall to the bat (asterisks in

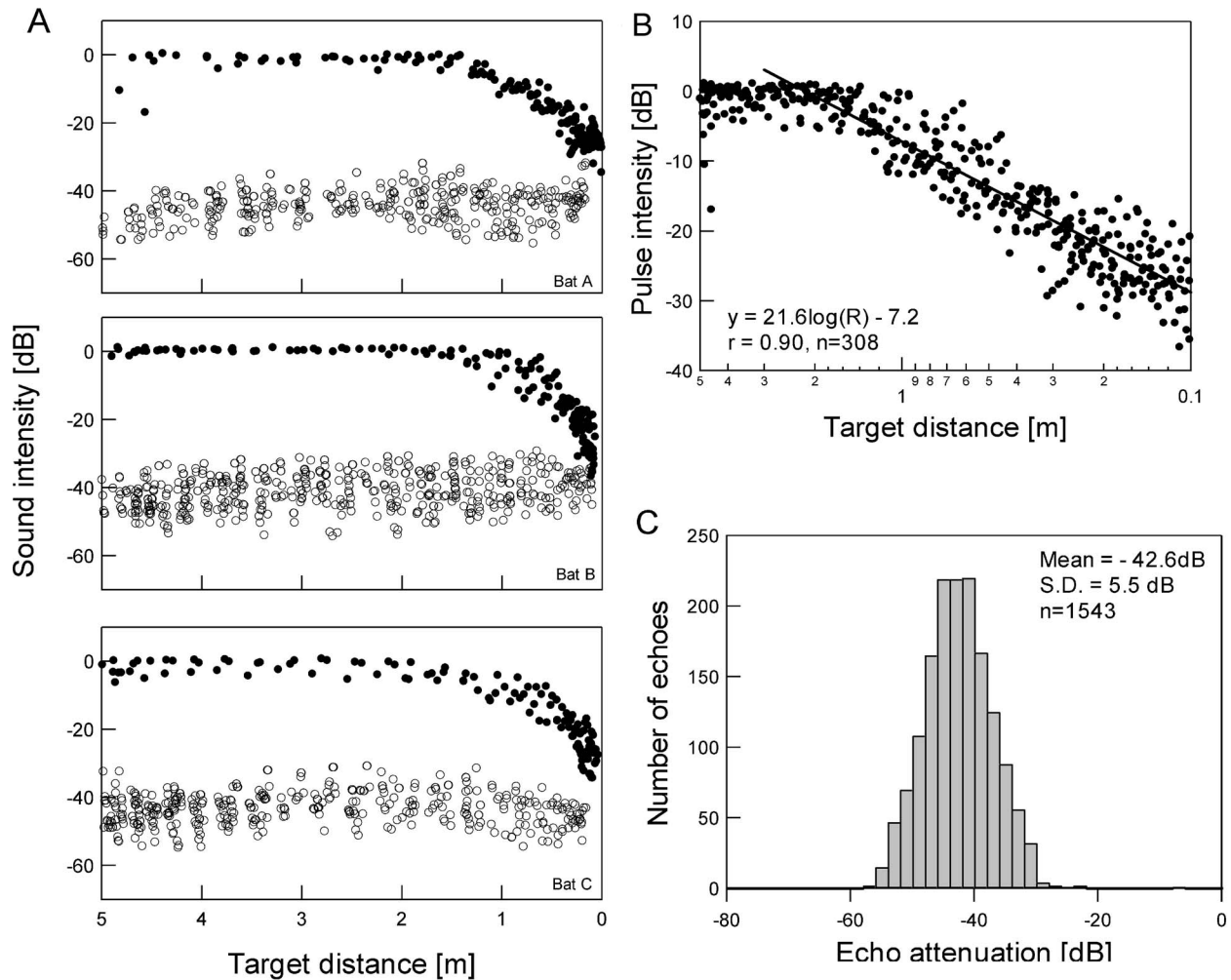


FIG. 5. Changes in intensity in the peak energy portion (the sonogram exhibited a peak in energy at about 50–60 kHz during flight, which was approximately 10 kHz higher than the terminal frequency) of the pulse-echo pairs, normalized to the average of intensity of the pulse emitted during the search phase. (A) Sound intensity of the pulse (solid circles) and echo (open circles) as a function of target distance for three bats. Data were taken from three flight sessions of each of three bats. (B) Decrease in pulse intensity for a total of nine flights of three bats. The solid line indicates the correlation line. Pulse intensity decreased by 6.5 dB on halving the target distance (21.6 dB decrease per decade of the target distance). (C) Distribution of observed echo intensity (relative to the average of the pulse intensity during the search phase) during the landing approach for nine flight sessions of three bats.

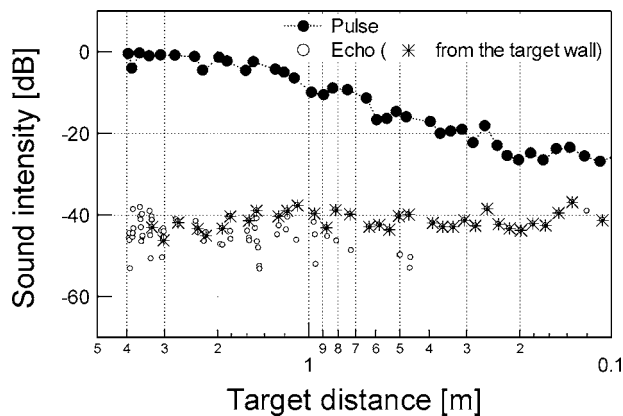


FIG. 6. Sound intensity of pulse-echo pairs as a function of target distance for one flight session by Bat A. Asterisks indicate echoes from the target wall, determined from sound and three-dimensional coordinate data (see Methods). The intensity of the echo returning from the target wall indicates a nearly constant intensity, while the bat considerably decreased the pulse intensity as it approached the target wall.

Fig. 6) indicates a nearly constant intensity, while the bat considerably decreased the pulse intensity as it approached the target wall.

IV. DISCUSSION

A. Echo-intensity compensation during flight

Bats are thought to adjust their pulse intensity in response to the intensity of the returning echoes that they attend to during echolocation. In this experiment, bats were expected to attend to the echo from the target wall. The echo intensity from the target wall (asterisks in Fig. 6) recorded directly by the Telemike was maintained at a constant level within a target distance of 2 m, whereas the pulse intensity considerably decreased. This indicates that the bat adjusted its pulse intensity depending on target distance such that the intensity of the echo returning from its destination was constant. In other words, this finding suggests that the echo from the target (destination) could be a prominent stimulus to adjustment of pulse intensity by bats and that a bat could rec-

ognize an echo from its target (the target wall) among a number of echoes reaching it [e.g., Fig. 2(A)].

A number of studies have attempted to measure pulse intensity for bats during flight (e.g., Griffin, 1958; Jen and Kamada, 1982; Vogler and Neuweiler, 1983; Kick and Simmons, 1984; Hartley *et al.*, 1989; Waters and Jones, 1995; Tian and Schnitzler, 1997; Boonman and Jones, 2002; Holdried and Helvesen, 2003). These studies reported a decrease in pulse intensity as a bat approached its target, using traditional fixed microphones that did not interfere with the flight path of the bat. In several studies, the reduction rate of pulse intensity as a function of distance to the target has been quantitatively estimated (e.g., Hartley *et al.*, 1989; Waters and Jones, 1995; Tian and Schnitzler, 1997; Boonman and Jones, 2002). For example, Tian and Schnitzler (1997) estimated the intensity of emitted pulses in *Rhinolophus ferrumequinum* during a landing approach, applying careful corrections to directional sounds recorded by the fixed microphone. Similarly, *Noctilio leporinus* decreased its pulse intensity while capturing a small target (mealworm) at a rate of about 6 dB on halving the target distance (Hartley, 1989). Although the reduction of emitted pulses by flying bats approaching a target varied slightly among the studies (e.g., Hartley *et al.*, 1989; Waters and Jones, 1995; Boonman and Jones, 2002), these studies suggest that the bats are supposed to adjust their pulse intensity in response to the intensity of returning echoes.

We found that *P. abramus* decreased their pulse intensity at a rate of 6.5 dB on halving the target distance and exhibited nearly constant echo intensity from the target wall, which should be attended to by a bat on its landing flight. This finding provides direct evidence of echo-intensity compensation by echolocating bats during flight to maintain the intensity of the resulting echo within the range necessary for optimal signal processing. The leaf-nosed bat *Hipposideros terasensis*, which emits a CF-FM pulse, decreased its pulse intensity by approximately 6 dB on halving the target distance using the same landing flight task as used here (Hiryu, 2005). Recently, echo-intensity compensation was demonstrated for the echolocation system in dolphins (Au and Benoit-Bird, 2003). Free-ranging dolphins decreased the amplitude of their echolocation signals as they approached the target by 6 dB per halving of distance. These results suggest that adjusting pulse intensity may be a common strategy of biosonar animals approaching a target.

The systems employed to compensate for variation in intensity and frequency of echoes are unique to biosonar animals. Extant artificial sonar systems are generally designed to emit a pulse with a fixed frequency to measure the Doppler shift in the returning echo in order to detect the velocity of the target. In contrast, CF-FM bats adjust their call frequency to maintain an echo frequency within the range they can hear best (Schnitzler, 1968; Suga, 1984). Consequently, the auditory system in the mustached bats can detect a shift in echo frequency with a high degree of accuracy, as small as 50 Hz which corresponds to less than 0.1% of the CF₂ frequency of the pulse emitted by the mustached bats (~61 kHz) (Suga, 1984; Riquimaroux *et al.*, 1991). Echo-intensity compensation allows bats to receive echoes within

a certain optimal intensity range, which may facilitate consistent and precise analysis of target information by the auditory system as seen in DSC (Kick and Simmons, 1984). These compensation behaviors associated with adjustments of pulse frequency and intensity may be a fundamental strategy employed by the animals for streamlining their echolocations, and will contribute to inspire the design of future artificial sonar systems or echo-sensing devices.

B. Multiple echoes reaching a bat during flight

We observed that a number of echoes from surrounding targets reached the flying bats with different echo delays and different intensities after a pulse emission (Fig. 2). When a bat was flying, the frequency of echoes received was Doppler-shifted depending on flight speed. A frequency sweep was observed with a slight difference among echoes from different targets [Fig. 2(A)], which may be due to differences in the extent of the Doppler shift according to the direction or relative velocity of the targets in relation to the flight path of the bat. We suggest that the difference in the frequency slope between returning echoes may provide FM bats in flight with direction and velocity information on multiple targets.

What specific behavioral strategies do echolocating bats use for processing multiple echoes? Bats are supposed to adjust their call parameters by feedback control in response to information from previous echoes. Analysis of the emitted pulse frequency and interpulse interval shows that bats may change the focus of their attention between echoes during flight (Hiryu *et al.*, 2005). By periodically focusing on different echoes, bats may process multiple auditory streams to perceive several targets in real time. In this flight experiment, the bats occasionally approached the wall attached with a landing mesh (referred to as the target wall in case of the landing scenario) and then returned to the starting point without landing. Interestingly, the bats during the U-turn did not decrease the intensity of emitted pulses as approached that wall, as seen when they were landing (Fig. 7). In addition, the observed intensity of echoes from that wall did not appear constant, showing a decrease of approximately 20 dB within a distance of 1 m before U-turn point [marked with U in Fig. 7(B)]. This suggests that the bats making a U-turn did not keep their attention on that wall as seen in case of landing scenario. It is likely that a bat may divert its attention to other targets depending on its flight direction in order to avoid colliding with surrounding walls while making a U-turn.

In this study, we focused on temporal changes in the sound intensity of pulse-echo pairs in the context of echo-intensity compensation. Further investigation of pulse-echo pairs to which a bat actually listens may lead to other adjustment mechanisms in echolocation systems for multiple targets in the immediate surroundings.

ACKNOWLEDGMENTS

This work was partly supported by a grant to the Research Center for Advanced Science and Technology (RCAST) at Doshisha University from the Ministry of Edu-

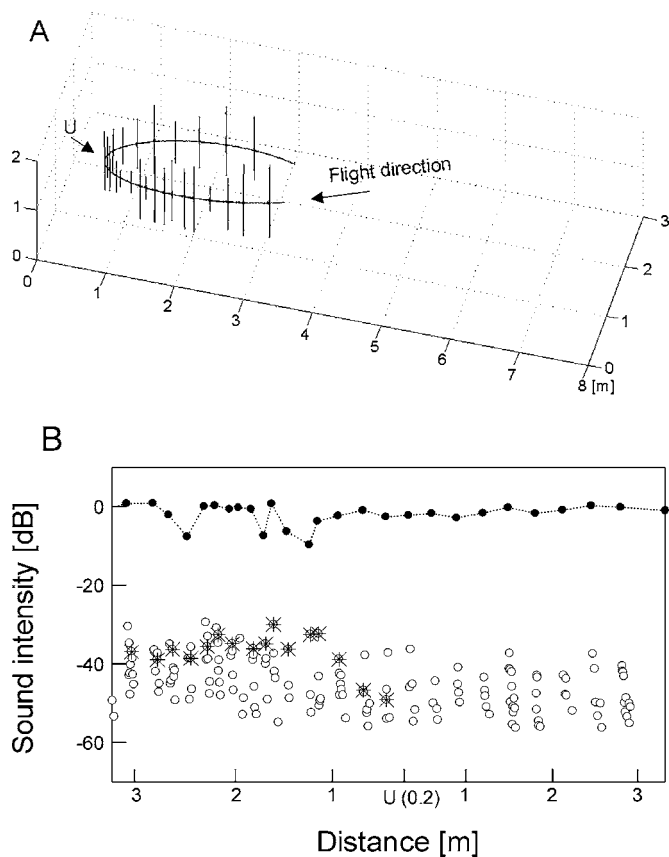


FIG. 7. Three-dimensional spatiotemporal reconstruction for a U-turn flight (A), and sound intensity of the pulse (solid circles) and echo (open circles) as a function of the distance to the wall with a landing mesh (referred to as “target wall” in case of the landing flight) while a bat making U-turn (B). Asterisk indicates the echo from that wall. The bat made a U-turn at a distance of 0.2 m (marked U) from that wall. The bat during the U-turn did not decrease the intensity of emitted pulses as approached that wall, as seen when they were landing.

cation, Culture, Sports, Science, and Technology (MEXT) of Japan: Special Research Grants for the Development of Characteristic Education from the Promotion and Mutual Aid Corporation for Private Schools of Japan and the Innovative Cluster Creation Project.

Au, W. W. L., and Benoit-Bird, K. J. (2003). “Automatic gain control in the echolocation system of dolphins,” *Nature (London)* **423**, 861–863.

Boonman, A., and Jones, G. (2002). “Intensity control during target approach in echolocating bats; stereotypical sensori-motor behaviour in Daubenton’s bats, *Myotis daubentonii*,” *J. Exp. Biol.* **205**, 2865–2874.

Gaioni, S. J., Riquimaroux, H., and Suga, N. (1990). “Biosonar behavior of mustached bats swung on a pendulum prior to cortical ablation,” *J. Neurophysiol.* **64**, 1801–1817.

Ghose, K., and Moss, C. F. (2003). “The sonar beam pattern of a flying bat as it tracks tethered insects,” *J. Acoust. Soc. Am.* **114**, 1120–1131.

Griffin, D. R. (1958). *Listening in the Dark* (Yale University, New Haven, CT).

Gustafson, Y., and Schnitzler, H. U. (1979). “Echolocation and obstacle avoidance in the hipposiderid bat *Assellia tridens*,” *J. Comp. Physiol. [A]* **131**, 161–167.

Hartley, D. J. (1992a). “Stabilization of perceived echo amplitudes in echolocating bats. I. Echo detection and automatic gain control in the big brown bat, *Eptesicus fuscus*, and the fishing bat, *Noctilio leporinus*,” *J. Acoust. Soc. Am.* **91**, 1120–1132.

Hartley, D. J. (1992b). “Stabilization of perceived echo amplitudes in

echolocating bats. II. The acoustic behavior of the big brown bat, *Eptesicus fuscus*, when tracking moving prey,” *J. Acoust. Soc. Am.* **91**, 1133–1149.

Hartley, D. J., Campbell, K. A., and Suthers, R. A. (1989). “The acoustic behavior of the fish-catching bat, *Noctilio leporinus*, during pre capture,” *J. Acoust. Soc. Am.* **86**, 8–27.

Henson, O. W., Jr., Bishop, A. L., Keating, A. W., Kobler, J. B., Henson, M. M., Wilson, B. S., and Hansen, R. (1987). “Bisonar imaging of insects by *Pteronotus p. parnellii*, the mustached bat,” *Natl. Geograph. Res.* **3**, 82–101.

Hiryu, S. (2005). “A study on biosonar engineering based on spatio-temporal measurement of bat’s behavior,” Ph.D dissertation, Doshisha University, Kyotanabe, Japan.

Hiryu, S., Katsura, K., Lin, L. K., Riquimaroux, H., and Watanabe, Y. (2005). “Doppler-shift compensation in the Taiwanese leaf-nosed bat (*Hipposideros terasensis*) recorded with a telemetry microphone system during flight,” *J. Acoust. Soc. Am.* **118**, 3927–3933.

Holderied, M. W., and Helversen, O. V. (2003). “Echolocation range and wingbeat period match in aerial-hawking bats,” *Proc. R. Soc. London, Ser. B* **270**, 2293–2299.

Hope, G. M., and Bhatnagar, K. P. (1979). “Electrical response of bat retina to spectral stimulation: comparison of four microchiropteran species,” *Experientia* **35**, 1189–1191.

Jen, P. H., and Kamada, T. (1982). “Analysis of orientation signals emitted by the CF-FM bat, *Pteronotus p. parnellii* and the FM bat, *Eptesicus fuscus* during avoidance of moving and stationary obstacles,” *J. Comp. Physiol. [A]* **148**, 389–398.

Kalko, E. (1995). “Insect pursuit, prey capture and echolocation in pipistrelle bats (Microchiroptera),” *Anim. Behav.* **50**, 861–880.

Keating, A. W., Henson, O. W., Jr., Henson, M. M., Lancaster, W. C., and Xie, D. H. (1994). “Doppler-shift compensation by the mustached bat: quantitative data,” *J. Exp. Biol.* **188**, 115–129.

Kick, S. A., and Simmons, J. A. (1984). “Automatic gain control in the bat’s sonar receiver and the neuroethology of echolocation,” *J. Neurosci.* **4**, 2725–2737.

Kobler, J. B., Wilson, B. S., Henson, O. W., Jr., and Bishop, A. L. (1985). “Echo intensity compensation by echolocating bats,” *Hear. Res.* **20**, 99–108.

Lancaster, W. C., Keating, A. W., and Henson, O. W., Jr. (1992). “Ultrasonic vocalizations of flying bats monitored by radiotelemetry,” *J. Exp. Biol.* **173**, 43–58.

Neuweiler, G. (1984). “Foraging, echolocation and audition in bats,” *Naturwiss.* **71**, 446–455.

Riquimaroux, H., Gaioni, S. J., and Suga, N. (1991). “Cortical computational maps control auditory perception,” *Science* **251**, 565–568.

Riquimaroux, H., and Watanabe, Y. (2000). “Characteristics of bat sonar sounds recorded by a telemetry system and a fixed ground microphone,” *The seventh western pacific regional acoustics conference (WEST-PRACVII)*, 233–238.

Schnitzler, H. U. (1968). “Die Ultraschallortungslaute der Hufeisen-Fledermause (*Chiroptera-Rhinolophidae*) in verschiedenen Orientierungssituationen [The Ultrasonic sounds of horseshoe bats (*Chiroptera-Rhinolophidae*) in different orientation situations],” *Zeitschrift für Vergleichende Physiologie* **57**, 376–408.

Schnitzler, H. U., Kalko, E., Miller, L., and Surlykke, A. (1987). “The echolocation and hunting behavior of the bat, *Pipistrellus kuhli*,” *J. Comp. Physiol. [A]* **161**, 267–274.

Schuller, G., Beuter, K., and Schnitzler, H. U. (1974). “Response to frequency shifted artificial echoes in the bat *Rhinolophus ferrumequinum*,” *J. Comp. Physiol. [A]* **89**, 275–286.

Simmons, J. A. (1974). “Response of the Doppler echolocation system in the bat, *Rhinolophus ferrumequinum*,” *J. Acoust. Soc. Am.* **56**, 672–682.

Simmons, J. A., Moffat, A. J., and Masters, W. M. (1992). “Sonar gain control and echo detection thresholds in the echolocating bat, *Eptesicus fuscus*,” *J. Acoust. Soc. Am.* **91**, 1150–1163.

Simmons, J. A., and Stein, R. A. (1980). “Acoustic imaging in bat sonar: echolocation signals and the evolution of echolocation,” *J. Comp. Physiol. [A]* **135**, 61–84.

Suga, N. (1984). “The extent to which biosonar information is represented in the bat auditory cortex,” in *Dynamic aspects of neocortical function*,

edited by G. M. Edelman, W. E. Gall, and W. E. Cowan (Wiley, New York), pp. 315–373.

Tian, B., and Schnitzler, H. U. (1997). “Echolocation signals of the greater horseshoe bat (*Rhinolophus ferrumequinum*) in transfer flight and during landing,” *J. Acoust. Soc. Am.* **101**, 2347–2364.

Trappe, M., and Schnitzler, H. U. (1982). “Doppler-shift compensation in

insect-catching horseshoe bats,” *Naturwiss.* **69**, 193–194.

Vogler, B., and Neuweiler, G. (1983). “Echolocation in the noctule (*Nyctalus noctula*) and horseshoe bat (*Rhinolophus ferrumequinum*),” *J. Comp. Physiol. [A]*, **152**, 421–432.

Waters, E., and Jones, G. (1995). “Echolocation call structure and intensity in five species of insectivorous bats,” *J. Exp. Biol.* **198**, 475–489.

Nonlinear acoustics in the pant hoots of common chimpanzees (*Pan troglodytes*): Vocalizing at the edge

Tobias Riede^{a)}

National Center for Voice and Speech at the Denver Center for the Performing Arts, and College of Liberal Arts and Sciences, Health, and Behavioral Sciences, University of Colorado at Denver, Colorado 80204

Adam Clark Arcadi

Department of Anthropology, Cornell University, Ithaca, New York 14853

Michael J. Owren

Department of Psychology, Language Research Center, and Center for Behavioral Neuroscience, Georgia State University, Atlanta, Georgia 30302

(Received 25 May 2006; revised 23 October 2006; accepted 24 November 2006)

Common chimpanzee (*Pan troglodytes*) “pant hoots” are multi-call events that build from quiet, consistently harmonic introductory sounds to loud, screamlike “climax” calls with acoustic irregularities known as “nonlinear phenomena” (NLP). Two possible functions of NLP in climax calls are to increase direct auditory impact on listeners and to signal physical condition. These possibilities were addressed by comparing climax calls from 12 wild chimpanzee males with “screams” and pant hoot “introduction” calls from the same individuals. Climax calls that included NLP were found to have higher fundamental frequencies (F0s) than introduction or climax calls that were purely harmonic. NLP onsets within climax calls were also specifically associated with local F0 maxima, suggesting vocalizers are vibrating their vocal folds at the upper limits of stability. Furthermore, climax calls showed far fewer NLP than did screams recorded from the same individuals, while showing equivalent or higher F0 values. Overall, the results are consistent with the hypothesis that the relative prevalence of NLP is a signal of physical condition, with callers “vocalizing at the edge” of regular, stable production while producing few NLP. The results are discussed in light of the initial hypotheses. © 2007 Acoustical Society of America.

[DOI: 10.1121/1.2427115]

PACS number(s): 43.80.Ka, 43.80.Nd, 43.80.Lb [JAS]

Pages: 1758–1767

I. INTRODUCTION

The common chimpanzee (*Pan troglodytes*) “pant hoot” is a distinctive vocalization [Fig. 1(a)] consisting of up to four acoustically and temporally separable sections, called the introduction, build-up, climax, and let-down (e.g., Marler and Hobbett, 1975). The function of pant hoots has not been easy to determine because the signal is produced in many contexts, both within and between groups. Within groups, two suggested functions are attracting conspecifics (e.g., allies and/or mates: Reynolds and Reynolds, 1965; Goodall, 1986; Mitani and Nishida, 1993) and displaying status and/or physical condition (Clark and Wrangham, 1993; Riede *et al.*, 2004). These possibilities are not mutually exclusive, as both high status and physical condition are likely to be important in attracting alliance or mating partners. In the current work, we evaluate these hypotheses in light of the acoustic properties of the calls themselves.

The basic vocal production process underlying pant hoots involves regular, synchronized vocal-fold vibration. Individual calls typically show a prominent fundamental fre-

quency (F0), corresponding to the basic rate of that vibration, as well as harmonically structured energy at higher frequencies. However, the call amplitudes and F0 values involved can vary significantly both within and across the four sections of a pant hoot. For example, introduction and build-up vocalizations are usually relatively quiet and low-pitched compared to climax calls, which are often extremely loud sounds with very high F0s. In earlier work, we demonstrated that in addition to harmonic components, pant hoots also routinely include irregularities known as “nonlinear phenomena” (NLP) (Riede *et al.*, 2004). Since the dynamics of the larynx as a whole are nonlinear, any and all sounds produced represent nonlinear phenomena, including purely harmonic vocalizations. However, for convenience, we will contrast “harmonic” versus “nonlinear” components of calls. Although not previously investigated in pant hoots, nonlinear phenomena are a normative feature of mammalian vocalizations and are traceable to the nonlinear nature of vocal-fold dynamics (e.g., Herzel *et al.*, 1994; Wilden *et al.*, 1998; Riede *et al.* 1997, 2000).

Four standard nonlinear phenomena were identified in the pant hoots, which were frequency jumps, subharmonics, deterministic chaos, and biphonation (illustrated in Fig. 2). As discussed by Riede *et al.* (2004), frequency jumps refer to sudden F0 changes in which vibration rate moves up or

^{a)} Author to whom correspondence should be addressed. National Center for Voice and Speech, 1101 13th Street, Denver, CO 80204. Electronic mail: tobiasriede@web.de

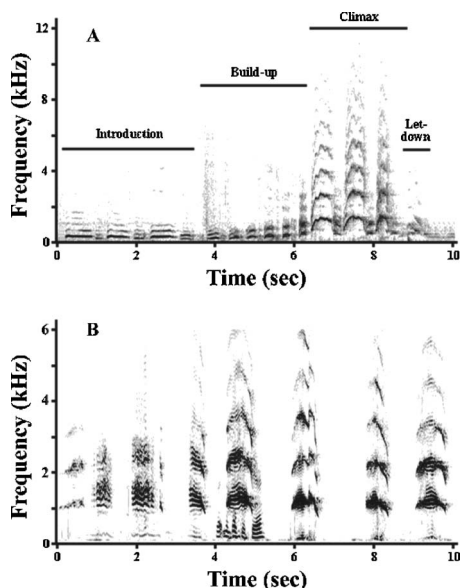


FIG. 1. (a) A narrow-band FFT-based spectrogram of a pant hoot by adult male chimpanzee AJ (44.1-kHz sampling rate, 512-point Hanning window). It includes three introduction calls, six built-up calls, three climax calls, and one let-down call. (b) A narrow-band FFT-based spectrogram of a scream bout by adult male chimpanzee SY (44.1-kHz sampling rate, 512-point Hanning window), including eight calls. For comparison, low-frequency components occurring between 4 and 5 s are speech produced in the background by an adult male human.

down abruptly and discontinuously. Subharmonics are additional spectral components that can suddenly appear at integer fractional values of an identifiable F_0 and as harmonics of these values. Deterministic chaos refers to episodes of nonrandom noise, which is produced through chaotic vocal fold vibration rather than through stochastic turbulence. Finally, biphonation refers to the occurrence of two simultaneous but independent F_0 's, which can be visible in a spectrogram as two distinct and autonomous frequency contours or as sidebands adjacent to harmonics that are associated with cyclic amplitude fluctuations in the time-series waveform. Neither deterministic chaos nor biphonation of the latter sort can be unequivocally identified by inspection of spectrograms alone, but were inferred in these calls based on evidence from a combination of mathematical modeling, human vocal production, experiments with excised larynges from nonhuman mammals, and in-depth analyses of similar vocalizations in monkeys, piglets, and dogs [see Riede *et al.* (2004), for further discussion].

Riede *et al.* (2004) reviewed three hypothesized functions for the occurrence of nonlinear phenomena that may be applicable to chimpanzee pant hoots. First, NLP may increase the individual's vocal distinctiveness (e.g., Wilden *et*

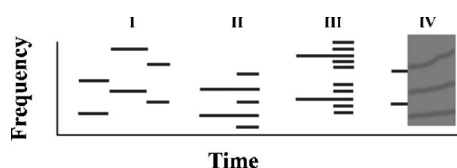


FIG. 2. Schematic narrow-band spectrograms illustrating frequency jumps (I), subharmonics (II), biphonation (III), and deterministic chaos (IV). See text for further details.

al., 1998; Volodina *et al.*, 2006), which could increase the signaling effectiveness of pant hoots regardless of what other functions are involved. However, this interpretation was deemed unlikely given the nature of the NLP found to be most prevalent in pant hoots (Riede *et al.*, 2004). Second, NLP may function to increase the auditory impact of calls (Owren, 2003; Owren and Rendall, 2001), which in pant hoots could be useful both for attracting allies or mates, as well as for signaling status or physical condition. If so, Riede *et al.* (2004) suggested that vocalizers should maximize the number of nonlinear phenomena occurring in their pant hoots in order to increase the impact of the sounds. Finally, the prevalence of nonlinear events could function as a cue of vocalizer health and physical condition, if the prevalence of NLP is inversely related to vocal-fold integrity. The condition of the vocal folds can be affected by systemic effects of disease or trauma (Riede *et al.*, 1997; Riede and Stolle-Malorny, 1999). By minimizing NLP and producing harmonically structured sounds at the highest possible frequencies and amplitudes, callers would signal physical vigor.

According to this third hypothesis, chimpanzee pant hoot climax calls could represent vocal-fold vibration being pushed to the limits of stability, thereby providing audible cues to the integrity and performance of the production system. Consistent with this "vocalizing at the edge" interpretation, the overall structure of pant hoots is intriguingly reminiscent of the "voice range profile" (VRP) test (Fig. 3) used to measure the condition of human voices (Seidner *et al.*, 1985; Awan, 1991; Akerlund *et al.*, 1992; Titze, 1992; Coleman, 1993). In this test, the vocalizer is asked to maintain a clear, even tone while singing or speaking both as softly and as loudly as possible across a range of the lowest to highest F_0 s that can be produced. In effect, the singer or talker attempts to maintain regular, synchronized vocal-fold vibration across the widest possible range of production conditions, and outcomes are displayed as a range of sound pressure levels (dB SPL) that can be reached at each F_0 value.

Pant hoots resemble a VRP in beginning with quiet, lower-frequency introduction calls and then building to a climax of much more extreme frequency and amplitude levels. Pabon and Plomp (1988) provide evidence that phonatory instability increases near the upper boundary of the VRP. In humans, access to vocal training can improve VRP outcomes (Heylen *et al.*, 2002), but the test remains sensitive to physical compromises brought on by conditions such as improper hydration (Soloman and Dimattia, 2000), fatigue (Welham and Maclagan, 2003), aging (Heylen *et al.*, 1998; Hacki and Heitmüller, 1999; Verdonck-de Leeuw and Mahieu, 2004), and disease (Ikeda *et al.*, 1999).

The current work had two goals. The first was to confirm whether the occurrence of nonlinearities in pant hoot climax calls is in fact specifically associated with elevated vocal effort. This possibility was examined at a global level by contrasting F_0 measures for pant hoot climax and introduction calls drawn from the same pant hoot bouts examined by Riede *et al.* (2004). Evidence of production extremes was also sought locally within calls. Our reasoning was that if irregularity is tied to vocal production conditions, then the onset of NLP should be reliably preceded by local maxima in

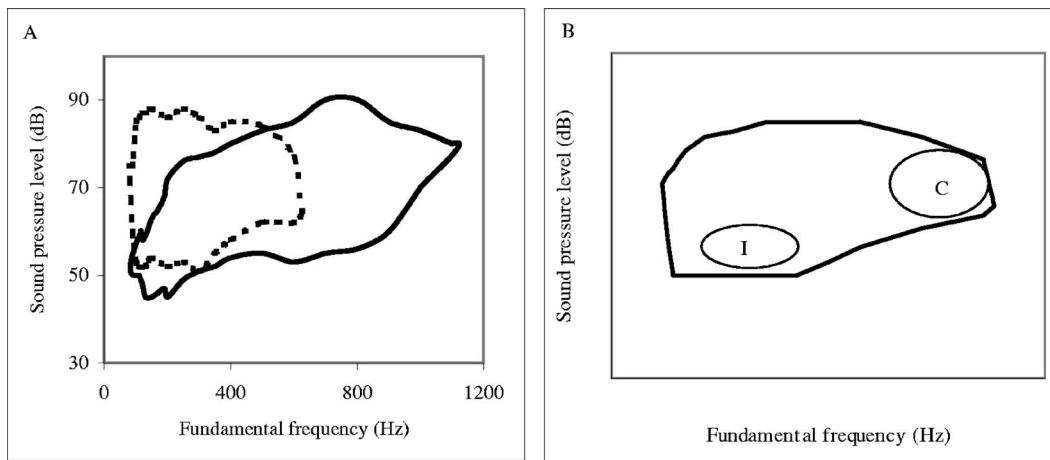


FIG. 3. (a) Two examples of voice range profiles from a human female (58 years) (solid line) and a human male (26 years) (dotted line). The encircled areas indicate the minimum and maximum sound pressure level values of a harmonic voice as a function of F0 [data after Pabon and Plomp (1988)]. F0 typically ranges from approximately 70 to 500 Hz for adult human males who have not had vocal training. (b) Schematic of a hypothetical chimpanzee voice range profile with the range of introduction calls (I) and climax calls (C) indicated. F0 values vary widely over the course of pant hoots from adult male chimpanzees, ranging from approximately 500 Hz in introductory calls to approximately 1100 Hz in climax calls. Corresponding amplitude ranges are not yet known. Our data suggest that climax calls are produced at the upper edge of an individual's F0 range.

critical production parameters (here F0), while NLP offset should be followed by decreases in the same parameters.

The second goal of the work was then to distinguish between the “auditory impact” and “vocal stability” hypotheses outlined earlier. Here, the approach was to contrast NLP occurring in pant hoot climax calls and screams. Screams were specifically of interest since they are also high-effort vocalizations with prominent acoustic irregularities. Preliminary comparisons showed screams to be a good match to climax calls in both duration and F0 [Fig. 1(b)], while other likely effortful calls such as “waas” and “barks” were not (Marler and Tenaza, 1977). Climax calls have in fact long been explicitly described as being similar to screams (e.g., Marler and Hobbett, 1975). Noisiness in screams produced by macaques has specifically been shown to be deterministic chaos rather than stochastic turbulence (Tokuda *et al.*, 2002). Thus, if the function of pant hoots is to have the greatest possible auditory impact on listeners, we expected that both F0 values and NLP rates would be maximized, and therefore comparable to or greater than those found in screams. On the other hand, if climax calls are primarily a vehicle for displaying vocal stability, climax calls should be marked by relatively low NLP rates compared to screams, while showing comparable or higher levels of vocal effort.

Overall, the results provided strong evidence that NLPs are associated with elevated F0 values, and less definitive but nonetheless clear evidence that by minimizing the occurrence NLP callers may signal physical condition via vocal stability. Follow-up comparisons did not reveal correlations between acoustic measures and either vocalizer age or rank, but pant hoots from two alpha males offer suggestive evidence of a link between maximum F0 values and dominance status.

II. METHODS

A. Call samples

Vocalizations produced by 12 males from the Kanjawara study group in the Kibale National Park (Uganda)

were recorded during two field seasons (1988–1989, 1996–1997; see also Riede *et al.*, 2004). Recordings were made using Sennheiser directional microphones (ME80 head, K3U power module) and windscreens, CrO2 tapes, and either a Sony Pro Walkman or Marantz PMD430 portable cassette deck. Two call samples were analyzed. The first was the same sample of 411 pant hoots screened for nonlinear phenomena by Riede *et al.* (2004), which was reanalyzed for the current work. This sample represented a subset of the calls identified as pant hoots in the field, which had been screened for defining acoustical and contextual features based on spectrograms, field notes, and tape-recorded commentary accompanying each call (Marler and Hobbett, 1975; Marler and Tenaza, 1977). The pant hoot sample included 268 introduction and 1052 climax calls [see Table 1 in Riede *et al.* (2004) for further details].

The second sample consisted of screams from the same vocalizers that had produced the pant hoots. A set of 60 scream bouts could be culled from the project recordings, representing 10 of the same 12 adult males. Screams were identified according to a combination of Goodall's (1986) perceptual description and the behavioral contexts coded at the time of audio recording. These calls had been identified as screams before the current project was actually conceived, based on a combination of auditory characteristics and context, and blind with respect to the goals of this project. Screams were vocalizations that had been produced by vocalizers showing submissive behavior and evident fear, were recorded at the same research site over the same time periods, and were classified using evidence from spectrograms, field notes, and recorded commentary.

Calls in both samples were screened for the occurrence of extraneous sounds, for example due to the animals moving through heavy vegetation or other chimpanzees vocalizing in the background. These kinds of events were common, as both pant hoots and screams are routinely associated with movement, are often produced while others are vocalizing, and can themselves trigger vocal responses. The extraneous

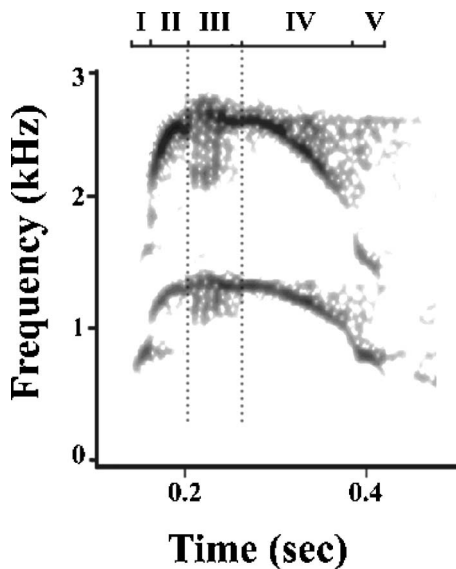


FIG. 4. A narrow-band FFT-based spectrogram of a climax call indicating a total of four segments (I, II, III, IV, V), including two frequency jumps between segments I and II, and between segments IV and V, one biphonation event (segment III). Note that fundamental frequency is increasing before the biphonation (segment II) and decreasing after the biphonation (segment IV).

sounds did not typically interfere with marking the beginnings and ends of bouts, calls, and segments, or assigning the sounds to categories. However, they could interfere with more detailed analyses. Sample sizes therefore vary among the analyses that follow, often depending on whether the level of comparison concerns entire calls, segments of calls, or shorter components within call segments.

B. Acoustic analysis

All calls were digitized at a 44.1-kHz sampling rate and 16-bit accuracy using CoolEdit 2000 software (Syntrillium Software, Scottsdale, AZ). After downsampling to 16 kHz, the files were analyzed primarily using ESPS/waves+5.3 “xwaves” software (Entropic Research Lab, Washington, DC), with additional analyses conducted using the Praat program (Boersma and Weeninck, 2005).

1. Segmenting calls

The pant hoot analyses conducted by Riede *et al.* (2004) were based on first marking onsets and offsets of pant hoots as a whole, individual calls within the bout comprising a pant hoot, and individual segments within each call. That process is illustrated in Fig. 4, with a segment being an interval within a call that shows homogeneous acoustic structure. For the current work, all pant hoot segmentation was reexamined and adjusted as needed, and the previously unanalyzed scream bouts were also labeled at each level. Labeling was based on visual inspection of narrow-band spectrograms (16-kHz sampling rate, 512-point Hanning window) and associated Fourier frequency spectra. Calls were further labeled as being either “harmonic” (showing no evidence of nonlinear phenomena) or “nonlinear” (exhibiting both nonlinear and harmonic components, or being wholly nonlinear).

Individual segments received mutually exclusive labels of “harmonic,” “subharmonic,” “chaotic,” and “biphonic,” as described by Riede *et al.* (2004).

Call and segment labeling allowed automated computation of acoustic measures of interest at both levels of organization, with analyses focused on F0 characteristics. Labeling by type was a necessary prelude to F0 extraction, which was restricted to harmonic calls and harmonic segments within calls. F0 values in a nonlinear call thus refer to harmonic segments found within that sound. While the concept of a fundamental frequency of vibration can be applied to some NLP, such as biphonation, it is not applicable to others. Most importantly, deterministic chaos is by definition aperiodic, meaning that there is no F0 to extract. F0 analysis was therefore not conducted with nonlinear segments. In addition to characterizing overall F0 values in calls and segments, analyses also focused on the segments immediately preceding and following a nonlinear event. Here, each targeted segment was divided into ten equal intervals, and values were extracted separately over each of these “subsegments.”

2. Dependent variables

Acoustic variables involved mean F0 over a segment (F0Mean) or subsegment (F0), and maximum F0 over a segment (F0Max).

3. F0 measurement

F0 values in chimpanzee pant hoots and screams routinely exceeded 1000 Hz and were not always adequately handled by the pitch-extraction algorithm implemented in xwaves. Linear predictive coding [LPC (Markel and Gray, 1976)] and associated peak-picking algorithms were therefore used instead [using LPC for F0 extraction is reviewed by Owren and Bernacki (1998)]. Values were computed at 0.01-s intervals using a 0.049-s cosine window and 12 LPC coefficients, which yielded approximately 2-Hz frequency resolution. Resulting F0 tracks were inspected after being overlaid on corresponding narrow-band spectrograms (0.040-s Hanning window). When visual inspection revealed deviations from the first harmonic, tracks were recomputed with different peak-extraction settings. Persistent tracking problems occurred almost exclusively at the beginnings and ends of target intervals, in which case F0 extraction was restricted to portions of segments or calls in which tracking was successful.

4. Subsegment analysis

The goal of subsegment analysis was to test for possible changes in F0 values associated with the occurrence of nonlinear events. One kind of analysis compared measurements from the last subsegment before a bifurcation from a harmonic to a nonlinear segment to the mean value for the entire preceding segment. Analogous comparisons were conducted for harmonic segments that followed a nonlinear event, contrasting measurements from the first subsegment after the bifurcation to mean values over all ten subsegments. Finally, F0 values were correlated with subsegment position for segments preceding and following bifurcations, with the ques-

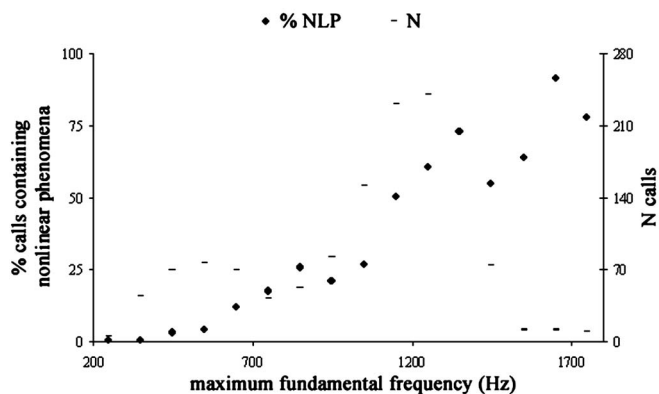


FIG. 5. Pant hoot calls of all individuals were categorized according to their maximum fundamental frequency in 100-Hz bins starting at 200 Hz. Each bin contains N calls. The percentage of calls in each bin that contains NLP increases with increasing maximum fundamental frequency.

tion of interest being whether the resulting regression line showed an overall slope different from 0. To help compensate for the lack of independence in making multiple measurements within each subsegment, a criterion of $P < 0.01$ was used when testing for statistical significance in these comparisons.

Sound amplitude, the second aspect of the voice range profile measured as the sound pressure level, was not considered in this study, given the difficulty of characterizing amplitude values of sounds recorded from free-ranging animals in an open acoustic environment. The amplitude of any field-recorded sound is necessarily affected by a number of sources of measurement error, including variation in the distance between source and microphone, in the relative orientations of source and microphone, in the effects of intervening vegetation and any other sound barriers, and in ambient background noise levels.

III. RESULTS

As noted earlier, the pant hoot sample yielded 295 introduction calls and 1052 climax calls. Of these, only 4 introduction calls were found to include NLP (1.3%), while 552 climax calls did so (55.2%) (Fig. 5) (Riede *et al.*, 2004). The 60 scream bouts included 600 total calls, of which 509 were nonlinear (84.8%).

A. Pant hoot acoustics

1. Introduction versus climax calls

Ignoring vocalizer identity, mean F0Max across all climax calls was 1170 Hz (SD=189 Hz) and was significantly higher than the mean of 595 Hz (SD=201) across all introduction calls, $t=45.7$, $P < 0.0001$. Comparing mean F0Max values for climax and introduction calls separately by individual allowed 6 matched comparisons in calls from the 1988 field season and 11 comparisons using calls from 1997 (that also included the same four individuals recorded earlier). F0Max values were again found to be higher in climax calls than in introduction calls, and the difference was statistically significant both when all 17 cases were included

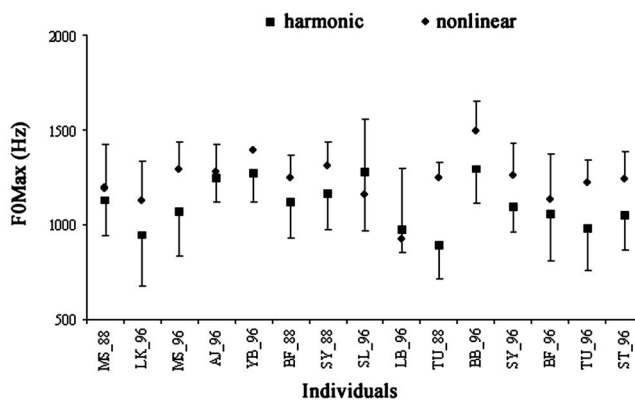


FIG. 6. Means and standard deviations of maximum fundamental frequency (F0Max) values from harmonic segments of climax calls that were either entirely harmonic or included one or more nonlinear events. Twelve different individual callers are labeled according to a two-letter identity code and the year of recording. The animals are ordered from youngest (MS_88, 11 years old) to oldest (ST_97, 41 years old).

(paired t -test, $t=24.6$, $P < 0.001$) and when only 1997 recordings were included (paired t -test, $t=15.9$, $P < 0.001$).

2. Harmonic versus nonlinear climax calls

Comparing F0Max values without regard to either caller identity or bout of origin showed a mean of 1127 Hz (SD=208) in the 500 climax calls that were purely harmonic, which was significantly lower than a mean value of 1240 Hz (SD=188) found in the 552 nonlinear climax calls ($t=-9.3$; $P < 0.001$). Limiting comparisons to a set of only the first harmonic and first nonlinear climax call from each bout produced a matched sample of 134 pairs from ten different callers. This sample also showed higher F0Max values in nonlinear calls (paired t -test, $t=-5.99$; $P < 0.001$). A total of 16 comparisons of mean F0Max outcomes could be made by individual caller over the two field seasons. These comparisons are illustrated in Fig. 6, with mean F0Max values again found to be higher in nonlinear than harmonic climax calls. The difference was statistically significant both when all 16 cases were included (paired t -test, $t=-4.55$, $P < 0.001$) and when the sample was restricted to the 12 unique callers recorded in 1997 (paired t -test, $t=-3.56$, $P < 0.01$).

3. Subsegment analyses

The mean F0 value of the last subsegment preceding bifurcation was 1134 Hz (SD=204) in 487 cases, which was significantly higher than the overall F0Mean value of 1044 Hz (SD=206) found in these segments (paired t -test, $t=26.7$, $P < 0.001$). Separate t -test comparisons were also conducted for each type of nonlinear event, which included 324 cases of biphonation, 71 cases of subharmonics, and 28 cases of deterministic chaos. In each instance, the F0 value of the final subsegment before bifurcation was significantly higher than the overall mean value for that segment. The mean slope of F0 contours in harmonic segments prior the bifurcation was 20.8 (19.5), which was significantly greater than 0 (one-tailed, paired t -test, $N=35$, $t=6.3$, $P < 0.001$). The mean slope of F0 in harmonic segments after a bifurca-

tion was -19.9 ($SD=27.9$), which was significantly lower than 0 (one-tailed, paired t -test, $N=36$, $t=-4.3$, $P<0.001$).

B. Screams versus climax calls

1. Proportion of NLP

As noted earlier, NLP were found in 85% of screams, compared to 55% of climax calls. Individual means from nine callers represented by both five or more climax calls and five or more screams showed an overall proportion of 75.1% of screams with one or more nonlinear events ($SD=22.1$, range=25% to 94%). This outcome was significantly higher than the corresponding value of 41.5% for climax calls ($SD=22.4$, range=3% to 96%, paired t -test, $t=3.3$, $P<0.01$). Nonlinear events were found to make up 84.8% ($SD=9.3$, $N=9$, range=66.2% to 98.1%) of the total duration of screams ($SD=9.3$, $N=9$, range=66.2% to 98.1%), which was significantly higher than the corresponding value of 29.5% ($SD=17.5$, $N=9$, range=19.6% to 61.7%) for nonlinear climax calls (paired t -test, $t=7.77$, $P<0.01$). The mean duration of nonlinear events relative to the total call duration was 85.5% across individual callers ($SD=8.9$, $N=12$, range=65% to 94%), which was also statistically different from the corresponding value of 43.0% for climax calls ($SD=8.0$, $N=10$, range=31% to 54%), comparing animals producing both call-types (paired t -test, $t=10.2$, $P<0.001$, $N=8$).

The overall proportion of each type of nonlinear event was generally similar in the two kinds of calls. As reported by Riede *et al.* (2004), the climax calls with nonlinear phenomena included 5% with frequency jumps, 13% with subharmonics, 6% with deterministic chaos, 68% with biphonation, and 8% with more than one type represented. For the screams, these proportions were 1% with frequency jumps, less than 1% with subharmonics, 21% with deterministic chaos, 66% with biphonation, and 12% with more than one kind of nonlinearity.

2. F0 comparisons

The mean F0Max value calculated without regard to caller identity was 1076 Hz ($SD=181$) in the 91 screams that were entirely harmonic and 1114 Hz ($SD=194$) in the 500 harmonic climax calls. This difference closely approached but did not reach statistical significance (t -test: $t=1.62$; $P=0.052$). Five individuals were represented by both five or more harmonic scream and five or more harmonic climax calls, and showed mean F0Max values of 1065 Hz ($SD=111$) and 1042 Hz ($SD=111$), respectively. This difference was clearly not significant (paired t -test: $t=-0.23$, $P=0.44$). However, F0Max in harmonic screams was significantly lower than the mean of 1191 Hz ($SD=218$) for the 245 nonlinear screams that included one or more harmonic segments ($t=-3.92$, $P<0.001$). Here, comparisons based on individuals could include five callers, showing mean F0Max values of 1079 Hz ($SD=127$) and 1181 Hz ($SD=119$) for harmonic screams and nonlinear screams, respectively. This difference suggested lower F0 values in the harmonic screams than in nonlinear screams, but the comparison was not statistically significant at the 1% level (paired t -test: $t=-2.6$, $P=0.02$).

Again disregarding individual identity, the mean F0Max value of these 245 nonlinear screams was significantly lower than the value of 1224 Hz ($SD=169$) in the 559 nonlinear climax calls ($t=2.39$, $P<0.01$). Corresponding F0Max values were available for nonlinear screams and nonlinear climax calls from a total of seven individuals, producing mean values of 1171 Hz ($SD=131$) and 1262 Hz (131), respectively. This difference suggested lower values in the screams than climax calls, but the comparison was not statistically significant (paired t -test: $t=-1.44$, $P=0.09$).

C. F0, age, and dominance rank

The last analyses conducted correlated F0 values for introduction and climax calls with age and dominance rank for 12 callers (see Table I). According to the four rank categories low, middle, high, and alpha, each male was assigned a number from 1 to 4. However, no significant relationships emerged, whether between F0Max in introduction calls and either caller age (Pearson's $r=0.03$, $P=0.9$) or dominance rank (Pearson's $r=0.03$, $P=0.9$), between F0Max in nonlinear climax calls and either caller age ($r=0.03$, $P=0.9$) or dominance rank ($r=0.05$, $P=0.9$), and between F0Mean in all climax calls and either age ($r=0.4$, $P=0.2$) or dominance rank ($r=0.24$, $P=0.5$).

Evidence from two individual alpha males suggested that a relationship might nonetheless exist between dominance rank and F0 (see Table I). In 1988 caller SY was the highest-ranking male, and at 1307 Hz also showed the highest mean F0Max values of any adult male in the group at that time. He subsequently lost his right hand to injury from a snare and had fallen to a low-ranking position in the hierarchy by 1997. His F0Max value had also fallen to a mean value of 1221 Hz, which was significantly lower ($t=4.45$, $P<0.001$). In 1997, the new alpha male was BB, who at 1366 Hz showed the highest mean F0Max value of any individual in the group at that time (YB's F0Max value of 1392 Hz comes from a single nonlinear climax call).

IV. DISCUSSION

A. Vocalizing at the edge

Taken together, the results show that nonlinear phenomena in pant hoot climax calls are associated with high rates of vocal-fold vibration. A very basic comparison first showed maximum F0 values to be higher in climax calls than in introduction calls, both for analyses with pooled calls and when caller identity was taken into account. More interestingly, harmonic components of nonlinear climax calls were found to show more extreme values than harmonic climax calls, whether pooling all calls, taking caller identity into account, or testing matched pairs from within the same pant hoots. Furthermore, nonlinear events were specifically associated with the local F0 maximum. These associations between extreme acoustics and NLP were found both globally across call types and locally within calls. F0 values in harmonic segments of nonlinear calls were also found to be rising when approaching the onset points of NLP.

Finding increases in measures of F0 just prior to the onset of irregularity is consistent with both computational

TABLE I. Overview of callers, years recorded, rank, and outcomes for key F_0 variables. Individuals arranged by age.

ID	Age (years)		Rank	Variable	1989 values (Hz)		1997 values (Hz)	
	1989/1997	1989/1997			Mean \pm SD	Mean \pm SD	Mean \pm SD	Mean \pm SD
LK	7/15	lowest/low	F ₀ Max NL climax				1103 \pm 211	
			F ₀ Mean All climax				777 \pm 138	
			F ₀ Mean Intro				489 \pm 94	
AJ	14/22	low/high	F ₀ Max NL climax				1259 \pm 151	
			F ₀ Mean All climax				980 \pm 158	
			F ₀ Mean Intro				426 \pm 114	
MS	14/22	low/high	F ₀ Max NL climax		1205 \pm 208		1204 \pm 141	
			F ₀ Mean All climax		917 \pm 182		922 \pm 169	
			F ₀ Mean Intro		508 \pm 85		475 \pm 111	
YB	16/24	low/low	F ₀ Max NL climax				1392	
			F ₀ Mean All climax				853 \pm 204	
			F ₀ Mean Intro				578 \pm 86	
SL	18/26	low/low	F ₀ Max NL climax				1182 \pm 270	
			F ₀ Mean All climax				939 \pm 179	
			F ₀ Mean Intro				673 \pm 217	
LB	21/29	middle/high	F ₀ Max NL climax				925 \pm 370	
			F ₀ Mean All climax				798 \pm 129	
			F ₀ Mean Intro				505 \pm 58	
BB	23/31	middle/alpha	F ₀ Max NLclimax				1366 \pm 132	
			F ₀ Mean All climax				878 \pm 167	
			F ₀ Mean Intro				657 \pm 107	
BF	23/31	high/middle	F ₀ Max NL climax		1218 \pm 110		1131 \pm 239	
			F ₀ Mean All climax		839 \pm 170		888 \pm 200	
			F ₀ Mean Intro		561 \pm 19		481 \pm 78	
SY	25/33	alpha/low	F ₀ MaxNL climax		1307 \pm 126		1221 \pm 141	
			F ₀ MeanAll climax		968 \pm 173		939 \pm 141	
			F ₀ Mean Intro		586 \pm 136		566 \pm 142	
TU	29/37	high/high	F ₀ Max NL climax		1244 \pm 84		1191 \pm 109	
			F ₀ Mean All climax		967 \pm 188		985 \pm 143	
			F ₀ Mean Intro		494 \pm 37		521 \pm 123	
ST	34/42	high/middle	F ₀ Max NL climax		1131 \pm 51		1225 \pm 129	
			F ₀ Mean All climax		905 \pm 81		897 \pm 142	
			F ₀ Mean Intro		494 \pm 47		513 \pm 64	

(Steinecke and Herzog, 1995; Mergell *et al.*, 2000) and physiological (Brown *et al.*, 2003) models of the vocal folds and their behavior, as well as with findings in human voice production (Gerratt and Kreiman, 2001). The evidence in each case points to more extreme production parameters as a causal trigger when vocal fold behavior bifurcates from stable, synchronized vibration to an irregular mode such as biphonation. Although there are few directly comparable data from other nonhumans, spotted hyenas (*Crocuta crocuta*) are known to produce a harmonic “hoo” vocalization whose increasing F_0 often leads into nonlinear events (Peters *et al.*, 2004; see also Wilden *et al.*, 1998).

Comparisons to scream acoustics were also revealing,

particularly if viewing pant hoots as a kind of voice range profile. On the one hand, overall comparisons between harmonic and nonlinear scream calls confirmed that NLP are associated with elevated F_0 values. On the other hand, harmonic components of screams showed lower F_0 values than did climax calls, even though NLP occurred more often in screams and made up a higher proportion of individual vocalizations. It should be noted that statistical testing was not as compelling as in earlier comparisons. Here, F_0 differences closely approached statistical significance when harmonic scream calls versus harmonic climax calls were contrasted, and did show statistical significance for nonlinear screams versus nonlinear climax calls. However, these results were

based on treating each call as an independent event and thereby inflating statistical power. Testing means that were calculated separately by individual animal did not produce statistically significant differences, although sample sizes were then small.

The critical point is nonetheless that the evidence unequivocally shows F0 values either to be higher in nonlinear climax calls than in nonlinear screams or to be statistically indistinguishable. The data thus show that while vocal effort was either comparable or greater in the climax calls than in the screams, both the prevalence and relative durations of NLP were unambiguously lower. In other words, elevating F0 to maximal values triggered fewer and shorter NLP when animals were producing pant hoots than when screaming. This evidence suggests at least indirectly that vocalizers were inhibiting the occurrence of NLP in climax calls, even as they pushed their vocal production to the upper limits of stable vocal-fold vibration.

B. Implications for function

These results help distinguish between the two explanations of NLP function in climax calls described earlier. Specifically, the results are not consistent with the interpretation that vocalizers are increasing their direct auditory impact on listeners (Owren, 2003; Owren and Rendall, 2001), as this hypothesis would predict that vocalizers should maximize the occurrence of NLP in pant hoots (Riede *et al.*, 2004). Vocalizers did the opposite, exhibiting fewer and shorter NLP, at comparable or higher F0 values, than in screams.

By contrast, the findings are consistent with the hypothesis that the relative prevalence of NLP is a measure of vocalizer condition. The four-component form of pant hoots is compatible with this view, with callers building from modest beginnings to great vocal exertion when producing these vocalizations. Screaming is similar in also being a high-amplitude and evidently effortful sound, but these calls are produced without preamble or build-up. The relative paucity of NLP in climax calls relative to screams therefore underscores the uniqueness of pant hoot form, and the likely importance of the transition from low- to high-effort vocalization. Interestingly, when one or more of the four pant-hoot components is missing, it is the intermediate build-up and terminal let-down sections that are omitted (Arcadi, 1996). In other words, the components that are always present are the ones that specifically make pant hoots reminiscent of human voice range profile (VRP) testing, namely the quiet, low-pitched introduction calls and the loud, high-pitched climax calls.

If the function of producing climax calls at the upper limits of vocal stability is to signal physical condition, one expects either that the signal itself is costly and hence more of a burden for individuals in poor condition (Zahavi, 1975), or that physiological constraints involved in production provide an index of signaler condition (Maynard Smith and Harper, 1995). These possibilities are not mutually exclusive, and both may apply in the case of climax calls. For example, minimizing NLP while producing high-F0 climax calls may significantly increase the effort required, and thereby become

a limiting factor for vocalizers in poor condition. That limit may become particularly pronounced if climax calls are being produced with maximum F0 values but minimal NLP prevalence.

Taken together, these acoustic characteristics may also reflect the vocalizer's competitive ability, for instance as an index of vocal-fold condition, laryngeal control, and aerobic capacity. Those characteristics could in turn be influenced by current systemic condition (Riede *et al.*, 1997; Riede and Stolle-Malorny, 1999). The index involved could also be more indirect, for example if internal hormonal effects associated with striving for or maintaining high rank also affect vocal production capabilities.

Finally, there may also be a role of practice effects. In humans, vocal training is known to increase performance on the VRP (Timmermans *et al.*, 2004). While chimpanzee males are unlikely to practice pant hoots intentionally, higher-ranking animals do in fact produce more of these calls than do lower-ranking animals (Mitani and Nishida, 1993; Clark, 1993). Repetition alone may therefore produce better vocal performance in some individuals compared to others, which could in turn help in achieving and holding higher rank.

While specific associations between F0 and either age or dominance rank were not found, the absence of clear-cut relationships may not be surprising. Sample size was small, involving only 10–12 individual callers. Furthermore, social dominance can be affected by factors other than age, size, or physical characteristics (e.g., Widdig *et al.*, 2001). In chimpanzees in particular, dominance competition is importantly influenced by support from coalition partners. The strongest conclusion to draw at this point is therefore that the current evidence is on several counts consistent with the argument that NLP in climax calls plays a functional role in signaling condition and is not consistent with a direct auditory impact function. More definitive evidence can be sought in future work, for example by increasing the sample of individuals and by being able to include multiple groups and their respective dominance hierarchies.

The more anecdotal findings reported for climax calls from the two alpha males represented in our data set do provide some grounds for optimism in this regard. It is at least suggestive that both individuals were displaying their highest maximum F0 values specifically during periods in which they had alpha rank. It is also notable that the individual SY showed clearly lower F0Max values after having fallen to a low rank from his previous alpha position. Losing his hand to a snare injury would have had many different effects on his competitive ability, including a general decline in overall condition through subsequent infection or disease, and decreased fighting ability due to having only one hand.

C. Related work

The condition-signaling function we have proposed for pant hoot acoustics is unusual in that most work on dominance signaling relates fitness and status to lower F0 values rather than to higher ones [reviewed by Owings and Morton (1998)]. However, one previous study in baboons (*Papio ha-*

madryas ursinus) found that F0 is higher in the loud, explosive “wa” of their “wahoo” calls when produced by higher- rather than lower-ranking adult males (Fischer *et al.*, 2004). This study thus represents a possible parallel, although the NLP that also occurs in wahoo calls were not examined [e.g., Fig. 1 in Fischer *et al.* (2004)]. F0 values in wahoos may function in dominance-related signaling as an index of vocalizer condition (Fischer *et al.*, 2004), as a correlate of greater vocal effort and direct auditory impact, or in conjunction with the presence or absence of NLP.

The filter function of the mammalian vocal tract is related to its cross-sectional area and length (Fant, 1970). Vocal tract length in turn is correlated with body size, for instance in domestic dogs (Riede and Fitch, 1999) and European red deer (Reby and McComb, 2003). As vocal tract length increases, resonance frequencies decrease. European red deer increase their vocal tract lengths while calling (Fitch and Reby, 2001), and tend to maximize this increase when challenged by a competing male (Reby *et al.*, 2005). This behavior suggests that a similar “vocalizing at the edge” mechanism is at work, affecting vocal tract length in the deer whereas it affects fundamental frequency in the chimpanzees and baboons. Although we have not considered supralaryngeal vocal tract acoustics in the chimpanzee, it seems unlikely that formants play a communicative role in the climax calls. Vocal tract resonances are difficult to perceive when high fundamental frequencies are involved, as is for instance well known in soprano singing (Benolen and Swanson, 1990). Nonlinear phenomena occurring within high-F0 calls frequently contribute only short segments of broadband noise that are unlikely to provide a good medium for reliably revealing resonance characteristics (e.g., Wilden *et al.*, 1998).

While there are relatively few studies investigating the effect of personality and social status on voice characteristics in humans (although see Gregory and Webster, 1996; Roy *et al.*, 2000; Puts *et al.*, 2006), there are many demonstrations that hormones can have direct effects on vocal attributes such as F0 (e.g., Vuorenkoski *et al.*, 1978; Akcam *et al.*, 2004). Androgens are particularly important in this regard, and these hormones are also known to influence dominance rank in nonhuman primates generally (e.g., Sapolsky *et al.*, 1997; Abbott *et al.*, 2003), and in chimpanzees in particular (Muller and Wrangham 2004a, b; Muehlenbein *et al.*, 2004). It is thus tempting to hypothesize that a causal relationship exists between rank and vocal characteristics.

Finally, we suggest that the acoustic analyses conducted here help underscore the importance of including NLP in the acoustic analysis of mammalian vocalizations. Although still relatively new and unfamiliar to many researchers interested in vocal behavior, NLPs can now routinely be identified and quantified as commonplace components of vocal communication (e.g., Wilden *et al.*, 1998; Riede *et al.*, 1997, 2000, 2004). As Wilden *et al.* (1998) have pointed out, because NLPs do not readily fit the analysis categories traditionally used in animal communication, it has been common to ignore them. Nevertheless, NLPs are visible in the spectrograms presented in recent work, including frequency jumps, subharmonics, deterministic chaos, and biphonation (e.g., Gouzoules and Gouzoules, 2000; Slocombe and Zuber-

bühler, 2005). Our results suggest that analysis of such irregularities may contribute to our understanding of the functional significance of many vocalizations.

ACKNOWLEDGMENTS

The authors thank the government of Uganda, as represented by the Ministry of Planning and Economic Development, and the Forestry Department, for permission to work in the Kibale National Park. Facilities were provided by the Makerere University Biological Field Station in Kibale Forest and the New York Zoological Society. The Department of Zoology, Makerere University, provided assistance throughout the study. We also thank John D. Anderson IV for assisting in acoustic analysis. Financial support for observation and recording of chimpanzees in the field came from grants awarded to R. W. Wrangham and M. D. Hauser by the following organizations: The National Science Foundation (BNS-8704458), the National Geographic Society (3603-87), and the Leakey Foundation. Funds were also provided by a MacArthur Foundation fellowship to R. W. Wrangham and two one-term Rackham dissertation fellowships (University of Michigan, 114650 and 114465) to A. Clark (Arcadi). TR was supported by the “Deutsche Akademie der Naturforscher Leopoldina” (BMBF-LPD 9901/8-127). MJO was partially supported by NIMH Prime Award 1 R01 MH65317-01A2, Subaward 8402-15235-X, and the Center for Behavioral Neuroscience under the STC Program of the National Science Foundation under Agreement No. IBN-9876754.

- Abbott, D. H., Keverne, E. B., Bercovitch, F. B., Shively, C. A., Mendoza, S. P., Saltzman, W., Snowdon, C. T., Ziegler, T. E., Banjevic, M., Garland, T., Jr., and Sapolsky, R. M. (2003). “Are subordinates always stressed? A comparative analysis of rank differences in cortisol levels among primates,” *Horm. Behav.* **43**, 67–82.
- Akcam, T., Bolu, E., Merati, A. L., Durmus, C., Gereke, M., and Ozkaptan, Y. (2004). “Voice changes after androgen therapy for hypogonadotropic hypogonadism,” *Laryngoscope* **114**, 1587–1591.
- Akerlund, L., Gramming, P., and Sundberg, J. (1992). “Phonetogram and averages of sound pressure levels and fundamental frequencies of speech: Comparison between female singers and nonsingers,” *J. Voice* **6**, 55–63.
- Arcadi, A. C. (1996). “Phrase structure of wild chimpanzee pant hoots: Patterns of production and interpopulation variability,” *Am. J. Primatol.* **39**, 159–178.
- Awan, S. N. (1991). “Phonetographic profiles and f0-SPL characteristics of untrained versus trained vocal groups,” *J. Voice* **5**, 41–50.
- Benolen, M., and Swanson, C. (1990). “The effect of pitch-related changes on the perception of sung vowels,” *J. Acoust. Soc. Am.* **87**, 1781–1785.
- Boersma, P., and Weenink, D. (2005). “Praat: Doing phonetics by computer,” computer program, retrieved from <http://www.praat.org/>
- Brown, C., Alipour, F., Berry, D. A., and Montequin, D. (2003). “Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*),” *J. Acoust. Soc. Am.* **113**, 2114–2126.
- Clark, A. P. (1993). “Rank differences in the production of vocalizations of wild chimpanzees as a function of social context,” *Am. J. Primatol.* **31**, 159–179.
- Clark, A. P., and Wrangham, R. W. (1993). “Acoustic analysis of wild chimpanzee pant hoots: Do Kibale forest chimpanzees have an acoustically distinct food arrival pant hoot?” *Am. J. Primatol.* **31**, 99–109.
- Coleman, R. F. (1993). “Sources of variation in phonetograms,” *J. Voice* **7**, 1–14.
- Fant, G. (1970). *Acoustic Theory of Speech Production* (Mouton, Paris).
- Fischer, J., Kitchen, D. M., Seyfarth, R. M., and Cheney, D. L. (2004). “Baboon loud calls advertise male quality: Acoustic features and their relation to rank, age, and exhaustion,” *Behav. Ecol. Sociobiol.* **56**, 140–148.
- Fitch, W. T., and Reby, D. (2001). “The descended larynx is not uniquely

- human," *Proc. R. Soc. London, Ser. B* **268**, 1669–1675.
- Gerratt, B., and Kreiman, J. (2001). "Toward a taxonomy of non-modal phonation," *J. Phonetics* **29**, 365–381.
- Goodall, J. (1986). *The Chimpanzees of Gombe* (Belknap, Cambridge, MA).
- Gouzoules, H., and Gouzoules, S. (2000). "Agonistic screams differ among four species of macaques: The significance of motivation-structural rules," *Anim. Behav.* **59**, 501–512.
- Gregory, S. W., Jr., and Webster, S. (1996). "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions," *J. Pers. Soc. Psychol.* **70**, 1231–1240.
- Hacki, T., and Heitmueller, S. (1999). "Development of the child's voice: Premutation, mutation," *Int. J. Pediatr. Otorhinolaryngol.* **49**(Suppl. 1), S141–S144.
- Herzel, H., Berry, D., Titze, I. R., and Saleh, S. (1994). "Analysis of vocal disorders with methods from nonlinear dynamics," *J. Speech Hear. Res.* **37**, 1008–1019.
- Heylen, L., Wuyts, F. L., Mertens, F., De Bodt, M., and Van de Heyning, P. H. (2002). "Normative voice range profiles of male and female professional voice users," *J. Voice* **16**, 1–7.
- Heylen, L., Wuyts, F. L., Mertens, F., De Bodt, M., Pattyn, J., Croux, C., and Van de Heyning, P. H. (1998). "Evaluation of the vocal performance of children using a voice range profile index," *J. Speech Lang. Hear. Res.* **41**, 232–238.
- Ikeda, Y., Masuda, T., Manako, H., Yamashita, H., Yamamoto, T., and Komiya, S. (1999). "Quantitative evaluation of the voice range profile in patients with voice disorder," *Eur. Arch. Otorhinolaryngol.* **256**(Suppl. 1), S51–S55.
- Markel, J. D., and Gray, A. H., Jr. (1976). *Linear Prediction of Speech* (Springer-Verlag, New York).
- Marler, P., and Hobbett, L. (1975). "Individuality in a long-range vocalization of wild chimpanzees," *Z. Tierpsychol.* **38**, 97–109.
- Marler, P., and Tenaza, R. (1977). "Signalling behavior of apes with special reference to vocalization," in *How Animals Communicate*, edited by T. Sebeok, Indiana U.P., Bloomington, IN, pp. 965–1003.
- Maynard Smith, J., and Harper, D. (1995). "Animal signals: Models and terminology," *J. Theor. Biol.* **177**, 305–311.
- Mergell, P., Herzel, H., and Titze, I. R. (2000). "Irregular vocal fold vibration: High speed observation and modelling," *J. Acoust. Soc. Am.* **108**, 2996–3002.
- Mitani, J. C., and Nishida, T. (1993). "Contexts and social correlates of long-distance calling by male chimpanzees," *Anim. Behav.* **45**, 735–746.
- Muehlenbein, M. P., Watts, D. P., and Whitten, P. L. (2004). "Dominance rank and fecal testosterone levels in adult male chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, Kibale National Park, Uganda," *Am. J. Primatol.* **64**, 71–82.
- Muller, M. N., and Wrangham, R. W. (2004a). "Dominance, aggression and testosterone in wild chimpanzees: A test of the challenge hypothesis," *Anim. Behav.* **67**, 113–123.
- Muller, M. N., and Wrangham, R. W. (2004b). "Dominance, cortisol and stress in wild chimpanzees (*Pan troglodytes schweinfurthii*)," *Behav. Ecol. Sociobiol.* **55**, 332–340.
- Owings, D. H., and Morton, E. S. (1998). *The Evolution of Vocal Communication: A New Approach* (Cambridge U.P., Cambridge).
- Owren, M. J. (2003). "Vocal production and perception in nonhuman primates provide clues about early hominids and speech evolution," *ATR Sympos. HIS Series, Vol. 1*, pp. 1–19.
- Owren, M. J., and Bernacki, R. H. (1998). "Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals," in *Animal Acoustic Communication*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer-Verlag, Berlin), pp. 129–162.
- Owren, M. J., and Rendall, D. (1997). "An affect-conditioning model of nonhuman primate vocal signaling," in *Perspectives in Ethology*, Vol. **12**: Communication, edited by D. Owings, M. Beecher, and N. Thompson (Plenum, New York), pp. 299–346.
- Owren, M. J., and Rendall, D. (2001). "Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling," *Evol. Anthro.* **10**, 58–71.
- Pabon, J. P. H., and Plomp, R. (1988). "Automatic phonetogram recording supplemented with acoustical voice-quality parameters," *J. Speech Hear. Res.* **31**, 710–722.
- Peters, G., East, M. L., Herzel, H., Henschel, J. R., Mills, M. G. L., Wilhelm, K., and Hofer, H. (2004). "Spotted hyaena whoops: Frequent incidence of vocal instabilities in a mammalian loud call," *Bioacoustics* **14**, 99–109.
- Puts, D. A., Gaulin, S. J. C., and Verdolini, K. (2006). "Dominance and the evolution of sexual dimorphism in human voice pitch," *Evol. Hum. Behav.* **27**, 283–296.
- Reby, D., and McComb, K. (2003). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," *Anim. Behav.* **65**, 519–530.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., and Clutton-Brock, T. (2005). "Red deer stags use formants as assessment cues during intrasexual agonistic interactions," *Proc. R. Soc. London, Ser. B* **272**, 941–947.
- Reynolds, V., and Reynolds, F. (1965). "Chimpanzees of the Budongo forest," in *Primate Behavior: Field Studies of Monkeys and Apes*, edited by I. DeVore (Holt, Rinehart & Winston, New York), pp. 368–424.
- Riede, T., and Fitch, W. T. (1999). "Vocal tract length and acoustics of vocalization in the domestic dog *J. Exp. Biol.* **202**, 2859–2867.
- Riede, T., and Stolle-Malorny, A. (1999). "The vocal change of a kitten with craniocerebellar trauma—a case study," *Bioacoustics* **10**, 131–141.
- Riede, T., Owren, M. J., and Arcadi, A. C. (2004). "Nonlinear acoustics in the pant hoot vocalizations of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," *Am. J. Primatol.* **64**, 277–291.
- Riede, T., Wilden, I., and Tembrock, G. (1997). "Subharmonics, biphonations, and frequency jumps—common components of mammalian vocalization or indicators for disorders," *Z. Säugetiere* **62**(Suppl. 2), 198–203.
- Riede, T., Herzel, H., Mehwald, D., Seidner, W., Trumler, E., Böhme, G., and Tembrock, G. (2000). "Nonlinear phenomena in the natural howling of a dog-wolf mix," *J. Acoust. Soc. Am.* **108**, 1435–1442.
- Roy, N., Bless, D. M., and Heisey, D. (2000). "Personality and voice disorders: A multitrait-multidimension analysis," *J. Voice* **14**, 521–548.
- Sapolsky, R. M., Alberts, S. C., and Altmann, J. (1997). "Hypercortisolism associated with social subordination or social isolation among wild baboons," *Arch. Ital. Otol., Rinol. Laringol. Patol. Cervico-Facciale* **54**, 1137–1143.
- Seidner, W., Krueger, H., and Wernecke, K. D. (1985). "Numerische Auswertung spektraler Stimmfelder," *Sprache-Stimme-Gehoer* **9**, 10–13.
- Slocombe, K. E., and Zuberbühler, K. (2005). "Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role," *J. Comp. Psychol.* **119**, 67–77.
- Steinecke, I., and Herzel, H. (1995). "Bifurcations in an asymmetric vocal fold model," *J. Acoust. Soc. Am.* **97**, 1874–1884.
- Timmermans, B., De Bodt, M. S., Wuyts, F. L., and Van de Heyning, P. H. (2004). "Training outcome in future professional voice users after 18 months of voice training," *Folia Phoniatr. Logop.* **56**, 120–129.
- Titze, I. R. (1992). "Acoustic interpretation of the voice range profile (phonetogram)," *J. Speech Hear. Res.* **35**, 21–34.
- Tokuda, I., Riede, T., Neubauer, J., Owren, M. J., and Herzel, H. (2002). "Nonlinear analysis of irregular animal vocalizations," *J. Acoust. Soc. Am.* **111**, 2908–2919.
- Verdonck-de Leeuw, I. M., and Mahieu, H. F. (2004). "Vocal aging and the impact of daily life: A longitudinal study," *J. Voice* **18**, 193–202.
- Volodina, E. V., Volodin, I. A., Isaeva, I. V., and Unck, C. (2006). "Biphonation may function to enhance individual recognition in the Dhole, *Cuon alpinus*," *Ethology* **112**, 815–825.
- Vuorenkoski, V., Lenko, H. L., Tjernlund, P., Vuorenkoski, L., and Perheentupa, J. (1978). "Fundamental voice frequency during normal and abnormal growth, and after androgen treatment," *Arch. Dis. Child* **53**, 201–209.
- Welham, N. V., and MacLagan, M. A. (2003). "Vocal fatigue: Current knowledge and future directions," *J. Voice* **17**, 21–30.
- Widdig, A., Nürnberg, P., Krawczak, M., Streich, W. J., and Bercovitch, F. B. (2001). "Paternal relatedness and age proximity regulate social relationships among adult female rhesus macaques," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13769–13773.
- Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (1998). "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," *Bioacoustics* **9**, 171–196.
- Zahavi, A. (1975). "Mate selection—a selection for a handicap," *J. Theor. Biol.* **53**, 205–213.

The perception of complex tones by a false killer whale (*Pseudorca crassidens*)

Michelle M. L. Yuen,^{a)} Paul E. Nachtigall, Marlee Breese, and Stephanie A. Vlachos
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 7 July 2006; revised 18 December 2006; accepted 2 January 2007)

Complex tonal whistles are frequently produced by some odontocete species. However, no experimental evidence exists regarding the detection of complex tones or the discrimination of harmonic frequencies by a marine mammal. The objectives of this investigation were to examine the ability of a false killer whale to discriminate pure tones from complex tones and to determine the minimum intensity level of a harmonic tone required for the whale to make the discrimination. The study was conducted with a go/no-go modified staircase procedure. The different stimuli were complex tones with a fundamental frequency of 5 kHz with one to five harmonic frequencies. The results from this complex tone discrimination task demonstrated: (1) that the false killer whale was able to discriminate a 5 kHz pure tone from a complex tone with up to five harmonics, and (2) that discrimination thresholds or minimum intensity levels exist for each harmonic combination measured. These results indicate that both frequency level and harmonic content may have contributed to the false killer whale's discrimination of complex tones. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436640]

PACS number(s): 43.80.Lb [WWA]

Pages: 1768–1774

I. INTRODUCTION

Complex tones are prevalent throughout the environment, from musical instruments to bird songs to sea lion pup calls to the human language. The perception of harmonics and complex tones has been a contentious and well-documented phenomenon in human auditory research. However, very limited research has analyzed this ability in other vertebrate species (Ward Tomlinson and Schwartz, 1988).

The evidence that nonhuman vertebrates possess the capability to discriminate individual components of a complex tone has been accumulating (Ward Tomlinson and Schwartz, 1988). The research thus far has primarily focused on terrestrial species. While no experimental data were collected about the perception of complex tones by a marine mammal species, there is a significant amount of information describing the auditory sensitivities (Nachtigall *et al.*, 2000) and acoustic production systems of odontocetes (Au, 1993; Au *et al.*, 2000).

There are three categories of sounds that odontocetes make. The first includes echolocation sounds of high intensity, high frequency, high repetition rate, and very short duration (Au *et al.*, 2000). The second category of odontocete sounds is comprised of pulsed sounds. Burst pulses are generally very complex and fast, with frequency components sometimes above 100 kHz and average repetition rates of 300 per second.

The final category of odontocete sounds is the narrow-band, low frequency, tonal whistles (Caldwell *et al.*, 1990; Au *et al.*, 2000). With most of their energy below 20 kHz, whistles have been observed with an extensive variety of

frequency patterns, durations, and source levels, each of which can be repeated or combined into more complex phrases (Tyack and Clark, 2000). The categorization of odontocete whistles can be subjective and detailed, and several differences are observed between habitats, species, and phylogenetic relationships. The classification of signature whistles is based on the frequency contours of the fundamental component. Whistles have been characterized based on general contours, such as whistle slope, downsweeps, and upsweeps (Bazua-Duran and Au, 2001). There also seems to be a common correlation that animals with longer length produce whistles with lower maximum frequencies (Ketten, 2000). Resident killer whales produce whistles that are more complex than most delphinid species, with longer durations and greater frequency modulations (Thomsen *et al.*, 2001). This complexity was suggested to have a function for close-range interactions. Beluga whales were observed to increase the high frequency components, such as the amplitude of their whistles with increasing depth (Ridgway *et al.*, 2001). The change in depth may also have changed the air flow and density for the sound production of the whistles. The interesting result was that hearing thresholds did not change with depth despite the changes in whistle production.

Whistles recorded from odontocetes in both laboratory and wild environments have an important role in odontocete social communication as individual signature calls or calls to synchronize group behaviors (Caldwell and Caldwell, 1965; Caldwell *et al.*, 1990; Tyack and Sayigh, 1997). Signature whistles have been observed to facilitate group cohesion (Janik and Slater, 1998), especially when mothers were separated from their calves and possibly facilitating their reunion (Caldwell *et al.*, 1990; Sayigh *et al.*, 1990). Signature whistles were recorded from bottlenose dolphins exclusively in an isolation context, additional evidence for the functional importance of group cohesion (Janik and Slater, 1998). Sig-

^{a)}Current affiliation: National Marine Fisheries Service, Pacific Islands Regional Office, 1601 Kapiolani Blvd., Suite 1110, Honolulu, HI 96814; electronic mail: myuen@hawaii.edu

nature whistles were also documented to contain context-related information and not only identification during discrimination tasks (Janik *et al.*, 1994).

However, some documented evidence disputed the signature call hypothesis, revealing a shared whistle type rather than distinct individual calls (McCowan, 1995; McCowan and Reiss, 1997; 2001). The limitations and discrepancies for categorization methods of whistles may be a significant factor contributing to contrasting measurements and perspectives (Janik, 1999).

Observations of wild populations of dolphins have included signature whistles that were correlated with coordinated group behaviors, such as foraging, feeding, courtship, and mating (Herzing, 2000). One proposed social role of whistles was that of matching interactions (Janik, 2000). Bottlenose dolphins were observed to use learned whistles to respond to conspecifics with the same whistle with both aggressive and affiliative interaction. A recent study suggested that the harmonic structure of whistles could provide an acoustic direction cue to aid the coordinated movement of a group of Hawaiian spinner dolphins (Lammers and Au, 2003). It is advantageous for species that are highly mobile to use cohesion calls or signals to maintain their association with the larger group, assuring recognition with specialized signals.

Despite the rich variety of tonal signals produced by whistling odontocetes, little is known not only about what behavioral and ecological information is perceived from whistle signals, but also about how the animals hear the complex sounds. Determining the relative contribution of each partial in the harmonic signals may clarify how these sounds are identified and resolved by odontocetes.

The echolocation and hearing systems of the false killer whale are adapted for efficient perception of the underwater environment (Madsen *et al.*, 2004; Thomas *et al.*, 1988; 1990; Thomas and Turl, 1990). The vocalizations produced by false killer whales include all three types mentioned earlier, some of which occur in distinct groups while other occurrences are on a gradual continuum transitioning from pulses that are widely spaced apart to whistles that are continuously sinusoidal (Murray *et al.*, 1998). However, no experimental evidence exists regarding odontocete detection of complex tones or their discrimination of harmonic frequencies. Therefore, the objectives of this research project were: (1) to examine the ability of a false killer whale to discriminate pure tones from complex tones, and (2) to determine the minimum intensity level of the harmonic tone required for the whale to discriminate a fundamental frequency and a fundamental frequency plus harmonics.

II. METHODS

A. Animal subject

This adult female, false killer whale named Kina was the subject of a variety of echolocation and audiometric experiments (Supin *et al.*, 2004; 2005; 2003; Thomas *et al.*, 1990; Yuen *et al.*, 2005), and since 1987, she resided at the Hawaii Institute of Marine Biology's Marine Mammal Research Program located in Kaneohe Bay, Oahu, HI. The whale had not

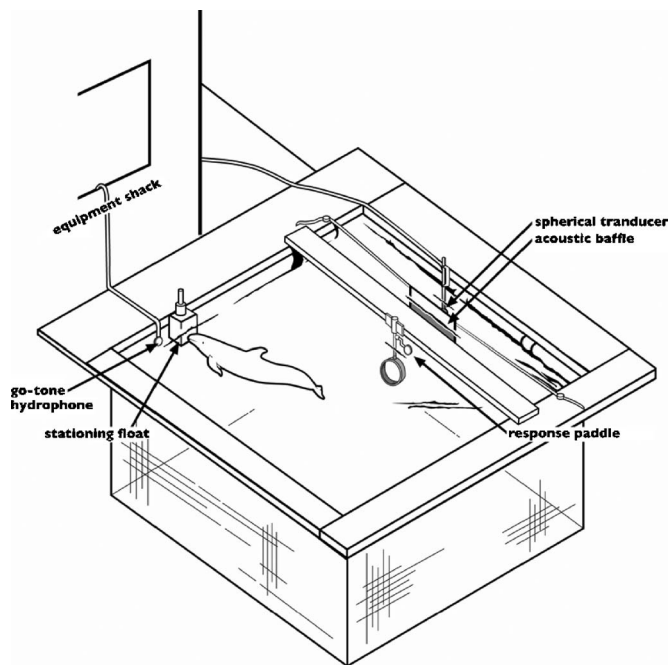


FIG. 1. Experimental configuration for the complex tone discrimination task by a false killer whale.

previously been trained to perform a passive hearing discrimination task and had no prior deliberate exposure to anthropogenic, complex tone stimuli.

B. Electronic equipment

Standard pure tones and complex tones were digitally synthesized with a customized LABVIEW 6I program from a desktop computer implemented with a National Instruments PCI-MIO-16E-1 DAQ card. Using an update rate of 500 kHz, each harmonic component of the complex tone was created and attenuated. Each harmonic component of the complex tone could be turned on or off depending on the stimulus combination presented. The signals were then projected through an ITC-1032 60 mm spherical piezoceramic transducer. A Techtronix TDS 1002 Oscilloscope monitored the signal as it was sent from the computer to the transducer. A second ITC 1032 transducer that had a flat frequency response (± 3 dB) up to 40 kHz was used to calibrate the frequency levels of the signal as it was received in the center of the hoop where the animal was positioned during the stimulus presentation.

C. Experimental setup

The study was conducted within a 6 \times 9 m floating pen in Kaneohe Bay, off the island of Oahu, HI (Fig. 1). This was the same test pen used for an earlier audiogram methodology comparison (Yuen *et al.*, 2005). This wire-fence enclosed pen was supported by floating buoys under the pen's wooden frame.

The transmitting hydrophone was suspended 1 m below the water surface and secured at the center of one side of the pen deck. An acoustic baffle was made of a 0.6 \times 0.9 m aluminum sheet, covered with neoprene on the side facing the transducer, and was hung at the surface of the water at the

half-distance (1 m) between the transducer and the animal. The acoustic baffle reduced the surface reflection of the sound reaching the animal underwater. The animal's hoop was placed 2 m from the sound source, and it was fixed firmly from a wooden beam that stretched across the pen deck. Above the surface of the water, a Styrofoam™ ball response paddle was attached to the wooden beam directly above the hoop. During the intertrial intervals, the animal waited next to a Styrofoam float at the water surface about 3 m away from the hoop, and about 5 m away from the transducer. A small transmitting hydrophone was placed in the water near this float, and projected only a 7 kHz tone. This tone was used to send the animal to the hoop at the beginning of each trial.

D. Discrimination task

There were two phases to this project: a training phase and an experimental phase. During the training phase, the whale was trained to respond to a standard 5 kHz pure tone with a constant amplitude level. The experimental phase included a comparison of the standard pure tone to a complex tone. The fundamental frequency was the same standard pure tone presented in the training phase, and the complex tone included up to five harmonic frequencies. The whale discriminated between the pure tone and the complex tone with different combinations of the components.

During the first part of the training phase, only the standard pure-tone (fundamental only) of 5 kHz was played for a duration of 5 s. The whale was rewarded for touching the response paddle when this tone was played. However, the challenge to this new paradigm was that the whale's only experience with hearing projects required the detection of any sound. She had to be retrained to wait for the entire duration of the tone in order to make a decision.

The next part of the training phase introduced the discrimination task of the standard pure-tone versus a variety of contrasting sounds in order to facilitate the training of the animal to "go" or respond only when the fundamental was heard and to "no-go" or not respond when any other comparison sound was heard. Three different comparison sounds were played: (1) computer-generated white noise, (2) sawtooth shaped tones (as opposed to sine waves), and (3) frequency-modulated tones. The whale was trained to reject these sounds and to remain in the hoop. Novel sounds were first introduced to the animal with very short durations of about 1 s. The duration of each trial was progressively increased as the animal improved her discrimination ability. The whale remained stationed on a target as the duration of the nonpure tone sound gradually increased to 3 s and then to 5 s. After the whale demonstrated that she was able to remain positioned in the hoop while these comparison sounds were played, the target was slowly faded out of each trial until no assistance was necessary for the animal to correctly reject all non-5 kHz comparison sounds.

The third part of training included a discrimination of different pure tones of varying frequencies, starting with higher frequencies that were most different from the standard 5 kHz. Three different frequencies were used: 20, 16, and

11 kHz. When the whale demonstrated that she could successfully discriminate the different pure tones, the final part of the training was to introduce a complex tone with five defined harmonics. Following the establishment of this discrimination, the experimental phase and data collection began.

During the experimental phase, the sequence of the trials was based on the Gellermann series, preventing more than three consecutive trials of the signal present or absent (Gellermann, 1933). A sequence with only half of the trials with a pure tone and the other half with a complex tone prevented response bias of the whale as well as any prediction of trial order.

The whale responded following a go/no-go modified staircase procedure. The whale was trained to remain in an underwater hoop with her pectoral fins touching the hoop at the signal of the trainer. When stationed correctly, the standard, 5 kHz pure tone was transmitted underwater. If the whale heard this tone, she exited the hoop and used her head to go and touch a response paddle suspended above the surface of the water. If a complex tone was transmitted, she was trained to remain stationed in the hoop and not to touch the response paddle, thereby making a correct rejection, or a no-go response. After 10 s, the trainer whistled to signal to the whale that she performed the correct response. For either correct response, the whale was rewarded with fish.

However, failure to respond to the 5 kHz pure tone was an incorrect rejection and termed a "miss." After the 10 s trial, she was called to the trainer and no reward was given. An incorrect detection, also called a "false alarm," occurred if the whale touched the response paddle when a complex tone was played instead of a 5 kHz pure tone, and no reward was given. Ten warm-up trials began each session and were used to gauge the whale's response behavior. During five of the trials a standard easy pure tone was transmitted according to the Gellermann series, and during the other five a complex tone was presented. The remainder of the session only proceeded if at least 90% of the warm-up trials were correct.

The whale discriminated the 5 kHz pure tone from a 5 kHz pure tone with added harmonics. The presence of the harmonics was the cue to the animal that this signal was different from a pure tone, and that she should not respond. The whale's ability to perceive the difference between pure tones and progressively decreasing components of the complex tones was then examined by selectively lowering the amplitude of the harmonic components. Therefore, after each correct rejection of a complex tone (no-go trial), the intensities of all of the harmonic components were jointly attenuated by 2 dB. After each miss, e.g., perceiving the complex tone as a pure tone, the intensities of the harmonic components were increased by 2 dB. In both instances, the trial was considered to be a "reversal," or a switch from increasing to decreasing intensity and vice versa.

At the beginning of each session, the standard pure tone was set at an intensity level 20 dB above sensation level (SL), and each of the harmonic components was also set at least 20 dB above the relative SL. SL is defined as the sound pressure level (SPL) of a sound above its auditory threshold for the individual (Yost, 1994). As in the A-weighted func-

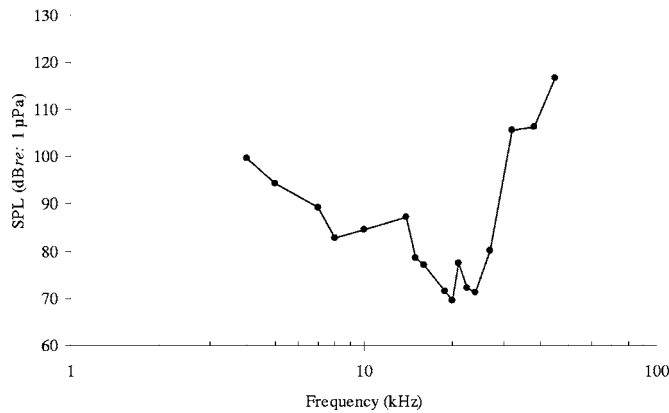


FIG. 2. Average hearing thresholds of three behavioral audiograms of a false killer whale. Thresholds were measured from 2001 to 2003 (from Yuen *et al.*, 2005).

tions used in human audiometrics, the advantage of using SL will eliminate the differences caused by variability in the SPL in an audiogram. The presentation of the data is more comparable and easier to comprehend. SLs reflect how the animal is hearing the relative contribution of the complex tones based on her individual hearing thresholds. The relative intensity levels of each of the SLs for each frequency were calculated from the false killer whale's behavioral audiogram hearing thresholds (Fig. 2) measured from this whale from 2001 to 2003 (Yuen *et al.*, 2005).

The relative intensity levels of each of the harmonic components were reduced equally and simultaneously following each correct trial, until five reversals were completed. The average attenuation value of the five reversals was calculated as the attenuation threshold for that session, and the procedure was repeated until two consecutive sessions occurred with an attenuation reversal average within 5 dB. This final value was calculated to be the amount of attenuation necessary for the whale to discriminate the complex from the pure tone.

When the intensity of each harmonic was too low to hear, the whale most likely heard the fundamental frequency and responded to what she perceived as the standard pure tone. The difference between the fundamental tone and the complex tone represented a threshold for the perception of the combined harmonics within the complex tone.

III. RESULTS

Each of the 13 different combinations of harmonic components resulted in different attenuation threshold values (Table I). Given that all of the combinations were presented to the animal, some of the combinations had only a single harmonic frequency besides the fundamental 5 kHz tone. A comparison of the one-harmonic complex tones is presented in Fig. 3. The discrimination of the 5 kHz pure tone from the set of complex tones that included only one harmonic component (5+10, 5+15, 5+20, and 5+25 kHz), resulted with intensity thresholds that varied considerably, from 6.8 to 29 dB above the relative SL (Fig. 3). There was a substantial difference between the threshold values of these one-harmonic complex tones, with the three higher fre-

TABLE I. Stimulus set of complex tones with a 5 kHz fundamental frequency and up to five harmonics used in the discrimination task by a false killer whale. Also included are the corresponding discrimination thresholds for each complex tone, measured as relative intensity above sensation level (SL).

Complex tone	Frequencies (kHz)	Intensity above SL (dB)
F1+H2	5+10	6.8
F1+H2+H3	5+10+15	12.6
F1+H2+H3+H4	5+10+15+20	9.2
F1+H2+H3+H4+H5	5+10+15+20+25	7.6
F1+H2+H3+H4+H5+H6	5+10+15+20+25+30	6.8
F1+H3	5+15	22.2
F1+H3+H4	5+15+20	14.4
F1+H3+H4+H5+H6	5+15+20+25+30	2.4
F1+H4	5+20	25.6
F1+H4+H5	5+20+25	16.4
F1+H4+H5+H6	5+20+25+30	8.8
F1+H5	5+25	29
F1+H5+H6	5+25+30	18.2

quency combinations having thresholds that differed by exactly 3.4 dB. It appeared that these discrimination thresholds worsened, i.e., the intensity threshold above SL increased, as the frequency of the single harmonic increased. The highest frequency harmonic (5+25) was played at the loudest average intensity level above SL (29 dB) in order for the whale to discriminate it from the pure tone 5 kHz. Therefore, as the one-harmonic component increased in frequency, the intensity level or amplitude of the harmonics portion of the complex tone had to be played at louder intensities above sensation level in order for the whale to distinguish them from the 5 kHz pure tone.

This general order reversed however when additional tones were added. When the intensity thresholds for each of the complex tone combinations were arranged in an ascending order, the whale's discrimination of the different complex sounds from the 5 kHz pure tone appeared to follow a similar pattern. The harmonic combinations were sorted into different sets based on the frequency of the first harmonic

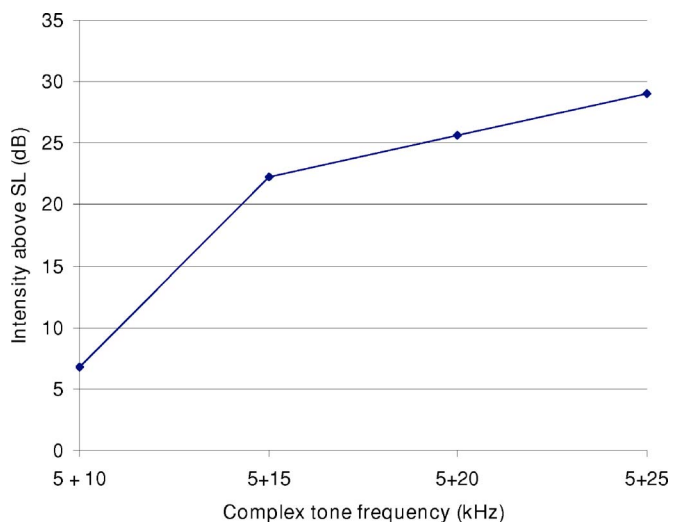


FIG. 3. Graph of the false killer whale discrimination thresholds for complex tones with one harmonic component of various frequencies.

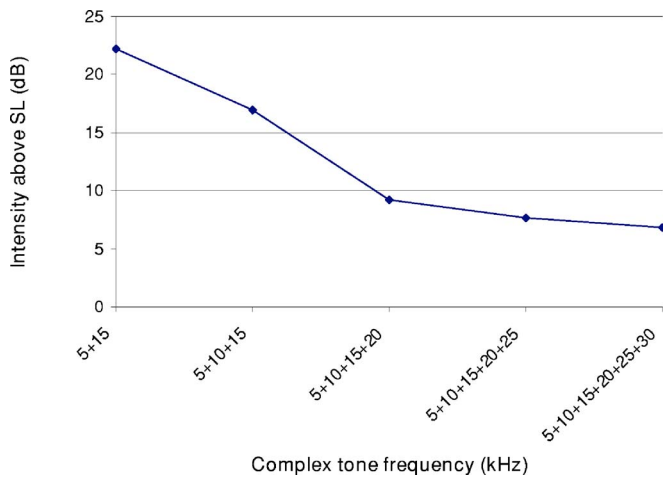


FIG. 4. Graph of the false killer whale discrimination thresholds for complex tones with harmonic components that contained 15 kHz.

added. For example, one set was organized by the harmonic combination of 5+15 kHz (Fig. 4). Each time a harmonic frequency was added to this complex tone, the discrimination ability of the whale improved. When 10 kHz was added, the intensity threshold decreased to 12.6 dB above SL, and then decreased again to 9.2 dB above SL when 20 kHz was added. Finally, as 25 and 30 kHz were added to the complex tone, the threshold again decreased to 7.6 and 6.8 dB above SL, respectively. It appeared that when a second harmonic was added, the discrimination became easier for the false killer whale. With the additional energy from the second harmonic, the complex tone was more discernible from 5 kHz, and the intensity levels were much closer to sensation level. This same pattern was observed as three, four, and five harmonics were added to the complex tone, where the harmonic components of the complex tone were played at lower intensities as more components were added.

Another combination of complex tones measured this effect of adding only the higher harmonics to 15 kHz without the first, lower harmonic of 10 kHz. Once again, the discrimination ability of the whale improved as higher frequencies were added (Fig. 5). The whale discriminated the 5 kHz pure tone from a 5+15 kHz complex tone at a threshold of 22.2 dB above SL. When the next higher harmonic of

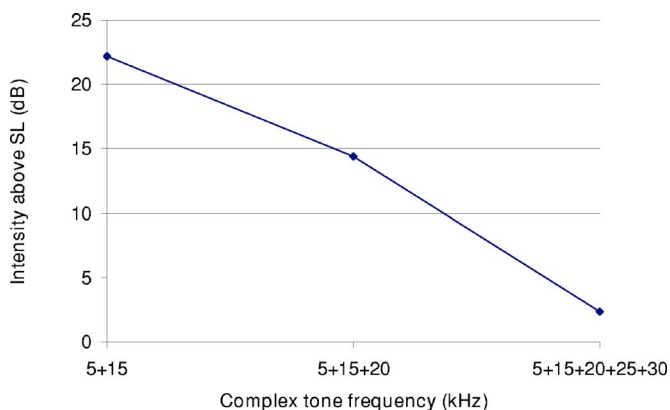


FIG. 5. Graph of the false killer whale discrimination thresholds for complex tones with harmonic frequencies of 15 kHz and higher.

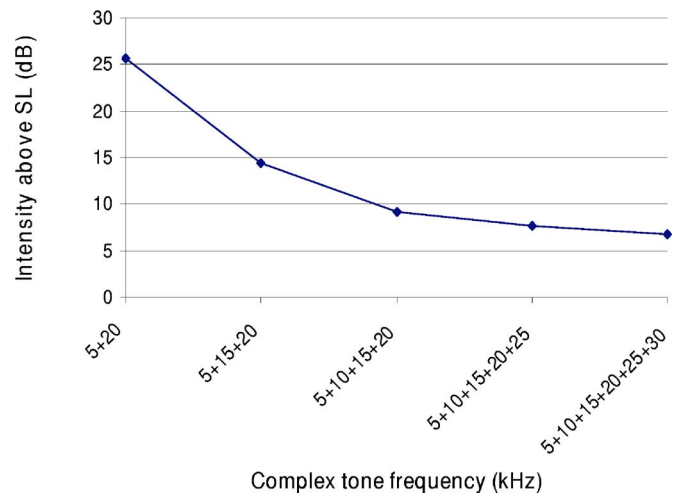


FIG. 6. Graph of the false killer whale discrimination thresholds for complex tones with harmonic components that contained 20 kHz.

20 kHz was added, the intensity threshold fell to 14.4 dB above threshold with a difference of 7.8 dB. Finally, when 25 and 30 kHz were included, the threshold decreased again to 2.4 dB above SL with a difference of 12 dB.

This relationship between the improved discrimination ability of the false killer whale and the increased number of harmonic components of a complex tone was evident in other data sets collected. When additional components were added to the complex tone of 5+20 kHz, the same trend of improved detection resulted (Fig. 6). And in yet another data set with the complex tone of 5+25 kHz, the addition of more harmonic components also resulted in more sensitive discrimination results by the animal (Fig. 7).

IV. DISCUSSION

The results from this complex tone discrimination task demonstrated: (1) that the false killer whale was able to discriminate a 5 kHz pure tone from complex tones with up to five harmonics, and (2) that discrimination thresholds or minimum intensity levels were measured for each harmonic combination. When the various complex tones were played to the false killer whale, she was able to discriminate each of them from a 5 kHz pure tone. Each harmonic combination was projected at different intensity levels above SL, each

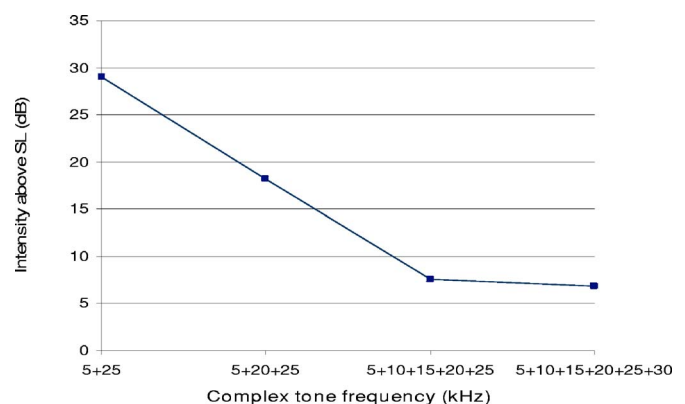


FIG. 7. Graph of the false killer whale discrimination thresholds for complex tones with harmonic components that contained 25 kHz.

with its own threshold level results at which the harmonics were barely audible and distinguishable from the pure tone. The results from this current investigation with a false killer whale were the first to demonstrate that a marine mammal has the ability to clearly identify a complex tone with up to five harmonics from a pure tone with the same frequency as the fundamental component.

It seemed likely that the false killer whale's ability to discriminate pure tones from complex tones improved possibly due to the detection of increased intensity and spectrum width from the additional harmonics (Figs. 4–7). As harmonics were added, the overall intensity required to discriminate the complex from the pure tone decreased. In other words, as the spectrum width of the complex tone increased, the required intensity to discriminate decreased. The complex tones may have been stimulating the inner ear differently as more harmonic components were added and as the intensity and spectrum width increased, explaining the observed trend in this experiment.

This may have an ecologically significant role for maintaining pod associations by providing cues and signals that may be perceived in these higher harmonics. The high frequency components of killer whale calls were measured to be more directional than the lower frequency components, thereby changing the received spectral content that corresponded to the movement of the signaler (Miller, 2002). If so, the higher frequency harmonics may have a significant role for facilitating group cohesion by containing important information for synchronized behaviors between signalers and receivers.

The adaptive values for odontocetes to discriminate harmonic vocalizations are found in additional theories of group cohesion and identification, including theories of harmonics as an acoustic by-product necessary for louder, higher intensity sounds that travel great distances. In a recent study by Lammers and Au (2003), the harmonic content of Hawaiian spinner dolphin whistles were observed to contain valuable information for a direction of movement cue. The broadband whistles are directional with frequency, and depending on the location of the receiving dolphin, information of the animals' positions in the pod are proposed by the authors to be inferred in the harmonic energy of the whistles. This may contribute to the well-coordinated movements observed over great distances of the Hawaiian spinner dolphin as well as other whistling, delphinid species. Although this particular model has not been applied for other odontocetes, it has been demonstrated from the present results that the discrimination capability for these complex tones exists for the false killer whale.

In addition to these observations, there appeared contradictory evidence demonstrating that the whale may have altered her strategy for sound perception, and that frequency or spectral analysis was performed when the fundamental frequency was presented with only one harmonic (Fig. 3). It was probable that the whale was not able to resolve the individual harmonic component as the harmonic to fundamental ratio increased, for example when 5+25 kHz was compared to the standard 5 kHz pure tone. This difference may have resulted from the higher frequency of the widely sepa-

rated, harmonic component perceived distinctly from the fundamental. If each area within the basilar membrane of the cochlea responded to a specific frequency, then it was possible that the combination of 5+10 kHz resulted with a greater stimulation of hair cells when discriminating the sound as the complex tone. However, the distance between 5 and 25 kHz may have been so large that each frequency was heard individually in the 5+25 kHz complex tone, thereby reducing the whale's ability to recognize this as a complex tone. This combination may not have been heard by the whale as a complex tone.

Although complex tones have been identified and characterized from odontocete vocalizations for decades, this was the first attempt to experimentally determine the ability of a marine mammal to perceive and distinguish harmonic combinations of sound. The results confirmed that a false killer whale, a species documented for producing complex sounds, can discriminate such complex sounds from simple, pure tones. Complex tones with as little as one harmonic to as many as five harmonics, and including various combinations in between, were all discernible from the fundamental frequency alone.

ACKNOWLEDGMENTS

The authors are grateful for the support of Whitlow Au, Dera Look, Vincent De Paolo, T. Aran Mooney, Kristen A. Taylor, Jeff Pawloski, Christopher Kam, and Maurice Radke for the illustrations, and the Marine Mammal Research Program of the Hawaii Institute of Marine Biology at the University of Hawaii. The study was supported by Office of Naval Research Grant No. N00014-98-1-0687, for which the authors thank Robert Gisiner. This work was conducted under Marine Mammal Permit 978-1567 issued to Paul E. Nachtigall by the NMFS NOAA Office of Protected Resources. This is contribution 1260 from the Hawaii Institute of Marine Biology.

- Au, W. W. L. (1993). *The Sonar of Dolphins* (Springer, New York).
- Au, W. W. L., Popper, A. N., and Fay, R. R. (2000). *Hearing by Whales and Dolphins* (Springer, New York).
- Bazua-Duran, C., and Au, W. W. L. (2001). "The whistles of Hawaiian spinner dolphins," *J. Acoust. Soc. Am.* **112**, 3064.
- Caldwell, M. C., and Caldwell, D. K. (1965). "Individualized whistle contours in bottlenosed dolphins, (*Tursiops truncatus*)," *Science* **207**, 434–435.
- Caldwell, M. C., Caldwell, D. K., and Tyack, P. (1990). "Review of the signature whistle hypothesis for the Atlantic bottlenose dolphin, *Tursiops truncatus*," in *The Bottlenose Dolphin: Recent Progress in Research*, edited by S. Leatherwood and R. Reeves (Academic, San Diego).
- Gellermann, L. W. (1933). "Chance orders of alternating stimuli in visual discrimination experiments," *J. Genetic Psy.* **42**, 206–208.
- Herzing, D. L. (2000). "Acoustics and social behavior of wild dolphins: Implications for a sound society," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 225–272.
- Janik, V. M. (1999). "Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods," *Anim. Behav.* **57**, 133–143.
- Janik, V. M. (2000). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," *Science* **289**, 1355–1358.
- Janik, V. M., Dehnhardt, G., and Todt, D. (1994). "Signature whistle variations in a bottlenosed dolphin, *Tursiops truncatus*," *Behav. Ecol. Sociobiol.* **35**, 243–248.
- Janik, V. M., and Slater, P. J. B. (1998). "Context-specific use suggests that

- bottlenose dolphin signature whistles are cohesion calls," *Anim. Behav.* **56**, 829–838.
- Ketten, D. R. (2000). "Cetacean ears," in *Hearing by Whales*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 43–108.
- Lammers, M. O., and Au, W. W. L. (2003). "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?," *Marine Mammal Sci.* **19**, 249–264.
- Madsen, P. T., Kerr, I., and Payne, R. (2004). "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*," *J. Acoust. Soc. Am.* **207**, 1811–1823.
- McCowan, B. (1995). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*)," *Ethology* **100**, 177–193.
- McCowan, B., and Reiss, D. (1997). "Quantitative comparison of whistles repertoires from captive adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): A re-evaluation of the signature whistle hypothesis," *Ethology* **100**, 193–209.
- McCowan, B., and Reiss, D. (2001). "The fallacy of 'signature whistles' in bottlenose dolphins: A comparative perspective of 'signature information' in animal vocalizations," *Anim. Behav.* **65**, 1151–1162.
- Miller, P. J. O. (2002). "Mixed-directionality of killer whale stereotyped calls: A direction of movement cue?," *Behav. Ecol. Sociobiol.* **52**, 262–270.
- Murray, S. O., Mercado, E., and Roitblat, H. L. (1998). "Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations," *J. Acoust. Soc. Am.* **104**, 1679–1688.
- Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (2000). "Psychoacoustic studies of dolphin and whale hearing," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 330–363.
- Ridgway, S. H., Carder, D. A., Kamolnick, T., Smith, R. R., Schlundt, C. E., and Elsberry, W. R. (2001). "Hearing and whistling in the deep sea: Depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti, Cetacea)," *J. Exp. Biol.* **204**, 3829–3841.
- Sayigh, L. S., Tyack, P., Wells, R. S., and Scott, M. D. (1990). "Signature whistles of free-ranging bottlenose dolphin, *Tursiops truncatus*: Stability and mother-offspring comparisons," *Behav. Ecol. Sociobiol.* **26**, 247–260.
- Supin, A. Y., Nachtigall, P. E., Au, W. W. L., and Breese, M. (2004). "The interaction of outgoing echolocation pulses and echoes in the false killer whale's auditory system: Evoked-potential study," *J. Acoust. Soc. Am.* **115**, 3218–3225.
- Supin, A. Y., Nachtigall, P. E., Au, W. W. L., and Breese, M. (2005). "Invariance of echo-responses to target strength and distance in an echolocating false killer whale: Evoked potential study," *J. Acoust. Soc. Am.* **117**, 3928–3935.
- Supin, A. Y., Nachtigall, P. E., Pawloski, J., and Au, W. W. L. (2003). "Evoked potential recording during echolocation in a false killer whale *Pseudorca crassidens*," *J. Acoust. Soc. Am.* **113**, 2408–2411.
- Thomas, J., Chun, N., Au, W. W. L., and Pugh, K. (1988). "Underwater audiogram of a false killer whale (*Pseudorca crassidens*)," *J. Acoust. Soc. Am.* **84**, 936–940.
- Thomas, J., Pawloski, J., and Au, W. W. L. (1990). "Masked hearing abilities of a false killer whale (*Pseudorca crassidens*)," in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 395–404.
- Thomas, J., and Turl, C. (1990). "Echolocation characteristics and range detection threshold of a false killer whale (*Pseudorca crassidens*)," in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York).
- Thomsen, F., Franck, D., and Ford, J. K. B. (2001). "Characteristics of whistles from the acoustic repertoire of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia," *J. Acoust. Soc. Am.* **109**, 1240–1246.
- Tyack, P. L., and Clark, C. W. (2000). "Communication and acoustic behavior of dolphins and whales," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 156–224.
- Tyack, P. L., and Sayigh, L. S. (1997). "Vocal learning in cetaceans," in *Influences on Vocal Development*, edited by C. T. Snowdon and M. Hausberger (Cambridge University Press, Cambridge), pp. 208–233.
- Ward Tomlinson, R. W., and Schwartz, D. W. F. (1988). "Perception of the missing fundamental in nonhuman primates," *J. Acoust. Soc. Am.* **84**, 560–565.
- Yost, W. A. (1994). *Fundamentals of Hearing: An Introduction* (Academic, San Diego).
- Yuen, M. M. L., Nachtigall, P. E., Supin, A. Y., and Breese, M. (2005). "Comparison of behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," *J. Acoust. Soc. Am.* **118**, 2695–2699.

Bottlenose dolphin (*Tursiops truncatus*) steady-state evoked responses to multiple simultaneous sinusoidal amplitude modulated tones

James J. Finneran

U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351,
53560 Hull St, San Diego, California 92152

Dorian S. Houser

Biomimetica, 7951 Shantung Drive, Santee, California 92071

(Received 6 April 2006; revised 15 November 2006; accepted 6 December 2006)

Auditory steady-state evoked potentials were measured in a bottlenose dolphin (*Tursiops truncatus*) in response to single and multiple sinusoidal amplitude modulated (SAM) tones. Tests were conducted in air using a “jawphone” sound projector. Evoked potentials were recorded noninvasively using surface electrodes embedded in suction cups. Sound stimuli consisted of SAM tones with 1, 2, 3, or 4 carrier frequencies (10, 20, 30, 40 kHz), each with a unique modulation frequency. Stimulus sound pressure levels were varied in 5-dB steps from approximately 120 to 60–75 dB re 1 μ Pa, depending on frequency. Evoked potentials followed the temporal envelope of each stimulus, resulting in spectral components at each unique modulation frequency. Spectral analysis was used to evaluate the response amplitude for each carrier as a function of stimulus level. There were no significant differences between thresholds obtained with single and multiple stimuli at 10, 30, and 40 kHz. At 20 kHz, thresholds obtained with three components were higher than those obtained with four components, possibly revealing interactions between stimuli with less than one octave frequency separation. The use of multiple SAM stimuli may offer substantial advantages for studies of marine mammal hearing, where testing time and access to subjects are typically limited. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2431330]

PACS number(s): 43.80.Lb, 43.64.Ri [WWA]

Pages: 1775–1782

I. INTRODUCTION

Auditory evoked potentials (AEPs) have been measured in a variety of marine mammal species using clicks (e.g., Bullock and Ridgway, 1972; Ridgway *et al.*, 1981; Popov and Supin, 1985), tone-pips (e.g., Bullock *et al.*, 1968; Popov and Supin, 1985; Popov and Supin, 1990; Szymanski *et al.*, 1999; Wolski *et al.*, 2003), and amplitude modulated tones [i.e., sinusoidal amplitude modulated (SAM) tones or two-tone (TT) stimuli] (e.g., Dolphin *et al.*, 1995; Supin and Popov, 1995; Dolphin, 1996; Nachtigall *et al.*, 2004; Yuen *et al.*, 2005; Finneran and Houser, 2006). Although short duration clicks and tone pips generate relatively large AEPs, SAM tones and TT stimuli possess narrower frequency spectra and are thus often preferred for estimating hearing thresholds and creating audiograms (e.g., Supin *et al.*, 2001; Nachtigall *et al.*, 2004; Yuen *et al.*, 2005). SAM and TT stimuli produce the so-called envelope following response (EFR) or auditory steady-state response (ASSR), a harmonic evoked potential with a fundamental frequency related to the SAM tone modulation frequency or TT difference frequency. Since the EFR/ASSR occurs at a discrete frequency, multiple SAM [or multi-tone (MT)] stimuli, each with a unique modulation frequency (or creating unique difference frequencies), would allow simultaneous measurements at multiple test frequencies. This phenomenon has been verified in humans and it is now relatively common for human ASSR measurements to employ multiple, simultaneous SAM

stimuli (e.g., Lins and Picton, 1995; John *et al.*, 1998; John *et al.*, 2002). If the responses do not interact, it is more efficient to measure responses to multiple stimuli rather than to measure separate responses to single stimuli (John *et al.*, 2002). Provided that sufficient frequency separation exists (at least one octave between each adjacent carrier frequency), up to four SAM stimuli may be presented to each ear simultaneously in humans without significantly affecting the response amplitudes (Lins and Picton, 1995; John *et al.*, 1998).

The use of multiple EFR/ASSR techniques in marine mammal audiology would benefit investigations by increasing the rate at which data may be collected. This may be particularly important in the study of stranded/rehabilitating individuals where the time available for testing is limited, or in temporary threshold shift (TTS) experiments where threshold shifts depend on the time following the fatiguing exposure. Investigation of multiple EFR/ASSR in marine mammals has been limited. Dolphin (1996) used multiple (2, 3, or 4) tones to generate ASSRs in bottlenose dolphins (*Tursiops truncatus*). These data verified that the dolphin auditory system could simultaneously track the multi-envelope stimuli created from 2–4 pure tones (1, 3, or 6 dominant envelope components). Popov *et al.* (1997; 1998) noted suppression of the ASSR to a 76 kHz SAM when a lower amplitude SAM was simultaneously presented at 85 kHz, a separation of approximately 1/6 octave (ASSR suppression was also observed when higher-frequency, lower-amplitude

pure tones were simultaneously presented). None of the previous studies specifically examined hearing thresholds using multiple stimuli.

The goal of the present study was to verify that multiple ASSRs could be elicited in a bottlenose dolphin using multiple, simultaneous SAM stimuli and to compare hearing thresholds obtained with single and multiple SAM tones. Evoked response amplitudes as functions of stimulus level and the number of SAM tones are presented and the potential applications of this technique to odontocete audiometry are discussed.

II. METHODS

The subject was a 41-y old female bottlenose dolphin (BLU, 200 kg) with extensive experience in cooperative psychophysical and electrophysiological tests. Previous tests revealed relatively poor high-frequency hearing above approximately 50 kHz (Finneran and Houser, 2006). The subject was housed in a 9×9 m netted enclosure in San Diego Bay, California. The study followed a protocol approved by the Institutional Animal Care and Use Committee of the Biosciences Division, Space and Naval Warfare Systems Center, San Diego, and all applicable U.S. Department of Defense guidelines for the care of laboratory animals.

AEP methods were similar to those described by Finneran and Houser (2006). Measurements were conducted in air with the subject resting on a foam mat. Sound stimuli were presented using a “jawphone” consisting of a spherical sound projector (ITC 1042) embedded in a silicon rubber (Rhodia V-1065) suction cup placed in the pan region of the lower left jaw. The jawphone calibration technique is described in detail in Finneran and Houser (2006) and thresholds resulting from this approach have been benchmarked against underwater behavioral audiograms in five dolphins (Finneran and Houser, 2006; Houser and Finneran, 2006). Expected differences between the in-air AEP data and underwater behavioral are revealed by the direct comparisons made in these studies. There has been no universally accepted technique for calibrating jawphones and the suitability of jawphone-measured thresholds as predictors of free field or direct field underwater thresholds is somewhat controversial, thus the thresholds presented should be considered as estimates. However, the present study is concerned with differences between thresholds obtained with single and multiple stimuli delivered through the same jawphone, which would not be affected by the jawphone calibration method.

Stimuli consisted of combinations of SAM tones containing discrete spectral peaks at the carrier frequency and the carrier \pm the modulation frequency. Seven stimulus waveforms were used, each containing one to four individual SAM tone components: (S_{1a} , S_{1b} , S_{1c} , S_{1d}) single SAM tones with carrier frequencies of 10, 20, 30, and 40 kHz and modulation rates of 0.7, 1.0, 1.1, and 1.2 kHz, respectively; (S_2) multiple SAM consisting of 20 and 40 kHz carriers modulated at 1.0 and 1.2 kHz, respectively; (S_3) multiple SAM consisting of 10, 20, and 40 kHz carriers modulated at 0.7, 1.0, and 1.2 kHz, respectively; and (S_4) multiple SAM consisting of 10, 20, 30, and 40 kHz carriers modulated at 0.7,

TABLE I. Parameters for single and multiple SAM tones. SPL values indicate the maximum levels tested; component levels were uniformly reduced by 5 dB for each successive measurement.

Stimulus	Number of components	Carrier frequencies (kHz)	Modulation frequencies (kHz)	SPL (dB re 1 μ Pa)
S_{1a}	1	10	0.7	118
S_{1b}	1	20	1.0	120
S_{1c}	1	30	1.1	120
S_{1d}	1	40	1.2	120
S_2	2	20 40	1.0 1.2	120 120
S_3	3	10 20 40	0.7 1.0 1.2	117 119 119
S_4	4	10 20 30 40	0.7 1.0 1.1 1.2	118 120 120 120

1.0, 1.1, and 1.2 kHz, respectively. Stimulus parameters are summarized in Table I. Carrier frequencies for S_2 and S_3 were chosen to maintain the full octave separation used as a guideline for humans (Lins and Picton, 1995; John *et al.*, 1998). Lateral suppression data from Popov *et al.* (1997; 1998) show negligible suppression of ASSRs to SAM tones from pure tones an octave higher. The inclusion of the 30 kHz SAM component in S_4 was specifically chosen to violate the full-octave guideline. Auditory filter widths estimated by Finneran *et al.* (2002a) for *Tursiops* suggest little interaction of a 30 kHz signal at 20 kHz, with equivalent rectangular bandwidths (ERBs) at 20 kHz of approximately 2–3 kHz. Filter widths at 30 kHz measured by Finneran *et al.* were larger than those at 20 kHz (ERBs approximately 2–9 kHz). Lemonds *et al.* (1997) estimated ERBs in *Tursiops* at 40 kHz to be about 7 kHz, so it is likely that an auditory filter centered at 40 kHz would overlap to some extent with the 30 kHz SAM. Modulation frequencies were chosen to keep rates near areas of high amplitude as seen in dolphin modulation rate transfer functions (MRTFs) (Dolphin *et al.*, 1995; Supin and Popov, 1995), to avoid harmonics of lower modulation frequencies, and provide sufficient frequency separation to resolve the responses during analysis. Individual stimulus components were digitally created and summed before digital-to-analog conversion using a 2-MHz update rate and 16-bit resolution (National Instruments PCI-6251). Stimuli were bandpass filtered (200 Hz to 150 kHz, Krohn-Hite 3C series module) before being applied to the jawphone. Sound stimuli were 62 ms in duration, including a 1-ms cosine envelope rise and fall. Measurements with each stimulus began with individual component sound pressure levels (SPLs) near 120 dB re 1 μ Pa (depending on frequency and the number of stimuli); the SPL of all components was reduced by 5 dB for each successive measurement.

Evoked potentials were recorded using three 10-mm gold cup electrodes embedded in silicon suction cups: the noninverting (+) electrode was placed just behind the blowhole and slightly contralateral to the jawphone, the inverting

(-) electrode was placed near the left external auditory meatus, and the common electrode was placed on the back near the dorsal fin. Conductive gel was applied to each electrode before placement. Electrode signals were differentially amplified and filtered (300–3000 Hz, Grass ICP511), then digitized (PCI-6251) at 10 kHz. Recordings were made over 71-ms epochs synchronized with the start of each stimulus, resulting in an intermittent stimulus presentation schedule that allowed AEP latencies to be observed at high stimulus SPLs. Stimuli were presented at a rate of approximately 14/s. Frequency analysis was performed on a 60-ms time window centered on the evoked response. This resulted in an analysis frequency resolution of 16.7 Hz, sufficient to resolve the 100 Hz or larger separation in stimulus modulation rates. Coherent averaging was performed in the frequency domain using a total of 500 epochs for each measurement. Epochs with peak voltages exceeding $12 \mu\text{V}$ were excluded from averaging.

As in Finneran and Houser (2006), magnitude-squared coherence (MSC) was used to objectively determine if the AEP component at each modulation frequency was statistically different from noise (Dobie and Wilson, 1989; Dobie, 1993; Dobie and Wilson, 1996). MSC calculations used 20 subaverages. Critical values for MSC (MSC_{crit}), using $\alpha = 0.01$, were obtained from Amos and Koopmans (1963) and Brillinger (1978). If the calculated MSC was greater than MSC_{crit} , the AEP at the modulation frequency was considered to be detected.

Measurements were repeated four times, yielding four threshold measurements for each stimulus frequency listed in Table I. Hearing thresholds derived from ASSR data are often estimated by linear regression of the AEP amplitude versus stimulus SPL data (e.g., Nachtigall *et al.*, 2004; Yuen *et al.*, 2005; Finneran and Houser, 2006) or defined as the stimulus SPL corresponding to the lowest detectable response (e.g., Lins, 1996; John *et al.*, 1998). In this study, we report thresholds based on the mean SPL of the lowest detected response and the next highest undetected response. Differences in mean thresholds were tested for significance using a one-way ANOVA (10, 20, 40 kHz) or *t* test (30 kHz, since there were only two experimental conditions) (GraphPad Software, 2003).

III. RESULTS

Figure 1 shows example time waveforms for sound stimuli and evoked responses. The stimulus and evoked response waveforms corresponding to S_{1a} , S_{1c} , S_{1d} are not shown, but were similar to those for S_{1b} . At the highest stimulus SPLs ($\sim 120 \text{ dB re } 1 \mu\text{Pa}$), evoked responses were visible in the time domain. At lower SPLs, responses were indistinguishable from noise. In each case, the evoked response followed the envelope of the sound stimulus, as expected. Latencies of 4–5 ms between stimulus and response onsets confirmed that the measured voltages were indeed evoked responses and not stimulus artifacts.

Figure 2 shows the frequency spectra from the time series shown in Fig. 1. Evoked responses corresponding to

each SAM component are visible at the corresponding modulation frequencies. The second harmonic of the 0.7-kHz modulation rate is also visible at 1.4 kHz.

Figure 3 shows the mean amplitudes (across the four measurements) of the detected responses as functions of the corresponding stimulus SPLs for SAM components with 10-, 20-, 30-, and 40-kHz carrier frequencies. The shapes of the curves are generally similar in each panel, though there are amplitude differences at specific SPLs. The 40-kHz curves featured shallow slopes near threshold, making a regression threshold technique problematic.

Estimated thresholds are plotted in Fig. 4. Previously measured AEP thresholds for BLU, obtained with single SAM tones and a regression technique (Finneran and Houser, 2006) are also shown for comparison.

There were no significant differences ($\alpha=0.05$) between thresholds at 10, 20, or 30 kHz. At 20 kHz, thresholds obtained with S_3 were 8.8 dB higher than those obtained with S_4 ($p < 0.01$). Differences between the other 20-kHz thresholds were not significant ($\alpha=0.05$).

IV. DISCUSSION

The increasing popularity of AEP methods primarily stems from the ability to test subjects that have not been specifically trained for behavioral hearing tests, including species not routinely kept under human care. In many circumstances, such as wild “catch- and-release” programs (Wells and Scott, 1990; Cook *et al.*, 2004) and stranded or rehabilitating animals (e.g., Ridgway and Carder, 2001; Nachtigall *et al.*, 2005; Cook *et al.*, 2006), competing priorities between providing health care and obtaining a variety of scientific data from species not normally available limit the time and access for audiometric testing. In other situations, such as TTS testing, the auditory phenomenon under investigation is transient, so hearing tests must be completed very quickly.

The use of the multiple ASSR technique may offer substantial advantages in these contexts since it would allow thresholds to be measured at multiple frequencies simultaneously. Studies in humans to evaluate the improvement in efficiency over single stimuli have shown that while the use of the multiple ASSR technique does not yield a 1:1 reduction in testing time (i.e., using two stimuli does not halve the testing time), testing time is always lower compared to single stimuli (John *et al.*, 2002). The technique may be particularly attractive for TTS testing, where hearing tests could be performed at multiple sound frequencies after exposure to a single fatiguing stimulus. Past investigations of marine mammal TTSs at multiple frequencies repeated exposures over multiple days so each frequency could be tested just after the exposure (Schlundt *et al.*, 2000; Finneran *et al.*, 2000; Finneran *et al.*, 2002b). The multiple ASSR technique would therefore not only reduce the time required to measure hearing thresholds, but also eliminate day-to-day variability in exposure levels and lessen the number of times individuals would be subjected to intense sound exposures.

Before widespread application of multiple ASSR techniques in marine mammals, the methods must be bench-

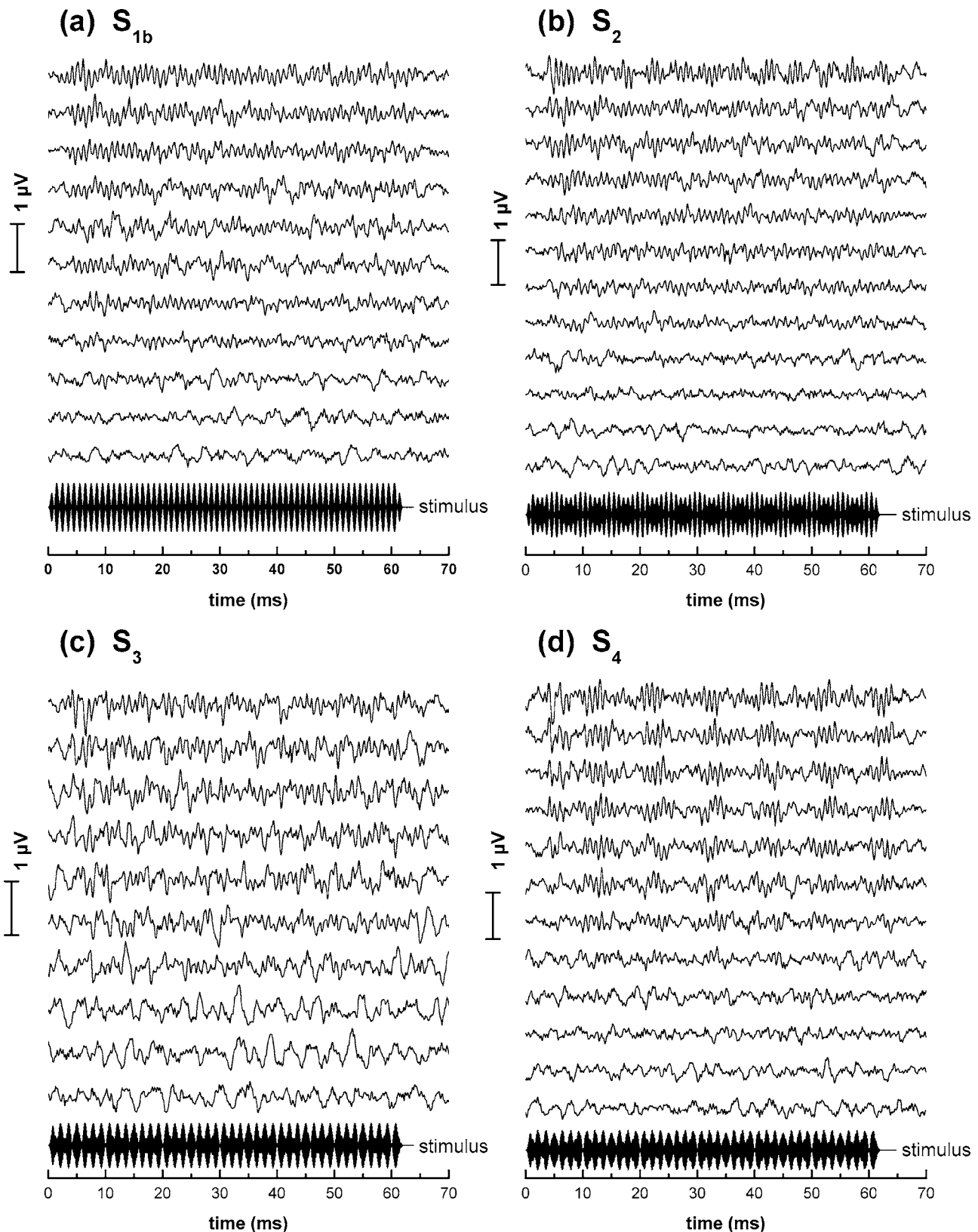


FIG. 1. Stimulus and evoked response waveforms for (a) S_{1a} , (b) S_2 , (c) S_3 , and (d) S_4 . Stimulus frequencies and maximum SPL for each component are listed in Table I. The stimulus SPL was reduced by 5 dB for each successive recording (from top to bottom within each panel).

marked versus more established AEP and behavioral data and the potential for interactions between multiple stimuli must be defined. Benchmarking of multiple ASSR methods in humans (e.g., Lins and Picton, 1995; John *et al.*, 1998;

John *et al.*, 2002) has established that up to four frequencies, separated by at least one octave, can be tested simultaneously in each ear without significantly affecting evoked response amplitudes. Bottlenose dolphins (as well as other

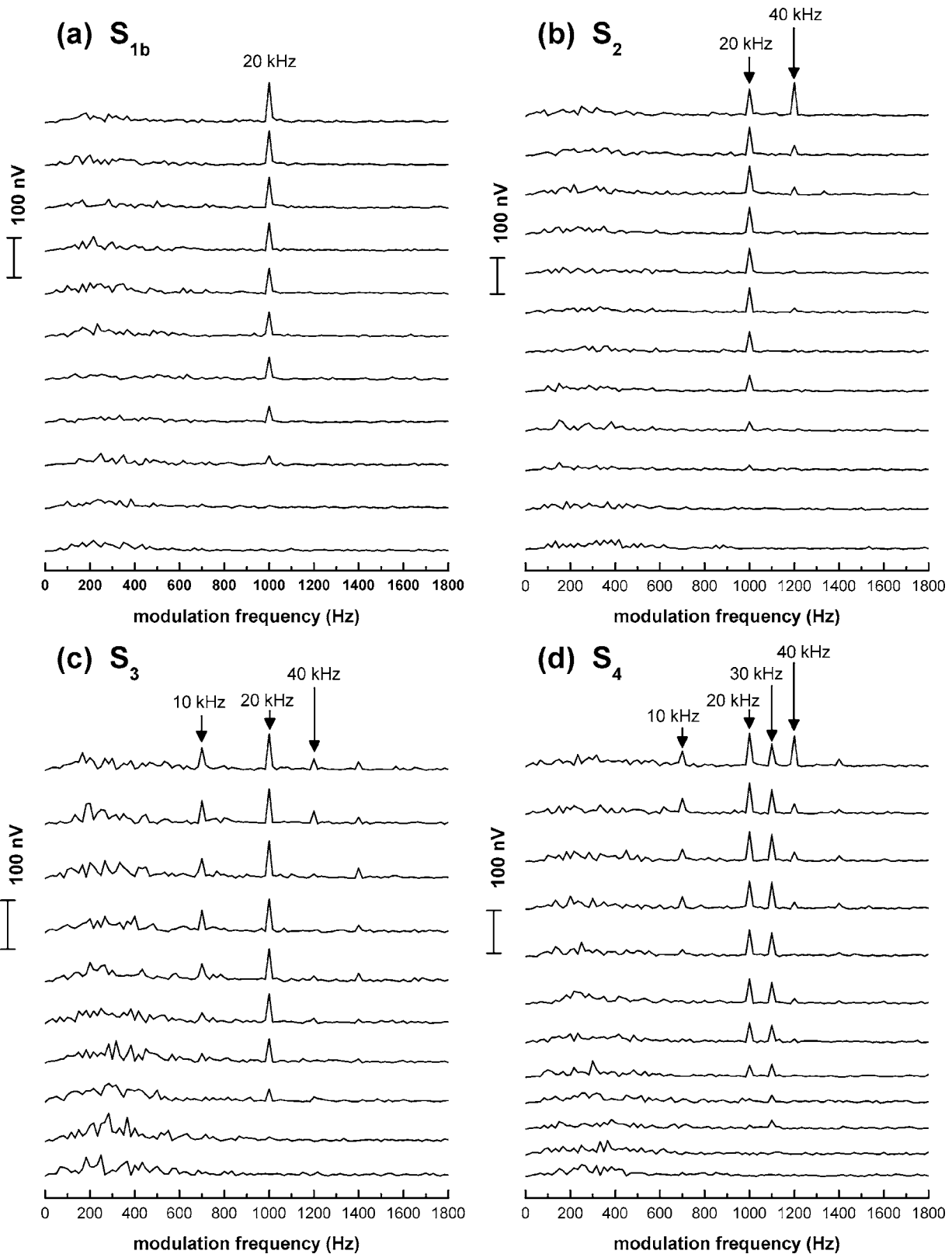


FIG. 2. Evoked response frequency spectra for stimuli (a) S_{1b} , (b) S_2 , (c) S_3 , and (d) S_4 . Stimulus frequencies and maximum SPL for each component are listed in Table I. The stimulus SPL was reduced by 5 dB for each successive recording (from top to bottom within each panel). The frequency resolution is 16.7 Hz.

marine mammals) possess a greater audible range than humans, so it may be possible to test more than four frequencies simultaneously, even while maintaining a full octave separation. In fact, Dolphin (1996) showed the dolphin audi-

tory system capable of following up to six primary envelope components created from MT stimuli (though potential interactions between the stimuli were not assessed). Whether the octave separation guideline is applicable to marine mammals

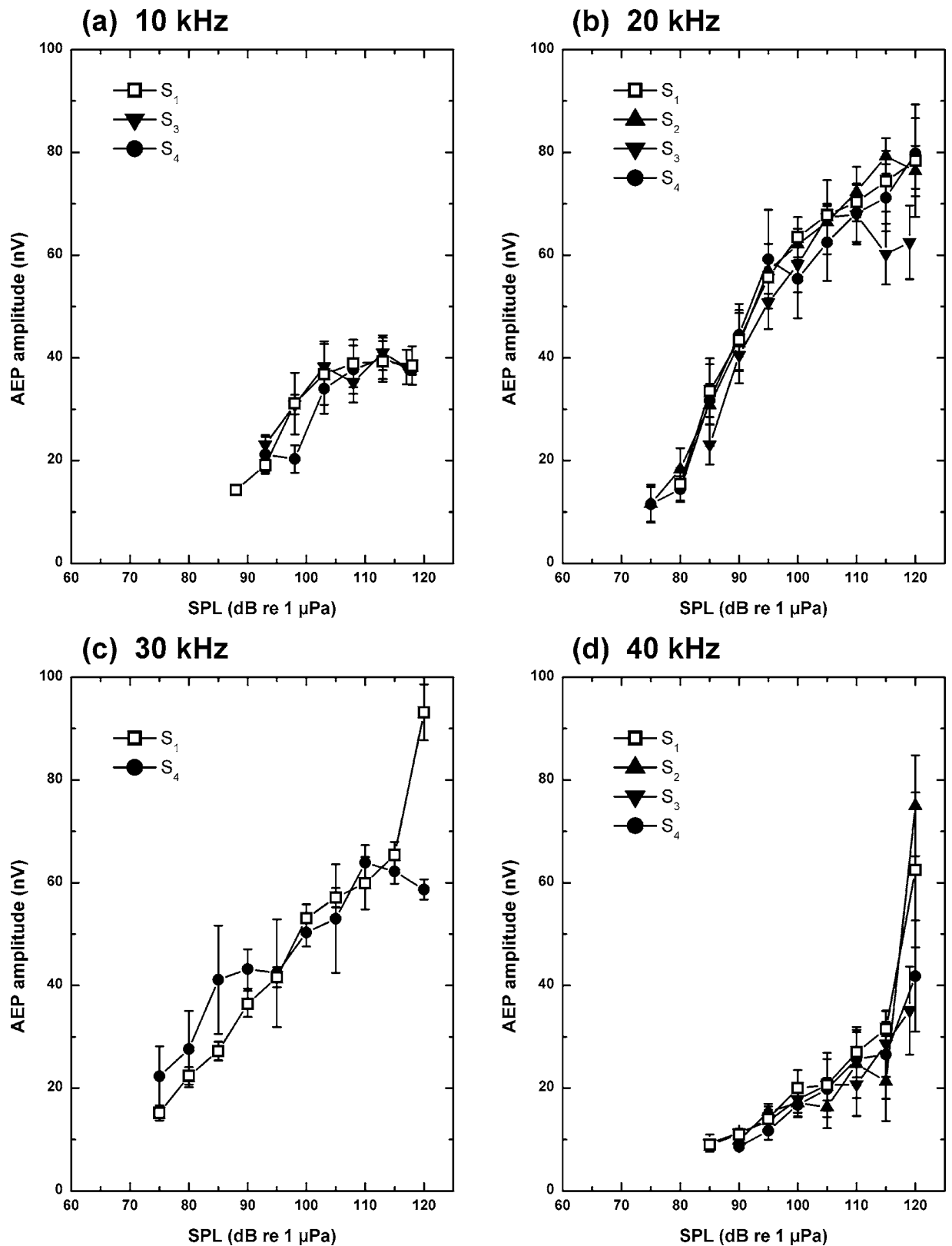


FIG. 3. AEP amplitudes as functions of stimulus component SPL for (a) 10 kHz, (b) 20 kHz, (c) 30 kHz, and (d) 40 kHz carrier frequencies. The mean \pm standard error is shown. For each sample, only detected responses ($MSC > MSC_{crit}$) were used ($2 \leq n \leq 4$).

is unknown. Popov *et al.* (1997; 1998) have shown that SPL and frequency-dependent interactions occur for simultaneous SAM tones with carrier frequencies of 76 and 85 kHz, about

1/6-octave separation. Pure tones one octave or higher did not significantly affect the ASSR amplitudes to SAM tones at lower frequency (Popov *et al.*, 1998).

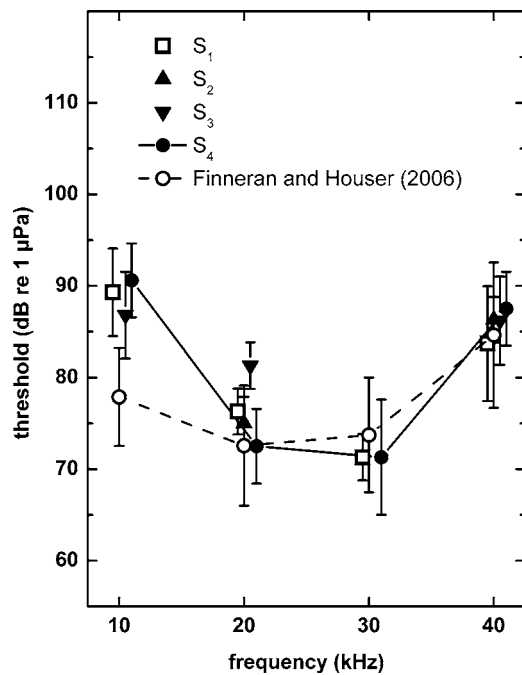


FIG. 4. AEP hearing thresholds versus sound frequency. The mean \pm standard deviation is shown. Frequency values are offset slightly for clarity. Individual thresholds were defined as the average SPL of the lowest detected/undetected response pair. AEP thresholds for BLU from Finneran and Houser (2006), based on a regression technique, are shown for comparison.

The current study represents a preliminary effort to verify the presence of the ASSR in dolphins presented with stimuli consisting of single and multiple SAM tones. Responses at the individual modulation rates were clearly observed using the recording and analysis parameters. Mean input-output functions corresponding to individual carrier frequencies were similar regardless of the number of SAM components, although there was a good deal of variability, especially at the higher SPLs. Increasing the number of averages or improved frequency resolution may have reduced some variability and allowed AEPs to be detected at lower stimulus SPLs for some stimuli. In fact, in human studies, it is common when defining thresholds by the lowest detectable response to require more averages to identify the absence of a response than the presence of a response (John *et al.*, 1998). In the present study the averaging time was limited by the maximum session time and the desire to collect data for all seven stimuli within the same session (to minimize differences in electrode and jawphone position).

Thresholds obtained with single stimuli were within ± 4 dB of those previously measured for BLU using AEP techniques and a jawphone (Finneran and Houser, 2006), except at 10 kHz, where thresholds were 11 dB higher. Some of the differences observed are likely a result of the different manner in which thresholds were defined: Finneran *et al.* (2006) used a regression technique, while the present study used the mean SPL of the lowest detected/undetected response pair. Depending on the slope of the input-output function, the two techniques could result in substantially different results. The particular modulation rates may also have affected the thresholds. Finneran and Houser (2006) used a

1-kHz modulation rate for all carrier frequencies, while the present study used a 700-Hz modulation rate at 10 kHz. Since background electrophysiological noise tends to increase with decreasing frequency, noise levels at 700 Hz would tend to be higher than those at 1 kHz, making the 10-kHz responses more difficult to detect and perhaps resulting in higher thresholds. Better results may be obtained in the future using a modulation rate closer to 1 kHz; with the 16.7-Hz analysis frequency resolution, responses could have been resolved with frequency separation of 50 Hz, allowing the modulation rates to be clustered more tightly around a more optimal region of the MRTF.

Thresholds obtained in the present study agreed closely, regardless of the number of SAM components. At 10, 30, and 40 kHz, all thresholds were within 4 dB of each other. Thresholds at 20 kHz obtained with S_{1b}, S₂, and S₄ were also within 4 dB; however, the 20-kHz threshold obtained with S₃ was significantly higher (8.8 dB) than the S₄ threshold. The reason for this difference, and no other differences, is not clear. Interactions between stimuli are complex and not symmetric—tones may have different effects on lower frequencies compared to higher frequencies with the same frequency separation. Although masking theory predicts that low-frequency tones would have a greater effect on high-frequency tones than vice versa, John *et al.* (1998) and Popov *et al.* (1997; 1998) observed the opposite effect in humans and dolphins, respectively. Overall, the present data indicate that the use of the multiple ASSR technique may not affect the resulting thresholds in some circumstances, though more data, especially from individuals with good high-frequency hearing, are needed to identify the manner of interactions between SAM tones in dolphins.

V. CONCLUSIONS

(1) Sound stimuli consisting of multiple SAM tones elicit evoked responses in bottlenose dolphins that follow the complex envelope of the stimulus. The response to each individual SAM tone can be determined by analyzing the evoked response at the SAM tone's unique modulation frequency.

(2) The multiple ASSR technique may be particularly advantageous for basic audiometry and TTS testing in marine mammals, where access to subjects is generally limited and/or testing speed is critical.

(3) Systematic effort is needed to establish limits for the number of test frequencies and required carrier frequency separation to achieve results comparable to those obtained with single SAM stimuli. Data need to be obtained in multiple individuals, preferably those with a full range of hearing.

ACKNOWLEDGMENTS

We thank Carolyn Melka, Randall Dear, Linda Green, and Laura Lewis for technical assistance and animal training. Financial support was provided by The US Office of Naval Research Marine Mammal S&T Program and the Space and Naval Warfare Systems Center, San Diego In-House Laboratory Independent Research (ILIR) Program.

- Amos, D. E., and Koopmans, L. H. (1963). *Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes* (Sandia Corporation, Livermore, California), p. 328.
- Brillinger, D. R. (1978). "A note on the estimation of evoked response," *Biol. Cybern.* **31**, 141–144.
- Bullock, T. H., Grinnell, A. D., Ikezono, E., Kameda, K., Katsuki, K., Nomoto, M., Sato, O., Suga, N., and Yanagisawa, K. (1968). "Electrophysiological studies of central auditory mechanisms in cetaceans," *Zeitschrift für Vergleichende Physiologie* **59**, 117–156.
- Bullock, T. H., and Ridgway, S. H. (1972). "Evoked potentials in the central auditory system of alert porpoises to their own and artificial sounds," *J. Neurobiol.* **3**, 79–99.
- Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., Houser, D., and Mann, D. A. (2006). "Beaked whale auditory evoked potential hearing measurements," *J. Comp. Physiol., A* **192**, 489–495.
- Cook, M. L. H., Wells, R. S., and Mann, D. A. (2004). "Auditory brainstem response hearing measurements in free-ranging bottlenose dolphins (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **116**, 2504.
- Dobie, R. A. (1993). "Objective response detection," *Ear Hear.* **14**, 31–35.
- Dobie, R. A., and Wilson, M. J. (1989). "Analysis of auditory evoked potentials by magnitude-squared coherence," *Ear Hear.* **10**, 2–13.
- Dobie, R. A., and Wilson, M. J. (1996). "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," *J. Acoust. Soc. Am.* **100**, 2236–2246.
- Dolphin, W. F. (1996). "Auditory evoked responses to amplitude modulated stimuli consisting of multiple envelope components," *J. Comp. Physiol., A* **179**, 113–121.
- Dolphin, W. F., Au, W. W., Nachtigall, P. E., and Pawloski, J. (1995). "Modulation rate transfer functions to low-frequency carriers in three species of cetaceans," *J. Comp. Physiol., A* **177**, 235–245.
- Finneran, J. J., and Houser, D. S. (2006). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," *J. Acoust. Soc. Am.* **119**, 3181–3192.
- Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (2000). "Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and a beluga whale (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions," *J. Acoust. Soc. Am.* **108**, 417–431.
- Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (2002a). "Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched noise," *J. Acoust. Soc. Am.* **112**, 322–328.
- Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (2002b). "Temporary shift in masked hearing thresholds (MTTS) in odontocetes after exposure to single underwater impulses from a seismic watergun," *J. Acoust. Soc. Am.* **111**, 2929–2940.
- GraphPad Software (2003). "GraphPad Prism," (GraphPad Software, San Diego, CA).
- Houser, D. S., and Finneran, J. J. (2006). "A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods," *J. Acoust. Soc. Am.* **120**, 1713–1722.
- John, M. S., Lins, O. G., Boucher, B. L., and Picton, T. W. (1998). "Multiple auditory steady-state responses (MASTER): stimulus and recording parameters," *Audiology* **37**, 59–82.
- John, M. S., Purcell, D. W., Dimitrijevic, A., and Picton, T. W. (2002). "Advantages and caveats when recording steady-state responses to multiple simultaneous stimuli," *J. Am. Acad. Audiol.* **13**, 246–259.
- Lemonds, D. W., Au, W. W. L., Nachtigall, P. E., Vlachos, S., and Roitblat, H. L. (1997). "Auditory frequency selectivity and masked hearing capabilities in an Atlantic bottlenose dolphin," *J. Acoust. Soc. Am.* **102**, 3102(A).
- Lins, O. G., and Picton, T. W. (1995). "Auditory steady-state responses to multiple simultaneous stimuli," *Electroencephalogr. Clin. Neurophysiol.* **96**, 420–432.
- Lins, O. G., Picton, T. W., Boucher, B. L., Durieux-Smith, A., Champagne, S. C., Moran, L. M., Perez-Abalo, M. C., and Martin V, S. G. (1996). "Frequency-specific audiometry using steady-state responses," *Ear Hear.* **17**, 81–96.
- Nachtigall, P. E., Supin, A. Y., Pawloski, J., and Au, W. W. L. (2004). "Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials," *Marine Mammal Sci.* **20**, 673–687.
- Nachtigall, P. E., Yuen, M. M. L., Mooney, T. A., and Taylor, K. A. (2005). "Hearing measurements from a stranded infant Risso's dolphin, *Grampus griseus*," *J. Exp. Biol.* **208**, 4181–4188.
- Popov, V. V., and Supin, A. Y. (1985). "Determining hearing characteristics in dolphins using evoked potentials of brain stem," *Dokl. Akad. Nauk SSSR* **283**, 496–499.
- Popov, V. V., and Supin, A. Y. (1990). "Auditory brainstem responses in characterization of dolphin hearing," *J. Comp. Physiol., A* **166**, 385–393.
- Popov, V. V., Supin, A. Y., and Klishin, V. O. (1997). "Paradoxical lateral suppression in the dolphin's auditory system: weak sounds suppress response to strong sounds," *Neurosci. Lett.* **234**, 51–54.
- Popov, V. V., Supin, A. Y., and Klishin, V. O. (1998). "Lateral suppression of rhythmic evoked responses in the dolphin's auditory system," *Hear. Res.* **126**, 126–134.
- Ridgway, S. H., Bullock, T. H., Carder, D. A., Seeley, R. L., Woods, D., and Galambos, R. (1981). "Auditory brainstem response in dolphins," *Neurobiology* **78**, 1943–1947.
- Ridgway, S. H., and Carder, D. A. (2001). "Assessing hearing and sound production in cetaceans not available for behavioral audiograms: experiences with sperm, pygmy sperm, and gray whales," *Aquat. Mamm.* **27**, 267–276.
- Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (2000). "Temporary shift in masked hearing thresholds of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones," *J. Acoust. Soc. Am.* **107**, 3496–3508.
- Supin, A. J., Popov, V. V., and Mass, A. M. (2001). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic Publishers, Boston).
- Supin, A. Y., and Popov, V. V. (1995). "Envelope-following response and modulation transfer function in the dolphin's auditory system," *Hear. Res.* **92**, 38–46.
- Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (1999). "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms," *J. Acoust. Soc. Am.* **106**, 1134–1141.
- Wells, R. S., and Scott, M. D. (1990). "Estimating bottlenose dolphin population parameters from individual identification and capture-release techniques," in *Individual Recognition of Cetaceans: Use of Photo-identification and Other Techniques to Estimate Population Parameters*, edited by P. S. Hammond, S. A. Mizroch, and G. P. Donovan (International Whaling Commission, Cambridge), pp. 407–415.
- Wolski, L. F., Anderson, R. C., Bowles, A. E., and Yochem, P. K. (2003). "Measuring hearing in the harbor seal (*Phoca vitulina*): Comparison of behavioral and auditory brainstem response techniques," *J. Acoust. Soc. Am.* **113**, 629–637.
- Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (2005). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," *J. Acoust. Soc. Am.* **118**, 2688–2695.

Comodulation detection differences in the hooded crow (*Corvus corone cornix*), with direct comparison to human subjects

Kenneth Kragh Jensen^{a)}

Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

(Received 13 July 2006; revised 4 December 2006; accepted 19 December 2006)

Envelope modulations have been shown important in determining the effectiveness of masking noises. For example, the threshold for detecting a signal flanked by maskers is lower if the maskers and the signal are modulated with different envelopes, rather than the same envelope (comodulation). This threshold change is called the comodulation detection difference (CDD). CDDs were studied in two wild-caught hooded crows, using a 1.5 kHz signal and two maskers at 0.9 and 2.1 kHz, presented at an overall level of 55 dB SPL (re 20 μ Pa). For direct comparison with human psychophysics, three human subjects were tested in the same setup. CDDs averaged 15 dB for the two crow subjects and 11 dB for the human subjects. The species difference between average CDDs was insignificant. The significance of the CDD effect in a natural setting is discussed.

© 2007 Acoustical Society of America. [DOI: 10.1121/1.2434246]

PACS number(s): 43.80.Lb, 43.66.Gf [JAS]

Pages: 1783–1789

I. INTRODUCTION

Acoustic communication plays a vital part in the behavioral ecology of birds. Birds sing and call in behaviorally important contexts such as group cohesion, individual recognition, mate attraction, territorial defense, and expression of aggression, fear, or sexual arousal. As a communication medium, sound has the advantages that a lot of information can be coded in both the spectral and temporal domains; it transmits far and fast; and natural obstacles like grass, trees, and bushes do not significantly obstruct it, within limits (Bradbury and Vehrencamp, 1998). Animals often communicate over long distances in noisy natural environments and extracting relevant information from background noise presents a challenge (Brumm and Slabbekoorn, 2005; Klump, 1996; Wiley and Richards, 1978, 1982). Songbirds communicate between (Catchpole and Slater, 1995) and even across territories (McGregor and Dabelsteen, 1996), which are generally 50–100 m in diameter for the smaller songbirds (Cramp, 1988). Thus, environmental noise and song degradation make song detection and discrimination important for avian survival and reproductive success (Brenowitz, 1982; Klump, 1996; Lohr *et al.*, 2003; Naguib and Wiley, 2001; Wiley and Richards, 1982). These acoustic constraints put special demands on both the sending and receiving parties (Brumm and Slabbekoorn, 2005). The sending party might strategically choose a position a certain height above the ground, from which sound transmits with minimized degradation (Dabelsteen *et al.*, 1993). The calling bird could also adjust its signal duration such that the receiving party achieves optimal temporal integration (Brumm and Slabbekoorn, 2005). The receiving party, which is the focus of the present paper, can similarly strategize or adapt to improve detection and discrimination. For instance, the receiver can move to a position relative to the sender that optimizes the signal-to-noise

ratio (Brumm and Slabbekoorn, 2005). Over evolutionary time, hearing capabilities may adapt for detection and discrimination in noise (Dooling *et al.*, 2000; Dooling and Saunders, 1975; Dooling and Searcy, 1979; Klump, 1996; Langemann *et al.*, 1998).

One interesting ability of the auditory system when it comes to detection is the exploitation of temporal fluctuations of masking noises especially when these are common across frequencies (see, e.g., Grose *et al.*, 2005). One such effect is the well-known *comodulation masking release* (CMR) where for instance a pure tone signal masked by an on-frequency masking noise and flanked by one or more additional maskers is easier to detect if all maskers are modulated simultaneously in amplitude (correlated) compared to when the on-frequency masker is modulated differently from the flanking maskers (uncorrelated). Similarly, in the detection of a narrow noise band signal masked by two flanking narrow noise bands, there is a release of masking if the two flanking bands are amplitude-modulated coherently but independently from the noise signal (uncorrelated) compared to if the flanking maskers and the signal are coherently modulated (correlated). The difference in detection threshold between these two conditions is termed the *comodulation detection difference* (CDD) (McFadden, 1987) and is the focus of the current study. Note that detection is best in the correlated condition in CMR but worse in CDD and vice versa.

The auditory system is modeled as a series of overlapping auditory filters (Moore, 2003). Detection in CMR and CDD is presumed to depend upon both integration of information between these filters (*across-channel cues*) and/or on information within a single filter (*within-channel cues*). In CDD within-channel cues like listening for the signal during masker envelope dips or a release from signal suppression during low masker amplitude have been suggested (Borrill and Moore, 2002; Moore and Borrill, 2002). The across-channel processes in CDD have focused on disadvantageous perceptual grouping processes (Borrill and Moore, 2002; Grose *et al.*, 2005; Hall *et al.*, 2006) where the signal is

^{a)}Electronic mail: kkj@jensenkk.net

difficult to detect in the correlated condition because it is perceptually fused (Bregman, 1994) with the flanking bands due to the common envelope fluctuation. Oppositely, in the uncorrelated condition the signal is more easily perceptually segregated (Bregman, 1994) from the flanking bands when the envelope fluctuations are not common.

Borrill and Moore (2002) and Moore and Borrill (2002) demonstrated that in CDD many aspects can be explained by a within-channel process like dip listening or suppression release rather than perceptual grouping/segregation processes. However, Hall *et al.* (2006) demonstrated that if the maskers were frequency roved and thus minimizing “energetic masking” perceptual grouping/segregation was demonstrated to play an important role in determining the size of the CDD. In contrast to Moore and Borrill (2002) they also found that even with fixed maskers the CDD of some listeners seemed to be affected by perceptual grouping/segregation processes. This means that depending on the listener and the condition CDD can apparently depend on both perceptual processes and/or more peripheral within-channel cues.

Hooded crows live year-round as monogamous breeding pairs in large territories ranging up to 5–700 m in diameter. The crows communicate between territories (Coombs, 1978; Jensen and Klokke, 2006), adding up to large distances compared to those faced by the smaller songbirds. In their habitat, hooded crows are especially challenged with detecting and discriminating faint long-distance vocalizations in environmental noise. This is reflected in the loudness, stereotypy, repetitiveness, long duration (typically 3–400 ms), and broadband nature of their calls, in contrast to the elaborate and complex vocalizations of many smaller songbirds. The natural situation of a crow detecting distant crow calls embedded in environmental noise shows similarities to an experimental CDD setup. Hooded crow calls are “noisy” in quality and after long-distance transmission (>300 m) they are more narrowband with a well-defined peak around 1.6 kHz and a 3 dB bandwidth of 450 Hz (Jensen and Klokke, 2006). Importantly, the call envelope shows short-term amplitude fluctuations at frequencies between approximately 25 and 75 Hz (Laiolo and Rolando, 2003). Sound that propagates through atmospheric turbulence exhibits marked amplitude fluctuations (Daigle *et al.*, 1983; Embleton, 1996) predominately at rates below 50 Hz (Richards and Wiley, 1980; Wiley and Richards, 1982). Thus, flanking environmental noise is likely to show some amplitude modulations (Klump, 1996). Since a major part of the amplitude modulation of crow calls is generated at the source, the amplitude modulation of the crow calls is expected to differ from that of the environmental noise. Thus like the experimental signal in the CDD setup, the natural signal to be detected is amplitude-modulated and flanked by (as well as embedded in) masking environmental noise that shows incoherent amplitude modulations (see Fig. 1). The author wished to determine whether hooded crows show a CDD effect and whether the effect is similar in magnitude to that found in humans.

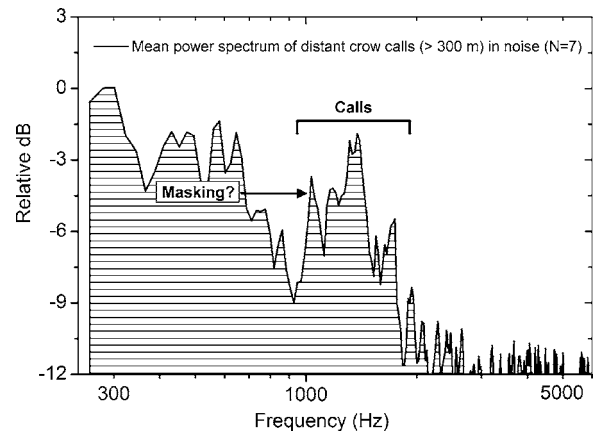


FIG. 1. Average power spectrum of hooded crow calls embedded in natural noise ($n=7$). The region of the power spectrum representing calls are indicated by the “calls” label. The “masking?” label indicates possible masking from low-frequency ambient noise. Modulation of the noise by, e.g., atmospheric turbulence, and source-generated modulation of the crow calls may resemble modulations in a laboratory CDD setup, causing a natural CDD effect. [Crow calls were recorded in open field habitats near Løkken, Denmark, at distances greater than 300 m, using a Sony DAT recorder (TCD-D10 Pro) and a handheld Beyer Dynamics shotgun microphone (MC737 PV)].

II. MATERIAL AND METHODS

A. Experimental subjects

Tests were performed on two hooded crows (labeled SP and CO) wild-caught as nestlings on the island Falster in Denmark in June, 2001. They were housed in an outside aviary with water ad libitum. Food was restricted to motivate them for the psychoacoustic testing. The crows were very experienced as psychoacoustic test subjects, since they had been subjects in studies of hearing sensitivity and critical ratios (Jensen and Klokke, 2006). Three human subjects aged 24 to 31 years were tested in the same setup [initials SK, CB, and KKJ the (author)]. Crow subject CO was first tested in the uncorrelated condition and then in the correlated one (see the following) and the order was reversed for crow subject SP. SP died from disease shortly after the experiments, making further elucidation of CDD in hooded crows untenable.

B. Test cage

Tests were performed in a custom-built soundproof test cage made of gypsum boards and mineral wool [for details see Jensen and Klokke (2006)].

C. Stimuli

The signal to be detected was a 20-Hz-wide noise band centered at 1.5 kHz. This signal frequency was chosen to be near the peak value of hooded crow calls (Jensen and Klokke, 2006), and it is also the signal frequency most often used in recent CDD studies on humans (Borrill and Moore, 2002; Moore and Borrill, 2002). Two 20-Hz-wide masking bands were centered at 0.9 and 2.1 kHz with an overall level of 55 dB SPL (which throughout the text is re 20 μ Pa), corresponding to a spectrum level of 42 dB/Hz [which throughout the text refers to an intensity spectrum density re

10^{-12} W/m² Hz (Moore (2003)]. This combination of frequencies produced some of the largest CDDs in the tests on human subjects (Borrill and Moore, 2002; Moore and Borrill, 2002).

The signal and masking bands were generated in a Real Time Processor [Tucker-Davis Technologies (TDT), type RP2.1] by multiplying a sinusoid of 1.5 kHz, 900 Hz, and 2.1 kHz, with a 10 Hz low-pass noise. This resulted in noises with a 3 dB bandwidth of 20 Hz centered at the sinusoid frequency as measured at the test position by a microphone (Brüel & Kjær, type 4133) and a frequency analyzer (Hewlett Packard, type 35665 A). The noise bands were thus amplitude-modulated as determined by the envelope of the low-pass noise. The tests were performed either by multiplying the signal and the two flanking bands by the same noise source (correlated); or by multiplying the signal by one noise source and the two flanking bands by another, independent source (uncorrelated; the two noise generators were each assigned different seed values). Threshold was estimated for both the correlated and uncorrelated conditions, and the CDD was calculated as the difference between the two thresholds.

To mask combination bands (Borrill and Moore, 2002; Moore and Borrill, 2002) which could potentially cue detection, a continuous, low-pass, Gaussian white noise was added at 400 Hz with a spectrum level of 28 dB/Hz using a noise generator (Hewlett Packard, type 3722A) filtered by a Stanford Research Systems filter (model SR650; 115 dB/oct roll-off). The spectrum level was adjusted to be flat within a 5 dB range by a graphic equalizer (*dbx* type 131).

The stimulus output was fed to a programmable attenuator (TDT, type PA5) to control levels. The signal, masking bands, and low-pass noise were then fed into a mixer connected to a stereo power amplifier (TDT, type SA1) and into a speaker (Vifa MG10MD09-08) in the test cage. The speaker was mounted approximately 50 cm directly above the position of the experimental bird's head during testing.

The flanking maskers at 0.9 and 2.1 kHz, as well as the combination band masker, were played back continuously. The noise signal had a 400 ms duration including a 5 ms rise and fall time (\cos^2 -gate) and was repeated in pulses separated by 30 ms (cf. Dooling and Okanoya, 1995). The sound level was calibrated at least once a day by a Brüel & Kjær sound level meter (type 2235) set to "linear weighting" and "random incidence," using a $\frac{1}{2}$ -in. Brüel & Kjær microphone (type 4176) perpendicular to the incoming sound, in the position normally occupied by the bird's head during testing. For further details including the variance in the sound field around the crow's head, please see Jensen and Klokke (2006).

The human subjects were tested in the same setup, sitting on the cage floor and resting their heads just above the test panel [see Jensen and Klokke (2006)]. The sound field was calibrated 5 cm in front of the meatus at both ears. The highest sound pressure value at the two ears was chosen as the calibration sound level. The sound pressure level at the two ear positions never varied more than 2.5 dB SPL. The noise level of the combination tone masker at the calibration position was adjusted to be flat within a 5 dB range.

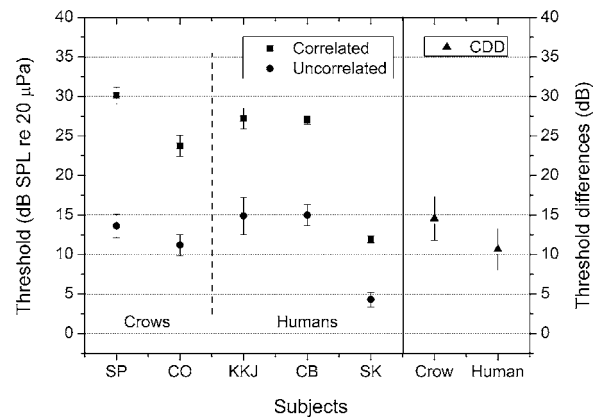


FIG. 2. Individual thresholds and comodulation detection differences (CDD) for two crows (SP and CO) and for three human subjects (KKJ, CB, and SK). The signal to be detected was a 20 Hz wide noise centered at 1.5 kHz (3 dB BW). The signal was masked by two flanking noises centered at 0.9 and 2.1 kHz, with an overall level of 50 dB SPL. A 400 Hz low-pass Gaussian noise was added at a spectrum level of 28 dB/Hz to mask potentially disturbing combination tones. "Correlated" indicates thresholds when the signal and the two maskers were modulated coherently, and "Uncorrelated" when the signal and maskers were modulated incoherently. The CDD is the threshold difference between the uncorrelated and correlated conditions. Error bars are standard deviations.

D. Psychoacoustic procedure and threshold determination

The test procedure for both the crows and the human subjects followed that of Jensen and Klokke (2006), who used a *yes/no procedure* with the *Method of Constant Stimuli* using five constant sound levels separated by a 3 dB step. The middle tone level was adjusted to lie around the assumed threshold before each session. The test signal or silence was presented with an equal probability of occurrence. Crows initiated a trial by pecking the middle key. They reported "yes" by pecking the right key if they heard a tone, and "no" by pecking the left. Correct answers (hits and correct rejections) were rewarded by food pellets, and wrong answers [false alarms (FAs) and misses] were punished by a 10–20 s blackout of cage illumination. From the hit percentages and FAs, d' values were calculated and plotted as a function of signal level. The thresholds were then read off at $d'=1$ by linear interpolation. The crows had unlimited response times and the threshold was corrected according to deferred decision theory (Birdsall and Roberts, 1965; Green and Swets, 1974; Jensen and Klokke, 2006; Swets and Birdsall, 1967). A single threshold was determined in a session of at least 200 trials. The final thresholds were averages of three consecutive sessions lying within a 3 dB range and not monotonically decreasing (i.e., based on a minimum of 600 total trials).

III. RESULTS

The average signal thresholds for individual crow and human subjects are shown in Fig. 2 for both the correlated and uncorrelated flanking band conditions. The right section of Fig. 2 shows the average differences between the thresholds, i.e., the CDDs, for the crow and human subjects (the rightmost part of Fig. 2). In the correlated condition, the

crows SP and CO showed average thresholds of 30 and 24 dB SPL, respectively. In the uncorrelated condition, SP and CO had average thresholds of 14 and 11 dB SPL, respectively. The CDDs for SP and CO were thus 16 and 13 dB, with a group average of 15 dB.

The human thresholds under the correlated condition were 27, 27, and 12 dB SPL for subjects KKJ, SK, and CB, respectively. Under the uncorrelated condition, thresholds were 15, 15, and 4 dB SPL for the same subjects, respectively. Thus the CDD was 12, 12, and 8 dB, respectively, for subjects KKJ, SK, and CB, with a group average CDD of 11 dB.

The mean CDD for crows and humans fall within 1.4 s.d. of each other, and the difference is nonsignificant ($P=0.19$; t-test) (Jandel Scientific, 1995). However, the power of the test is low due to the small n (power=0.144).

IV. DISCUSSION

This experiment demonstrates that hooded crows show a CDD effect and suggest by direct comparison that this is not significantly different from that in humans. In an abstract, Langemann *et al.* (2005) reported a CDD up to an average of 5 dB in the European starling (*Sturnus vulgaris*) using a 100 Hz wide, 400 ms duration signal band centered at 2 kHz masked by six 100 Hz flanking bands at a spectrum level of 50 dB/Hz. This CDD is close to human values in a roughly comparable study (McFadden and Wright, 1990), further supporting the hypothesis that passerine and human CDDs are similar.

The CDD effects obtained in the current study were generally large, however, the thresholds in the uncorrelated condition were relatively low for all subjects and in particular for the human subject SK. This may indicate that the CDDs may have been limited by the absolute thresholds. In another study (Jensen and Klokke, 2006) the pure tone absolute thresholds tested, which were closest to the tested 1.5 kHz signal in the current study, were for subject SK -2 dB SPL at 1 kHz, 2 dB SPL for subject KKJ at 1 kHz, and 0 and 6 dB SPL at 0.375 and 5.6 kHz, respectively, for subject CB. Since human free-field hearing thresholds generally tends to be lower at 1.5 kHz than at the tested frequencies at 0.375, 1, and 5.6 kHz (ISO 389-7:1996) this suggests that the thresholds in the uncorrelated condition were approximately 13 dB above absolute threshold for both subject CB and KKJ and approximately 6 dB for subject SK. For the crow subjects SP and CO the pure tone absolute thresholds were measured to be 2 and 1 dB SPL, respectively, at 1.4 kHz (Jensen and Klokke, 2006) suggesting that thresholds in the uncorrelated condition were approximately 12 and 10 dB above threshold. According to certain models on the additivity of masking there can be a relative large effect on the masked threshold due to the additivity of the "internal noise" corresponding to the absolute threshold, depending on the compression factor P (Humes *et al.*, 1988; Humes and Jesteadt, 1989), whereas oppositely, according to a simple model of linear energy summation there is little effect 6 dB above threshold (Humes and Jesteadt, 1989). However, without data on the exact absolute threshold for the signal noise band and the P factor in

a comodulation task it is inappropriate to make any quantitative evaluations and it suffices here to make the notion. Nevertheless, CDD in closeness of the absolute or masked threshold is most relevant to crow territorial communication when detecting barely audible calls in a natural setting and it may be worthy to investigate the effect of masking additivity further.

A. Comparison to CDD values in the literature

CDDs from studies similar to the current one are shown in Table I. To ensure comparability, data were selected with the following criteria: The stimulus had only one signal band; maskers were not roved in frequency; the signal duration was long (≥ 240 ms); the signal center frequency (f_c) did not differ by more than 1 kHz from our 1.5 kHz signal; and the frequency deviation (Δf) between the signal and the first masking band(s) ranged from 400 to 600 Hz. The measurements are sorted by spectrum level in Table I. Table I shows a trend that the CDD increases with the spectrum level and increasing numbers of flanking bands, both within and across studies. However, the CDD of the current study (Experiment 2) and that of Cohen and Schubert (1987) (Experiment 8) do not conform to this pattern since they produced large CDDs with a low spectrum level and a low number of bands. The contradiction may be due to the application of multiplied noise instead of Gaussian noise.

To evaluate the significance of the noise type and the other factors in Table I, a multiple regression with backward elimination was performed (Jandel Scientific, 1995; Zar, 1996). Each noise type was assigned an arbitrary number in order to perform the numerical test. The result showed, with a multiple regression correlation coefficient of $R=0.96$, a power of 1, and a passing normality and constant variance test, that the significant variables influencing the CDD were the spectrum level ($P<0.001$), bandwidth ($P=0.018$), number of bands ($P<0.001$), and noise type ($P<0.001$). Thus, the noise type significantly effects the CDD and therefore seems to explain the larger CDDs obtained with multiplied noises in the current study and that of Cohen and Schubert (1987), despite low spectrum levels and a low number of masking bands.

In addition, the multiple regression is an across-study, statistical support of the general findings that the spectrum level and the number of bands have a significant effect on the CDD (Borrill and Moore, 2002; McFadden, 1987; Moore and Borrill, 2002). The effect of bandwidth was probably significant due to its effect on the overall level of the maskers, rather than due to the consequences of bandwidth alone, since this has been found not to influence the CDD (Fantini and Moore, 1994; Moore and Borrill, 2002). As assumed in our selection criteria, the frequency deviation (Δf) and the signal center frequency (f_c) did not show a significant effect, confirming the comparability of studies using frequencies within this range.

Schooneveldt and Moore (1989a) studied some properties of masking noise in a comodulated masking release (CMR) study. In an analysis of 50 Hz wide noise maskers,

TABLE I. Comparison of comodulation detection differences (CDDs) between studies with comparable experimental setup. The data are sorted with ascending spectrum levels (N_0). BW is the bandwidth of the signal and maskers; f_c is the center frequency of the signal band; Δf is the frequency deviation between the center frequency of the signal band and the nearest maskers (when two numbers are shown the first number refers to the low frequency band and the second to the high frequency band); “Bands” refers to the total number of masking bands; “Noise type” refers to how the signal and masking bands were prepared: either by summing a number of sinusoids with random amplitude and phase drawn from a Rayleigh and rectangular distribution, respectively (“Gaussian”), or by modulation of a sinusoid by multiplication with a low pass noise (“Multiplied”).

Experiment	CDD (dB)	N_0 (dB/Hz)	BW (Hz)	f_c (Hz)	Δf (Hz)	Bands (No.)	Noise type	Study
1	7	33	30	2000	460/570	6	Gaussian	Hall <i>et al.</i> (2006)
2	11	42	20	1500	600	2	Multiplied	Current
3	4	45	20	1500	600	2	Gaussian	Borrill and Moore (2002)
4	4	50	100	2500	500	2	Gaussian	McFadden (1987)
5	9	50	100	2500	500	4	Gaussian	McFadden (1987)
6	7	50	50	1250	200	4	Gaussian	McFadden and Wright (1990)
7	6	50	100	1250	400	4	Gaussian	McFadden and Wright (1990)
8	10	53	100	1500	500	1	Multiplied	Cohen and Schubert (1987)
9	6	55	20	1500	600	2	Gaussian	Borrill and Moore (2002)
10	10	63	20	1500	429/600	2	Gaussian	Moore and Borrill (2002)
11	13	63	20	1500	429/600	6	Gaussian	Moore and Borrill (2002)
12	10	65	20	1500	600	2	Gaussian	Borrill and Moore (2002)
13	10	65	20	1500	400	2	Gaussian	Borrill and Moore (2002)
14	9	65	20	1500	600	2	Gaussian	Borrill and Moore (2002)

^aExperiment 1: taken as an average of fringed and nonfringed experiments in the fixed masker condition (cf. their Table I).

^bExperiments 3, 9, and 12: taken from their experiment 2 (see their Fig. 3); experiments 13 and 14: taken from their experiment 1 (cf. their Fig. 1).

^cExperiments 4 and 5: average of values at scrambled and nonscrambled conditions with cue bands center frequencies 2 and 3 kHz and cue bands 1.5, 2.0, 3.0, and 3.5, respectively (cf. his Fig. 6).

^dExperiment 6: average CDD of their experiment 4 (cf. their Fig. 5).

^eExperiment 7: average of all fringe durations and all individuals (cf. their Fig. 4).

^fExperiment 8: taken at signal frequency 1.5 kHz (cf. their Fig. 2).

^gExperiments 10 and 11: taken from their experiment 3 as an average of the fringed and nonfringed two band and six band flanking band condition, respectively (cf. their Fig. 5).

they showed that the envelope distribution (probability density of the envelope fluctuations relative to the rms value of the noise or the noise + signal) was greatly skewed toward lower rms values in multiplied noise as compared to Gaussian noise. In other words, the multiplied noise envelope spends more time at low values. The lower the modulator frequency, the more the distribution was skewed toward lower rms values. A CMR study by Ernst and Verhey (2005) also shows that multiplied noises spend a great amount of time at low amplitude values, compared to a Rayleigh-distributed Gaussian noise. In general, slower and larger envelope dips normally result in larger CMR effects (Verhey *et al.*, 2003). The greater amount of time that multiplied noises have low envelope values would provide more opportunities for within-channel cues like dip listening and suppression release. This clearly suggests, from a within-channel perspective, that the envelope distribution is the reason why CDDs obtained with multiplied noises are larger than those obtained with Gaussian noises.

B. Implications for the functionality of CDD in the natural environment

The current study showed that hooded crows exhibit a CDD effect similar to that observed in humans. The question

remains, how useful is the CDD to crows in a natural detection situation? Natural scenarios are probably quite similar to the current experimental setup in that crow calls peak around 1.6 kHz and are singularly modulated at the source (Laiolo and Rolando, 2003), whereas flanking environmental noises are likely to be modulated in a different pattern by atmospheric turbulence (Daigle *et al.*, 1983; Embleton, 1996; Nelken *et al.*, 1999; Piercy *et al.*, 1977; Wiley and Richards, 1978, 1982). Clearly, the crow call will be modulated in a similar manner by the atmospheric turbulence, but the dissimilar inherent modulations of crow calls will probably play an important role in segregating calls from the environmental noise if perceptual grouping/segregation is a significant part of CDD in a natural setting, whereas they are likely to be less important if within-channel cues dominates. Furthermore, even sound propagation paths that do not differ by more than a few degrees relative to the receiver will lead to very different turbulence regimes (Mcleod *et al.*, 2004). If sounds produced by a group of vocalizing animals, by a singing bird, or by rustling leaves of a tree propagate toward the receiver from one direction they will be modulated differently than, for instance, crow calls emitted from a sender in another direction. Consequently there may be many natural

situations in which the signal is modulated differently from masking sounds, resulting in a masking release like the CDD observed in the laboratory (Fig. 1).

However, there are at least three important differences between the natural condition and the experimental setup that need to be taken into consideration for future evaluation of natural communication. First, the sounds and noises come from many different directions, and therefore binaural masking level differences (BMLDs) need also to be taken into consideration. Studies on CMR in humans show that a significant CMR can occur in addition to a BMLD. It has been suggested that different cues are responsible for the two effects (Hall *et al.*, 1988; Schooneveldt and Moore, 1989b). BMLDs of up to 8 dB at 2 and 4 kHz have been demonstrated in the budgerigar (*Melopsittacus undulatus*) (Dent *et al.*, 1997). Thus when spatial separation of the target sound and masking noises is likely to produce a BMLD, birds may *in addition* make benefit of a CMR and probably also a CDD effect.

Second, in the experimental setup, envelope distributions with dips down to zero rms are possible (see earlier text). The CDD and CMR have been mostly studied under such conditions in humans. But in a natural setting there is omnipresent noise from all directions that may interfere with and limit the effective depth of the envelope dips. Thus, in the natural environment the CDD and the CMR may be different and smaller than observed in laboratories.

Third, it has been found that sounds modulated coherently in frequency is common in natural settings partly attributed to atmospheric turbulence (Nelken *et al.*, 1999). However, the modulating nature and frequency dependence of turbulence needs to be established in more detail. When turbulence comodulates sounds, does this affect a broad or limited range of frequencies?

In conclusion, the hooded crow shows a CDD of about 15 dB SPL, which is quantitatively similar to that observed in humans by direct comparison. Multiplied noises, which spend more time at low envelope values, result in larger CDDs than observed with Gaussian noises, according to multiple regression analysis. Coherently modulated sounds are common in the natural environment and there are many situations where masking environmental noise is differently modulated from the crow calls. Therefore, it seems likely that the crows can exploit the demonstrated CDD in natural listening situations.

ACKNOWLEDGMENTS

The current study was performed with permission and counseling from the Danish Ringing Centre, The Danish Forest and Nature Agency, and the Danish Animal Experiments Inspectorate. Experiments with human subjects were performed with permission from the Danish National Committee for Biomedical Research Ethics (region Vejle and Fyns Amt). The author is grateful to gamekeeper Benny Larsen for his invaluable advice and help on catching crows and Signe Klokker and Christian Brandt for volunteering as experimental subjects. Brian C. J. Moore is acknowledged for very useful comments and discussion on an earlier version of the

manuscript. Ole Næsbye Larsen, James Simmons, Taffeta Elliott, and Beth Brittan-Powell are acknowledged for valuable comments and language corrections on earlier versions of the manuscript. Work was supported by a grant from the Danish National Research Foundation to the Center for Sound Communication.

- Birdsall, T. G., and Roberts, R. A. (1965). "Theory of signal detectability - Deferred-decision theory," *J. Acoust. Soc. Am.* **37**, 1064–1074.
- Borrill, S. J., and Moore, B. C. J. (2002). "Evidence that comodulation detection differences depend on within-channel mechanisms," *J. Acoust. Soc. Am.* **111**, 309–319.
- Bradbury, J. W., and Vehrencamp, S. L. (1998). *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA).
- Bregman, A. S. (1994). *Auditory Scene Analysis - The Perceptual Organization of Sound* (MIT, Cambridge, MA).
- Brenowitz, E. A. (1982). "The active space of red-winged blackbird song," *J. Comp. Physiol.* **147**, 511–522.
- Brumm, H., and Slabbekoorn, H. (2005). "Acoustic communication in noise," *Adv. Stud. Behav.* **35**, 151–209.
- Catchpole, C. K., and Slater, P. J. B. (1995). *Bird Song - Biological Themes and Variations* (Cambridge University Press, Cambridge).
- Cohen, M. F., and Schubert, E. D. (1987). "The effect of cross-spectrum correlation on the detectability of a noise band," *J. Acoust. Soc. Am.* **81**, 721–723.
- Coombs, C. J. F. (1978). *The Crows: A Study of the Corvids of Europe* (Batsford, London).
- Cramp, S. (1988). *Handbook of the Birds of Europe, the Middle East and North Africa - The Birds of the Western Palearctic* (Oxford University Press, New York).
- Dabelsteen, T., Larsen, O. N., and Pedersen, S. B. (1993). "Habitat-induced degradation of sound signals: Quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song," *J. Acoust. Soc. Am.* **93**, 2206–2220.
- Daigle, G. A., Piercy, J. E., and Embleton, T. F. W. (1983). "Line-of-sight propagation through atmospheric turbulence near the ground," *J. Acoust. Soc. Am.* **74**, 1505–1513.
- Dent, M. L., Larsen, O. N., and Dooling, R. J. (1997). "Free-field binaural unmasking in budgerigars (*Melopsittacus undulatus*)," *Behav. Neurosci.* **111**, 590–598.
- Dooling, R. J., Lohr, B., and Dent, M. (2000). *Hearing in Birds and Reptiles*, in *Comparative Hearing: Birds and Reptiles Vol. 13*, edited by R. J. Dooling, R. R. Fay, and A. N. Popper (Springer, New York), pp. 308–359.
- Dooling, R. J., and Okanoya, K. (1995). *Psychophysical Methods for Assessing Perceptual Categories*, in *Methods in Comparative Psychoacoustics Vol. 6*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser, Basel), pp. 307–318.
- Dooling, R. J., and Saunders, J. C. (1975). "Hearing in parakeet (*Melopsittacus-undulatus*) - Absolute thresholds, critical ratios, frequency limens, and vocalizations," *J. Comp. Physiol. Psychol.* **88**, 1–20.
- Dooling, R. J., and Searcy, M. H. (1979). "Relation among critical ratios, critical bands, and intensity difference limens in the parakeet (*Melopsittacus undulatus*)," *Bull. Psychon. Soc.* **13**, 300–302.
- Embleton, T. F. W. (1996). "Tutorial on sound propagation outdoors," *J. Acoust. Soc. Am.* **100**, 31–48.
- Ernst, S. M. A., and Verhey, J. L. (2005). "Comodulation masking release over a three octave range," *Acta. Acust. Acust.* **91**, 998–1006.
- Fantini, D. A., and Moore, B. C. J. (1994). "Profile analysis and comodulation detection differences using narrow bands of noise and their relation to comodulation masking release," *J. Acoust. Soc. Am.* **95**, 2180–2191.
- Green, D. M., and Swets, J. A. (1974). *Signal Detection Theory and Psychophysics* (Krieger, Huntington, NY).
- Grose, J. H., Hall, J. W., and Buss, E. (2005). "Across-channel spectral processing," *Int. Rev. Neurobiol.* **70**, 87–119.
- Hall, J. W., Buss, E., and Grose, J. H. (2006). "Comodulation detection differences for fixed-frequency and roved-frequency maskers," *J. Acoust. Soc. Am.* **119**, 1021–1028.
- Hall, J. W., Cokely, J. A., and Grose, J. H. (1988). "Combined monaural and binaural masking release," *J. Acoust. Soc. Am.* **83**, 1839–1845.
- Humes, L. E., Espinozavaras, B., and Watson, C. S. (1988). "Modelling sensorineural hearing loss. I. Model and retrospective evaluation," *J. Acoust. Soc. Am.* **83**, 188–202.
- Humes, L. E., and Jesteadt, W. (1989). "Models of the additivity of mask-

- ing," *J. Acoust. Soc. Am.* **85**, 1285–1294.
- Jandel Scientific. (1995). SigmaStat. [2.0.173] (computer program).
- Jensen, K. K., and Klokke, S. (2006). "Hearing sensitivity and critical ratios of hooded crows (*Corvus corone cornix*)," *J. Acoust. Soc. Am.* **119**, 1269–1276.
- Klump, G. M. (1996). "Bird communication in the noisy world," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 321–338.
- Laiolo, P., and Rolando, A. (2003). "The evolution of vocalisations in the genus *Corvus*: Effects of phylogeny, morphology and habitat," *Evol. Ecol. Res.* **17**, 111–123.
- Langemann, U., Bee, M. A., and Klump, G. M. (2005). "Comodulation effects in a songbird: Comodulation detection difference," *J. Assoc. Res. Otolaryngol.* **28**, 253–254.
- Langemann, U., Gauger, B., and Klump, G. M. (1998). "Auditory sensitivity in the great tit: Perception of signals in the presence and absence of noise," *Anim. Behav.* **56**, 763–769.
- Lohr, B., Wright, T. F., and Dooling, R. J. (2003). "Detection and discrimination of natural calls in masking noise by birds; estimating the active space of a signal," *Anim. Behav.* **65**, 763–777.
- McFadden, D. (1987). "Comodulation detection differences using noise-band signals," *J. Acoust. Soc. Am.* **81**, 1519–1527.
- McFadden, D., and Wright, B. A. (1990). "Temporal decline of masking and comodulation detection differences," *J. Acoust. Soc. Am.* **88**, 711–724.
- McGregor, P. K., and Dabelsteen, T. (1996). "Communication network," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 409–425.
- McLeod, I. D., Don, C. G., and Swenson, G. G. (2004). "Acoustic pulse propagation in an outdoor turbulent atmosphere," *J. Acoust. Soc. Am.* **116**, 2855–2863.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (Academic, San Diego).
- Moore, B. C. J., and Borrill, S. J. (2002). "Tests of a within-channel account of comodulation detection differences," *J. Acoust. Soc. Am.* **112**, 2099–2109.
- Naguib, M., and Wiley, R. H. (2001). "Estimating the distance to a source of sound: Mechanisms and adaptations for long-range communication," *Anim. Behav.* **62**, 825–837.
- Nelken, I., Rotman, Y., and Bar Yosef, O. (1999). "Responses of auditory-cortex neurons to structural features of natural sounds," *Nature (London)* **397**, 154–157.
- Piercy, J. E., Embleton, T. F. W., and Sutherland, L. C. (1977). "Review of noise propagation in the atmosphere," *J. Acoust. Soc. Am.* **61**, 1403–1418.
- Richards, D. G., and Wiley, R. H. (1980). "Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication," *Am. Nat.* **115**, 381–399.
- Schooneveldt, G. P., and Moore, B. C. J. (1989a). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration," *J. Acoust. Soc. Am.* **85**, 273–281.
- Schooneveldt, G. P., and Moore, B. C. J. (1989b). "Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands," *J. Acoust. Soc. Am.* **85**, 262–272.
- Swets, J. A., and Birdsall, T. G. (1967). "Deferred decision in human signal detection - Preliminary experiment," *Percept. Psychophys.* **2**, 15–28.
- Verhey, J. L., Pressnitzer, D., and Winter, I. M. (2003). "The psychophysics and physiology of comodulation masking release," *Exp. Brain Res.* **153**, 405–417.
- Wiley, R. H., and Richards, D. G. (1978). "Physical constraints on acoustic communication in atmosphere - Implications for evolution of animal vocalizations," *Behav. Ecol. Sociobiol.* **3**, 69–94.
- Wiley, R. H., and Richards, D. G. (1982). "Adaptations for acoustic communication in birds: Sound transmission and signal detection," in *Acoustic Communication in Birds: Communication and Behavior*, edited by D. E. Kroodsma (Academic, New York), pp. 131–181.
- Zar, J. H. (1996). *Biostatistical Analysis* (Prentice-Hall, Upper Saddle River, NJ).

Singular spectrum analysis applied to ultrasonic detection and imaging of brachytherapy seeds

Jonathan Mamou^{a)} and Ernest J. Feleppa

*Frederic L. Lizzi Center for Biomedical Engineering, Riverside Research Institute,
156 William Street, New York, New York 10038*

(Received 24 August 2006; revised 22 November 2006; accepted 4 January 2007)

Ultrasound-guided brachytherapy using titanium-shelled radioactive seeds is a popular, effective means of treating prostate cancer. Unfortunately, implantation using needles inserted transperitoneally causes gland movement and distortion, which often results in seed misplacement and dosimetry errors. If actual seed locations could be determined in the operating room, then corrections to dosimetry errors could be made immediately. However, seed specularity, shadowing, and tissue clutter make imaging seeds difficult using conventional ultrasound. Singular spectrum analysis (SSA) shows promise for reliably imaging radioactive seeds implanted in the prostate and enabling additional corrective implantations to be made in the operating room. SSA utilizes eigenvalues derived from the diagonalized correlation matrix of envelope-detected radio-frequency echo signals to yield a P value indicative of the likelihood of a seed-specific repetitive signal. We demonstrated the potential of SSA for seed detection and imaging and illustrated the trade-off considerations for optimization of SSA in clinical applications using simulations assessing performance as a function of different levels of noise and the presence of repetitive signals with various repetition periods; experiments in an ideal scattering environment; and experiments using seeds implanted in beef. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2436713]

PACS number(s): 43.80.Vj, 43.60.Lq, 43.35.Yb [FD]

Pages: 1790–1801

I. INTRODUCTION

Implantation of seeds that contain a radioactive material (e.g., iodine or palladium) into the prostate is becoming a commonly used means of treating gland-confined cancer.^{1,2} Seed implantation also is gaining acceptance as a means of treating other cancers as well, e.g., breast cancer.^{3–5} The standard imaging modality for treatment-planning dosimetry and for guiding and monitoring seed implantation is transrectal ultrasound (TRUS).^{6,7} However, movement of the prostate during seed insertion, shadowing caused by hemorrhage or by other seeds, clutter from calcifications and other hyper-echogenic scattering objects, and the specularity of seeds themselves make detecting implanted seeds, determining their actual location, and verifying proper dosimetry difficult in an intraoperative time frame. As a consequence, real-time, intraoperative corrections cannot be made for insufficient treatment doses that result from seed misplacements. Currently, information regarding actual seed location typically is obtained from x-ray computed tomography or magnetic resonance imaging performed 3 h to 4 weeks after the implantation procedure.^{8–10} At that time, dose corrections can be made only through subsequent external-beam radiation therapy. Radiation oncologists would find a real-time imaging system that is capable of providing accurate, postinsertion, seed-location information to be very valuable because the information would enable timely feedback for intraoperatively correcting deficiencies in the radiation dose delivered to the gland. Implanting dose-correcting seeds in the operating room would permit an optimal treatment effect to be

achieved by brachytherapy alone. Therefore, a remaining challenge for TRUS-guided brachytherapy of prostate cancer is accurate detection and localization of the seeds upon their insertion. This article describes initial investigations of a method that may offer a means of meeting this challenge.

Figure 1 shows a cross section of a commonly used seed that contains radioactive palladium (Pd¹²⁵). Seeds used in the prostate nominally are 5 mm long by 0.8 mm in diameter. (The actual length of the Pd seed is 4.5 mm.) Like the Pd seed shown in Fig. 1, most types of prostate-brachytherapy seeds consist of a thin, rigid, cylindrical, titanium shell with enclosed radioactive material; the smooth metal shell causes strong specular reflections of ultrasound waves. However, other commonly occurring strongly reflecting entities, such as calcifications, also often exist in the prostate and may produce false-positive indications if seed detection were to be based solely on backscattered echo-signal strength. Furthermore, the cylindrical shape of most seeds makes their echo strength dependent on the angle of beam incidence. During brachytherapy, seeds are inserted into the prostate using needles that are orthogonal to the scan plane, i.e., to the ultrasound-beam axis. However, because of gland distortion and movement during insertion, an orthogonal orientation with respect to the beam axis can be difficult to maintain after needle withdrawal even when stranded seeds are used.

The difficulties in achieving optimal seed placement and dose distribution in prostate brachytherapy led us to formulate this study with the following general goal: to improve detection and imaging of brachytherapy seeds using ultrasound signals. This paper describes our initial efforts to identify a seed “signature” and to develop a straightforward, robust, signal-processing method for reliably detecting that

^{a)}Electronic mail: mamou@rrinyc.org

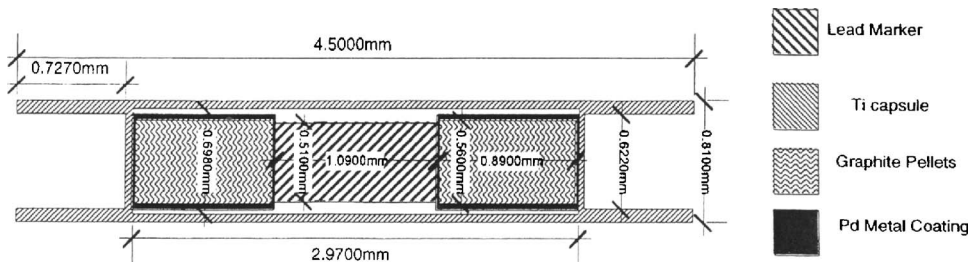


FIG. 1. Structure of the brachytherapy seeds used in our study. (Printed with the permission of American Association of Physicists in Medicine.)

signature and depicting the corresponding seed location. The article is organized into the following sections: Section II establishes that a distinctive seed signature exists; Sec. III details the signal-processing approach we have developed for signature detection; Sec. IV presents our results from simulations and *in vitro* experiments; and Sec. V discusses the study and draws relevant conclusions.

II. SEED SIGNATURE

Our initial aim in this study was to determine whether a seed signature exists that would be amenable to detection and localization. To determine whether we could detect a seed ultrasonically, we conducted a simple experiment. We inserted a seed into an acoustically transparent gel standoff pad (Aquaflex, Parker Laboratories, Fairfield, NJ). Following seed insertion, we scanned the gel pad using three, different, single-element, focused ultrasound transducers. The nominal center frequencies of the imaging transducers were 3.5, 5.0, and 10.0 MHz. These frequencies were chosen in the range of those used to clinically image prostate (near 6 MHz). An Acqiris DP-110 (Acqiris, Monroe, NY) A/D board digitized 8-bit samples of rf echo signals at a sampling rate of 50 MS/s, and a desktop computer stored the digitized signals for subsequent processing, as described in the following. Figure 2 shows the *B*-mode image reconstructed from rf echo-signal data obtained with a nominally 5-MHz transducer (V326, Panametrics, Inc., Waltham, MA). The dynamic range of the image is approximately 50 dB. The trans-

ducer has an F number of 5 with a focal distance of 5 cm and an aperture of 1 cm. The shadow below the seed in Fig. 2 suggests that the backscattered signals from the seed have a very long “tail” of repeating echoes. This phenomenon also was consistently present for all three transducers.

Figures 3(a) and 3(d) show typical backscattered rf signals from the seed acquired with the 5- and 3.5-MHz (V383, Panametrics Inc., Waltham, MA) transducers, respectively. These signals seem to show a first echo followed about $1 \mu\text{s}$ later by a second echo that is very similar in shape to the first echo. Figures 3(b) and 3(e) show the log power spectra of the backscattered signals shown in Figs. 3(a) and 3(d). The spectra contain periodic ripples that are about 1 MHz apart. Scallops in spectra indicate repetitions in the corresponding time signal; they typically are caused by interference between echoes from closely spaced parallel reflecting surfaces. A standard method to test for repetitions in time signals is to compute the cepstrum of the time signal.¹¹ (The cepstrum is the inverse Fourier transform of the log power spectrum.) Peaks in the cepstrum indicate repetitions in the time signals. The locations of the cepstral peaks in time denote the period of the repetitions. Figures 3(c) and 3(f) show the computed cepstra derived from the rf signals of Figs. 3(a) and 3(d). These cepstra contain a single, very clear peak near $1 \mu\text{s}$, which indicates that repetitions exist in the original signal with a repetition period (i.e., time separation) of approximately $1 \mu\text{s}$. In particular, both rf signals seem to have the same repetition period even though they were acquired using two different transducers. Thus, a repetition frequency near

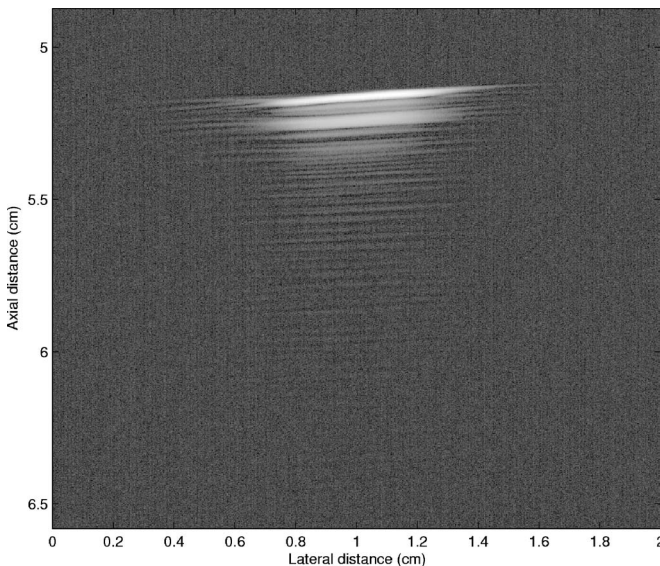


FIG. 2. *B*-mode image of a seed in an acoustically transparent gel pad obtained with a 5-MHz center-frequency transducer.

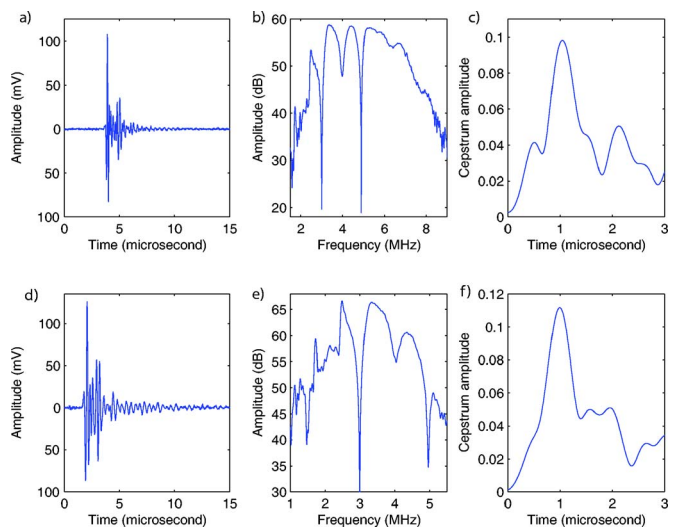


FIG. 3. (Color online) rf signals (left), log spectrum (middle), and cepstrum (right) obtained from a seed with a 5-MHz center-frequency transducer (top) and a 3.5-MHz center-frequency transducer (bottom).

1 μ s may be typical of (and possibly specific for) these titanium-shelled, Pd, brachytherapy seeds. This determination achieved our initial aim by establishing that a signature in fact exists in the form of the distinctive pattern of echo-signal repetitions that we observed. Accordingly, the second aim of this study was to determine whether seeds can be detected and located by using a signal-processing technique that is able to recognize specific repetitions in the echo signals from seeds.

III. METHODS

A. Approaches considered

We considered four approaches for sensing characteristic signal repetitions and specifying their locations: cepstral, autocorrelation, spectral autocorrelation, and singular-spectral methods.

The use of cepstra to detect and characterize repetitions in signals that arise from repetitive spacings in tissue architecture is not new to the medical ultrasound community. Cepstra have been used for tissue characterization for more than two decades. In particular, cepstra were used for estimation of mean-scatterer spacing, which is a diagnostically significant parameter for diffuse liver diseases.¹² Several methods subsequent to the original cepstral approach developed by Lizzi *et al.*¹¹ were applied successfully. Wear also showed that the use of autoregressive analysis to derive the spectrum for cepstrum analysis improved the accuracy of the cepstral estimate of the mean-scatterer spacing (i.e., provided a lower bias and smaller variance) with a smaller time gate (i.e., improved spatial resolution).¹³ The cepstrum is a valuable tool for detecting repetitions in signals (as illustrated in Fig. 3), but cepstra have several drawbacks that make them unsuitable for seed imaging. First, cepstrum computation is sensitive to the bandwidth of the log spectrum used prior to Fourier inversion. For example, to compute the cepstra displayed in Figs. 3(c) and 3(f) only the 2–7.5 and 1–6 MHz bandwidths of the log spectra displayed in Figs. 3(b) and 3(e) were used, respectively. Frequency-dependent attenuation decreases the usable bandwidth for computing the cepstrum as a function of depth into tissue, and thus obtaining reliable cepstra may necessitate schemes to correct for bandwidth reduction as a function of depth. Second, specific to our imaging problem, detecting a maximum near 1 μ s and then applying a decision criterion for the presence or absence of a seed would be difficult. Third, the presence of an unambiguous cepstral peak near 1 μ s would establish only that a repetition exists, but it would not establish the location of the seed within an A line, as this study seeks to provide. Using a sliding window to compute short-time cepstrum would be computer intensive. Furthermore, cepstrum quality degrades quickly with decreasing signal length.¹³

Autocorrelation functions are a second candidate method for detecting and characterizing periodicity in tissue. Wear recently compared cepstral and autocorrelation approaches for evaluating the cortical thickness of the tibia.¹⁴ The thickness of the cortical layer of bone is an indicator of fracture risk, and ultrasound offers a less-expensive, nonion-

izing alternative to x-ray methods. In his study, Wear demonstrated that both techniques performed comparably for this application.

Varghese and Donohue refined the autocorrelation technique by introducing the spectral autocorrelation function.^{15,16} Spectral autocorrelation is a third candidate for assessing signal periodicity and Varghese and Donohue showed it to be clearly superior to basic autocorrelation methods when strong, diffuse (i.e., unresolved) scatterers were present. In particular, they were able to obtain reliable estimates using shorter signals which provided improved spatial resolution.

Singular spectrum analysis (SSA) is a fourth candidate method for detecting and characterizing signal repetitions. Vautard *et al.* introduced the SSA approach and successfully used it for atmospheric sciences.^{17,18} The SSA method was used to extract trends in periodicity from 130 years of temperature data.¹⁹ Pereira *et al.* applied the SSA approach in ultrasound,²⁰ and it proved to be effective for characterizing properties of trabecular bones.²¹ The SSA method provided information about bone microstructures that could not be appreciated in B-mode images. Furthermore, Pereira established that his SSA algorithm performed at least as well as the advanced autocorrelation methods.

When used alone, none of these methods (cepstrum, autocorrelation, spectral autocorrelation, and SSA) provides localization information, which is needed for clinically useful imaging of brachytherapy seeds. All of these methods return an estimate of the signal-repetition period over a given region of interest (ROI), but all are devoid of information regarding the location of structures of interest within the ROI. In estimating the mean-scatterer spacing of tissue, specific localization information is not expected or needed because the scatterers are smaller than the wavelength and are not resolved. In such applications, the mean-scatterer estimates serve as useful statistical descriptors of the periodic organization of the underlying stochastic tissue architecture. However, to image seeds, we require location information and therefore we are developing a theoretical framework that is based on SSA but that is augmented by a method to retrieve localization information.

B. Singular spectrum analysis

The SSA framework is very well-suited to our problem because it allows detecting repetitions with very little sensitivity degradation due to noise, and a heuristic criterion easily can be formulated to detect specific repetition periods (e.g., in our case near 1 μ s). A good sensitivity to a specific periodic echo-signal pattern is necessary because structures other than seeds may exist in tissue (e.g., liver, prostate, etc.) that generate repetitions in the backscattered signals that need to be distinguished from the repetition signals associated with seeds.

The SSA framework assumes that the time signal is a linear combination of the orthonormal basis that is deduced from the correlation matrix of the time signal. In particular, this orthonormal basis is different for each rf time signal. The diagonalization of the correlation matrix allows for optimal

denoising and easy detection of signal repetitions. Theoretical proofs of these mathematical statements can be found in Vautard and Ghil,¹⁷ and they are outside the scope of this paper.

Assume that our time signal, s , is composed of N samples and that a sliding window of $M (< N)$ samples is used to estimate the correlation matrix. In our SSA framework for seed detection, s is the zero-mean envelope-detected backscattered rf signal. The correlation matrix, A , of s is of size M by M and element (i, j) is estimated by

$$A_{i,j} = \frac{1}{N-M} \sum_{k=1}^{N-M} s(k)s(k+|i-j|). \quad (1)$$

Equation (1) shows that values of the matrix elements depend only on the absolute value of $i-j$ (for a given signal s), i.e., the matrix A is Toeplitz and symmetric. The diagonalization of A is the first step of our SSA method. After diagonalization, the eigenvalues are ordered from the largest to the smallest, $\lambda_1 > \dots > \lambda_M$, (the eigenvalues are real and positive because A is symmetric). The eigenvectors, E^1, \dots, E^M , are called the empirical orthonormal functions (EOFs), and the projections of the original time signal onto the EOF yields the principal components (PCs), a^1, \dots, a^M .^{17,19}

Each eigenvalue is equal to the variance of the signal in the corresponding PC direction. Therefore, only the largest eigenvalues are really signal components, and the smallest eigenvalues can be interpreted as noise. In order to filter our time signal, we selected a heuristic criterion to reject the eigenvalues assumed to correspond to PCs of the noise. We decided to discard all the PCs that corresponded to eigenvalues beyond 90% of the total variance of the signal (i.e., the sum of all eigenvalues: $\sum_{i=1}^M \lambda_i$). Eigenvalue λ_k was rejected if

$$k > \text{Min} \left\{ l \quad \text{such that} \quad \frac{\sum_{i=1}^l \lambda_i}{\sum_{i=1}^M \lambda_i} > 0.9 \right\}. \quad (2)$$

A particularly relevant property of SSA is that periodic components (i.e., repetitions) generate eigenvalue pairs.^{17,19} Therefore, from eigenvalues that remained after rejection of noise-associated eigenvalues, we retained only the eigenvalues that existed in pairs. However, problems with automated implementation are that (1) two eigenvalues are not likely to be exactly equal, and (2) noise in the system can create spurious pairs. Therefore, we adopted a heuristic criterion for detecting “actual” eigenvalue pairs: λ_i and λ_j were selected as a non-noisy actual periodic pair if the following conditions are simultaneously true:

- (1) $i < k$ and $j < k$, where k is defined in Eq. (2).
- (2) $|1 - \lambda_i / \lambda_j| < 0.04$.
- (3) The frequencies, f_i and f_j , at which the spectra (i.e., the magnitude of the Fourier transform) of eigenvectors E^i and E^j reach their maxima, respectively, are such that $|1 - f_i / f_j| < 0.03$.

Condition (1) discards all eigenvalues assumed to be associated with noise. Condition (2) allows for 4% tolerance to detect equal eigenvalues. Finally, condition (3) is added be-

cause eigenvalue pairs corresponding to periodic behavior should have EOFs with a very similar spectral content. The set of selected eigenvalues is denoted by Γ .

Following eigenvalue selection, a new time signal, s_r , is constructed from the selected pairs by summing the remaining PCs:¹⁷

$$s_r(i+j) = \sum_{k \in \Gamma} a_i^k E_j^k, \quad (3)$$

where the subscripts i and j represent the i th and j th component of the PC and EOF, respectively. As stated earlier, the PCs are obtained by projection of s over the EOF:¹⁷

$$a_i^k = \sum_{j=1}^M s_{i+j} E_j^k. \quad (4)$$

C. Seed detection

The signal, s_r , constructed from the selected eigenvalue pairs is used first to determine seed presence and location. According to the selection criteria applied to the eigenvalues, the possibility exists that no eigenvalue pair can be selected. In that case, s_r vanishes, and the algorithm determines that no seed is present in this time signal.

If eigenvalues are selected, the Fourier transform, $S_r(f)$, of s_r is computed. The spectral power, P , in the bandwidth of interest (i.e., that of the repetition frequency) is computed:

$$P = \int_{f_0}^{f_1} |S_r(f)|^2 df. \quad (5)$$

P represents the power of the periodic part of the signal after optimal denoising in the frequency range close to that expected to be generated if a seed were present. Therefore, P can be interpreted as a score for the likelihood of a seed being present. However, P is specific to the entire portion of the original signal, s , within the analysis window, and information on the location of the seed within the window does not exist; that is, the seed signal can be located anywhere in the analysis window. We observed experimentally that the first part of a repetition signal has a greater magnitude than the second part, and we determined that the seed actually is located at the first echo. This is illustrated in Fig. 3(a). Therefore, we assume that the location of the seed is where s_r achieves its maximum amplitude.

To obtain a color-coded image that depicts seed location based on our detection criteria, each backscattered signal is evaluated using a window size of N samples over which the SSA processing is conducted and leads to the assignment of a P value. This P is displayed at the location where s_r achieves its maximum amplitude. The window is then shifted to the next location and the computation is repeated. Windows are designed to provide an 87.5% overlap between two consecutive windows. After the SSA processing has been conducted over the entire A line, the algorithm moves to the next A line. Processing ends when all the A lines are analyzed. The resulting P map is encoded in false color and imposed over the conventional B -mode image. This color-coded image is called a P -mode image. Higher values of P

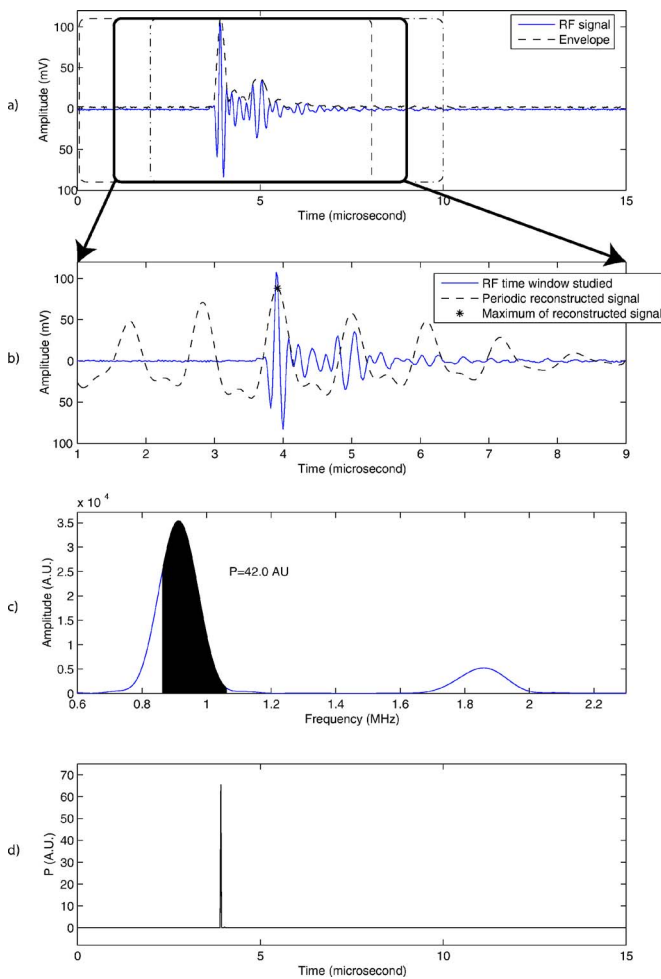


FIG. 4. (Color online) Illustration of the proposed SSA method. (a) rf signal, envelope, and three 8- μ s windows with 7- μ s overlap (i.e., 87.5%). (b) Reconstructed signal from the SSA method. (c) Spectrum of the reconstructed signal, and illustration of the evaluation of Eq. (5). (d) P line.

indicate a higher likelihood of the presence of a seed. A discussion of how to choose and interpret the values of N , M , and the overlap percentage is provided in Sec. V.

To illustrate the method, we applied it to the center A line of the B -mode image of Fig. 2. The corresponding rf signal is shown in Fig. 3(a). The first N samples were analyzed within the window indicated by the dashed rectangle in Fig. 4(a); in this case, N is 400 and the sampling frequency was 50 MHz, so that 400 samples span 8 μ s. M was chosen to be the half of N (i.e., 200 samples and 4 μ s). The method is illustrated on the second 400-sample window with 87.5% overlap (i.e., spanning samples 51–450) as indicated by the bold rectangle in Fig. 4(a). The 200-by-200 correlation matrix of this signal was estimated using Eq. (1), diagonalized, and eigenvalues were ordered from the largest to the smallest. Then, we detected the cut-off eigenvalue by finding the lowest value of index, k , such that Eq. (2) is true. For this example, k was 17. Among these 17 eigenvalues, two pairs of eigenvalues met the three requirements of Sec. III B. The reconstructed signal using these two pairs of eigenvalues is shown in Fig. 4(b). The asterisk indicates where the reconstructed signal achieved its maximum amplitude.

The spectrum (i.e., magnitude of the Fourier transform)

of the reconstructed signal is shown in Fig. 4(c). For this signal of a seed in a gel pad, the power integration of Eq. (5) was conducted over a frequency range of 0.86–1.06 MHz because we experimentally noticed a repetition period of approximately 1.04 μ s (which corresponds to a repetition frequency of 0.96 MHz) as illustrated by the location of the cepstral peak in Fig. 3(c). The dark area under the curve in Fig. 4(c) depicts this integration, and the value of P for this segment was 42.0 in arbitrary units (AU). This value was added to the P plot at time sample 3.92 μ s because it was the time sample at which s_r reached its maximum, as denoted by an asterisk in Fig. 4(b).

The algorithm continued by analyzing the next N -point window with 87.5% overlap (i.e., samples 101–500) and repeated itself until the entire signal was processed. The output of the algorithm was a P line, i.e., a signal with a length equal to the length of the original signal and with values indicating the relative likelihood of a seed being present. This P line is shown in Fig. 4(d). It contained a very sharp peak near 3.94 μ s. The amplitude of the peak was 65.5 AU, which is greater than 42.0 AU because the other N -sample windows also contributed to the P value at this time point. Every A line was analyzed; the resulting P lines were merged together to form a two-dimensional matrix of P values. Then, the P values were normalized, log-compressed, color-encoded, and superimposed on the B -mode image to form the P -mode image. (The corresponding P -mode image is presented in greater detail in Sec. IV.)

IV. RESULTS

A. Simulations

To demonstrate and validate the proposed detection method, we conducted several simulations with increasing complexity and levels of noise. All the simulations were performed using a single, computer-generated rf signal (i.e., one A line) over which our SSA algorithm was conducted. The output of the simulation was a P line, i.e., the plot of P as a function of time. For all these simulations, the P integration was conducted over the frequency range of 0.9–1.1 MHz because we were looking for a repetition period near 1 μ s.

1. One repetition and white noise

The purpose of this first set of simulations was to assess the robustness of the method in the presence of white noise. A rf signal that contained an obvious repetition was generated. It contained two, five-cycle, Hanning-windowed pulses at a center frequency of 5 MHz. The first five-cycle pulse had an amplitude of 1 while the second had an amplitude of 0.4. These numbers were consistent with what is observed in Fig. 3(a). The two pulses were separated by 1 μ s; the first five-cycle pulse always started at 4.25 μ s, and its peak was at 4.5 μ s. Seven simulations were conducted. The first simulated signal was free of noise, and increasing amounts of white noise were added to the simulated signal in subsequent simulations. The signal-to-noise ratio (SNR) in noise-containing signals ranged from 28 to 12.4 dB. For each case, ten signals were generated and the resulting ten P lines were

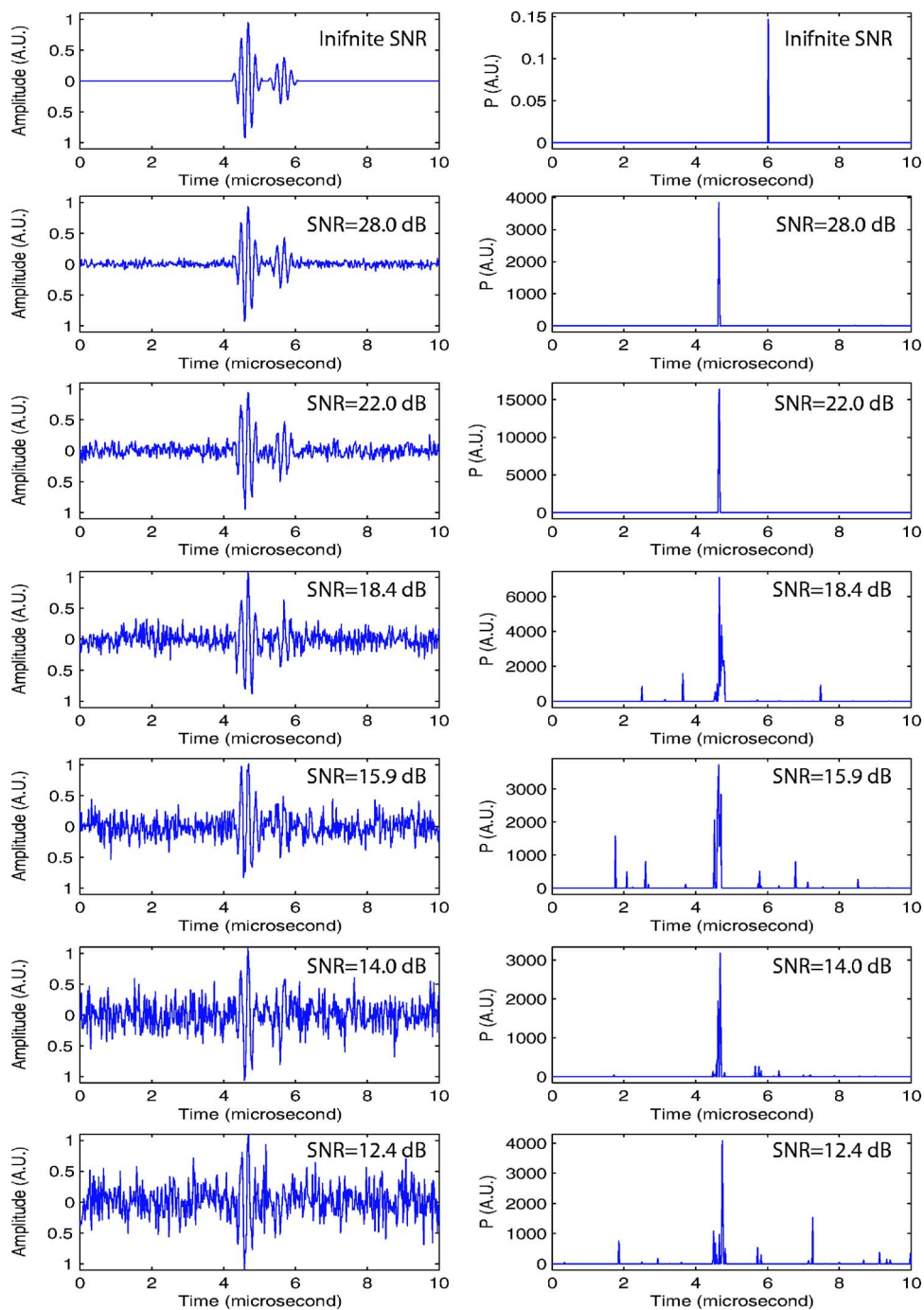


FIG. 5. (Color online) First set of simulations: One repetition in white noise. The left column displays one realization at a specific SNR, the right column displays resulting averaged P lines from ten realizations at the same SNR.

averaged. The left side of Fig. 5 shows examples of signal realizations with a prescribed SNR, and the right side shows the corresponding, averaged P line.

The first striking observation that can be made from these results is that the worst case was actually the noise-free case for which the P line virtually vanished at every sample; in this case, the peak at $6 \mu\text{s}$ was several orders of magnitude smaller than the other peaks in the other cases. The reason for this virtually-zero P line was our criterion for the selection of eigenvalues in which we discarded eigenvalues that were beyond 90% of the total variance of the signal (selection criterion 1), but for a noise-free signal, all eigenvalues are meaningful and important descriptors. Consequently, important information about the noise-free signal

was lost and s_r did not contain a significant amount of spectral energy near the repetition frequency of 1 MHz.

The next two P lines (SNRs of 28.0 and 22.0 dB) showed excellent results; a single and sharp peak existed near $4.5 \mu\text{s}$, which is where the maximum amplitude of the original signal is located. These two cases show that the algorithm is robust in the presence of noise, and it easily achieves excellent results even with only a moderate SNR.

The last four P lines showed less-satisfactory results. In particular, two noise effects degraded the results. First, the presence of noise led to the appearance of several side peaks that did not correspond to the actual repetition. Second, the peaks that did correspond to the repetition were broader, i.e., P values spread to the sides and widened the peak. This

leakage had the detrimental effects of decreasing the effective SNR (the peak magnitude was lower, i.e., closer to the noise plateau), and degrading the resolution of the reconstructed P -mode image (i.e., it introduced uncertainty in the seed location depicted by the P -mode image). Nevertheless, the presence of a peak near the actual repetition location demonstrates that the algorithm continues to perform at these low SNRs. However, the possibility exists that the side peaks presented by signals with very low SNRs could cause false-positive seed determinations in some P -mode images.

This set of simulations established that the threshold used for discarding the eigenvalues needs to be a function of noise. The lower the noise level is, the closer the threshold should be to 100% so that the greatest amount of information is retained. However, a number of eigenvalues express the noise, and therefore the threshold needs to be reduced as the SNR becomes lower so that the noise can be separated effectively from the useful information. The threshold value of 90% adopted here was selected simply by trial and error on several seed-imaging cases. Nevertheless, this set of simulations demonstrates that the algorithm behaves very well at reasonable noise levels (e.g., for a SNR > 22 dB), and we expect that the algorithm will perform satisfactorily with the SNRs of signals acquired under clinical conditions.

2. Two different repetitions and white noise

The purpose of this second set of simulations was to evaluate how the algorithm performs when a repetitive signal from a structure other than the seed is present in addition to the signal from the seed. For this set of simulations, the noise was kept at a constant SNR of 22.0 dB. This noise level was chosen based on the results of the previous simulations. The signal consisted of two repetitions: the first one was the same as in the seed-based signal used in the previous simulation, i.e., two, five-cycle pulse with a 1 μ s time difference. The second, nonseed repetition also was composed of two, five-cycle pulses; for the nonseed signal, the first five-cycle pulse was always located at 14 μ s, the second pulse was located from 14.6 to 16.4 μ s in 0.2- μ s increments giving time differences that ranged from 0.6 to 2.4 μ s (i.e., the repetition frequency ranged from approximately 1.67 to 0.42 MHz).

Figure 6 presents the results in the same format that was used for the previous set of simulations. The left side of Fig. 6 shows one realization of each different case, and the right shows the averaged, reconstructed P line. Figure 6 demonstrates that the selectivity of the algorithm seems excellent because it shows no significant P values near 14 μ s except for the two cases with delays of 0.8 and 1 μ s for the second repetition signal. For the 1 μ s, these relatively large P values occurred because the interval between pulses was within the repetition band (0.90–1.10 MHz) to which the algorithm was sensitive. Just as in the previous set of simulations, P leakage existed into the location of the second cycle (at 15 μ s) of the repetition. This probably was an effect of noise. The presence of a strong peak for the case in which the delay was 0.8 μ s was surprising because the algorithm should not be sensitive to this repetition frequency (i.e., $1/0.8 \mu\text{s} = 1.25 \text{ MHz}$ is outside the integration range of 0.9–1.1 MHz). The existence of this peak means that the

algorithm found significant energy near 1 MHz; it also suggests that the combination of two, five-cycle bursts led to the creation of a spurious energy component near 1 MHz.

This set of simulations demonstrates that the SSA algorithm generally is very selective and capable of differentiating among repetitions having different frequencies. These capabilities are important for many clinical, imaging situations where periodic scattering structures may exist that are regularly organized and capable of creating systematic repetitions in the rf signals.

3. One repetition and speckle noise

The purpose of this third set of simulations was to test the SSA approach in a setting that would occur when imaging biological tissues. For this set of simulations, speckle noise was added to the simulated signal containing the original repetition. Speckle was obtained by acquiring signals from degassed beef. We used the 5-MHz transducer and a sampling frequency of 50 MHz; these parameters are the same as those used for generating the simulated repetition. A signal with a single repetition was generated, and the acquired “speckle noise signal” was added. We also added white noise so that the SNR was kept at 22 dB. We varied the signal-to-speckle-noise ratio (SSNR) from 60 to –6 dB. We defined the SSNR as the ratio of the maximum of the simulated signal to the standard deviation of the speckle noise signal.

Figure 7 shows the results. The 60-dB SSNR case was essentially equivalent to the case of Fig. 5 for which the SNR was 22 dB. The seed is clearly detected and no artifacts or false-positives are found on the P line. The next two cases (i.e., SSNR of 26.0 and 14.0 dB) showed satisfactory results with a large peak at the actual seed location; however, a smaller, but significant, peak also was present at a location far away from the actual seed (near 9 μ s). The 6-dB case shows an important peak at the location of the second part of the repetition (near 6 μ s). This phenomenon also is apparent on certain P lines in Fig. 5. The –1.22-dB case showed several peaks near 4, 6, 8, and 9 μ s. The algorithm apparently broke down at this SSNR. Surprisingly, fairly satisfactory results were obtained for the last two cases with even lower SSNRs.

This set of simulations demonstrates that the algorithm is quite robust even in the presence of speckle noise acquired from biological tissues. However, it also shows that, when a repetition is present, it may assign a high P value to the second pulse of the repetition, which in turn may result in a P -mode image showing a seed very slightly deeper than its actual location. (Fortunately, the resulting error would be less than a millimeter, and would not affect the delivered dose in a clinically significant manner.)

B. Seed in gel pad

We evaluated the SSA method in the ideal case of a 5-MHz scan of a seed in an acoustically transparent gel pad (as shown in Fig. 2). The reconstructed P -mode image is displayed in Fig. 8(a). This image was reconstructed by retaining the P values that fall within a 40-dB dynamic range.

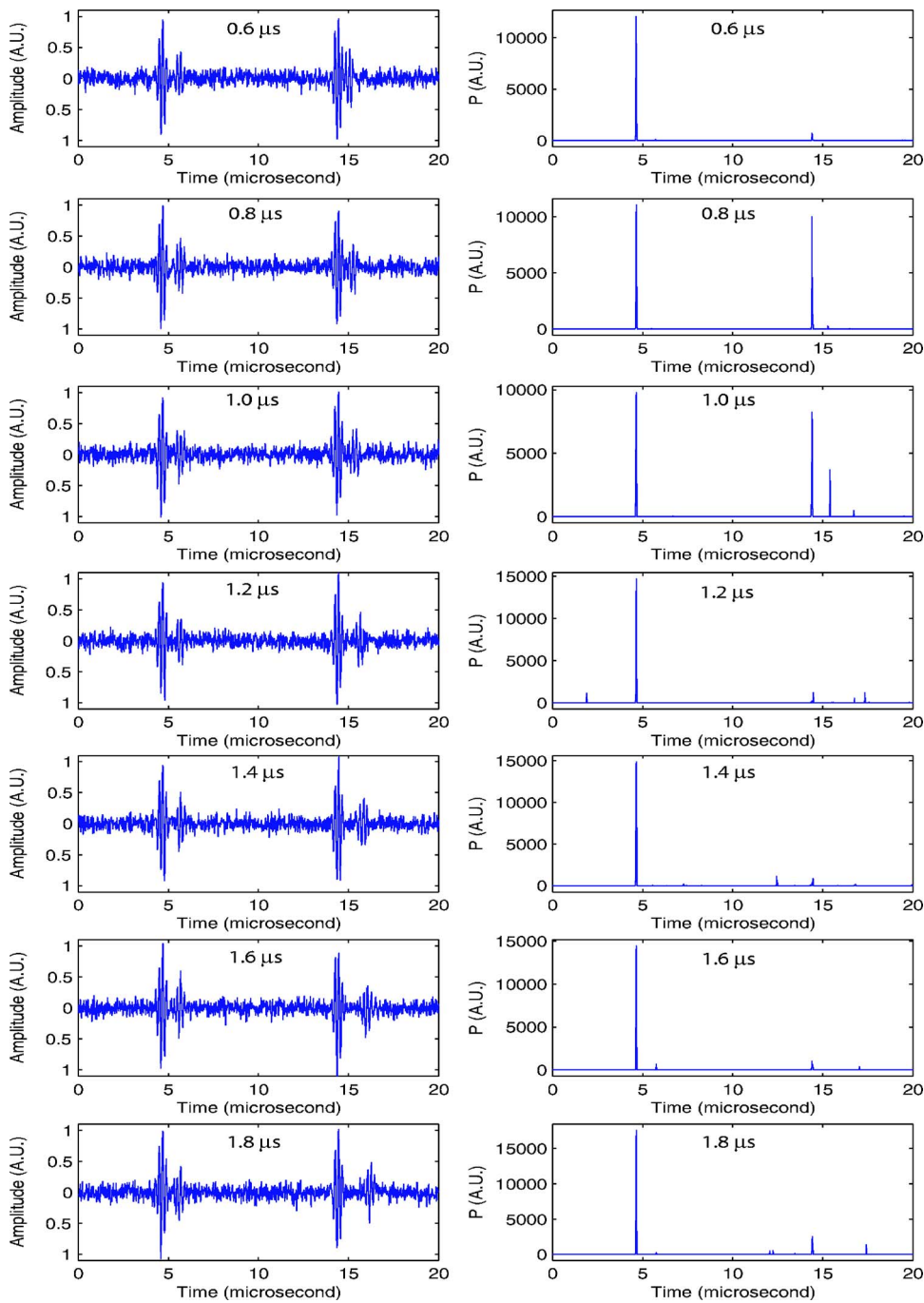


FIG. 6. (Color online) Second set of simulations: Two repetitions in white noise. The left column displays one realization for specific time difference for the second repetition, the right column displays resulting averaged P lines from ten realizations. SNR was kept at 22 dB.

Specifically, if a value of P was outside that dynamic range, then it was replaced by the B -mode data. As shown in Fig. 8(a), all of the P values in the 40-dB dynamic range are very well localized in the red rectangle. For this case, basically no leakage of P values occurred outside the actual seed location. Unfortunately, the color-encoded information shown in Fig. 8(a) is difficult to appreciate; Fig. 8(b) shows an expanded version of the red rectangle. However, even in the expanded version, the color-coded information remains somewhat difficult to assess.

To improve the reconstructed P -mode image, we consecutively applied two nonlinear image-processing steps. First, we used a morpho-mathematical operation called dilation with a square mask of side length $100\ \mu\text{m}$. This operation essentially expands the color-coded region. A more-

mathematical description of dilation is outside the scope of this paper, but can be found in Serra.²² The resulting dilated image is shown in Fig. 8(c). Second, we applied a median filter with a square mask of side length $200\ \mu\text{m}$. This operation reduced the noise and also allowed for the “shadow” of the seed (i.e., the lower intensity P values contained below the main line of large P values in the red rectangle) to be somewhat mitigated. In some cases, the seed shadow appears because the reconstructed signal has its maximum at the second repetition echo complex instead of the first one. However, the shadow P values are less connected so that the median filter is able to reduce their effect as shown in Fig. 8(d). In the remainder of this paper, the displayed P -mode images always include these two image-processing operations.

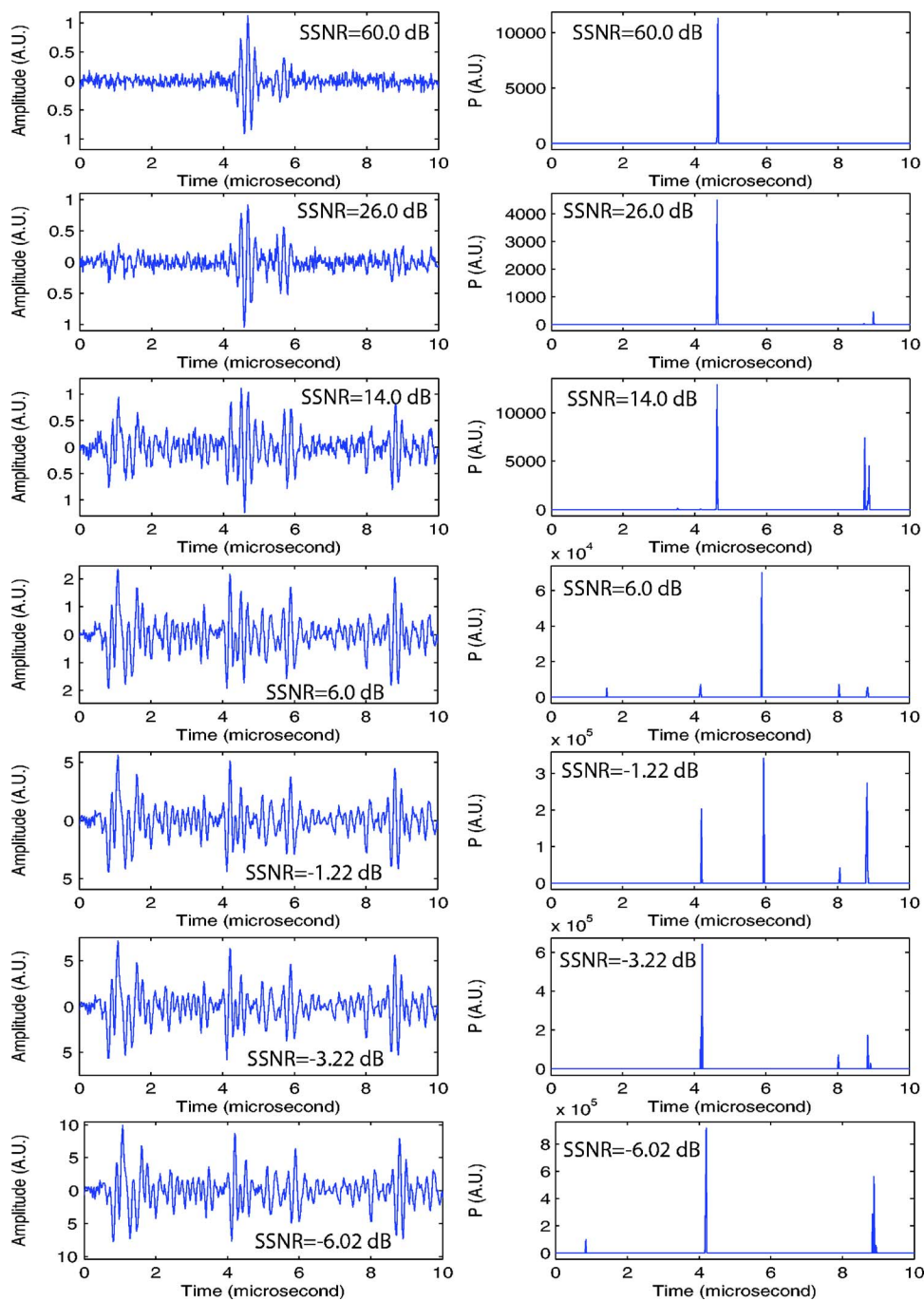


FIG. 7. (Color online) Third set of simulations: One repetition in speckle noise. The left column displays one realization at a specific SSNR, the right column displays resulting averaged P lines from ten realizations at the same SSNR. Speckle noise signal was acquired with 5-MHz center frequency transducer from a degassed beef steak.

The final P -mode image [Fig. 8(d)] shows excellent results, the seed is clearly visible and the SSA algorithm did not produce any false-positive indications. The shadow persists, but with P values approximately 20 dB lower than those at the true seed location. This example shows that the SSA method performs very well for this controlled case.

C. *Ex vivo* experiment

The last step of this preliminary study was to evaluate the SSA algorithm *ex vivo*. A seed was implanted in a piece of fresh, degassed beef, and several planes were scanned with the 3.5-MHz transducer and processed using our SSA method. In this setting, the true location of the seed was known because it was implanted into the piece of beef with a

needle after the beef was degassed. This allowed for direct assessment of the method. The experiment also is interesting because biological tissues are strong generators of speckle and also the small, spatial variations of acoustical properties that exist in tissue create scattering and phase aberrations. This experiment was meant to evaluate the performance of the method in a more-challenging setting that is more representative of conditions to be encountered in clinical applications.

Figure 9 shows a P -mode image obtained *ex vivo*. The seed location is indicated by an arrow. The image shows that, for this more-challenging case, P -value leakage clearly exists as indicated by the scattered light to dark blue squares present on the image. However, the leakage P -values remain

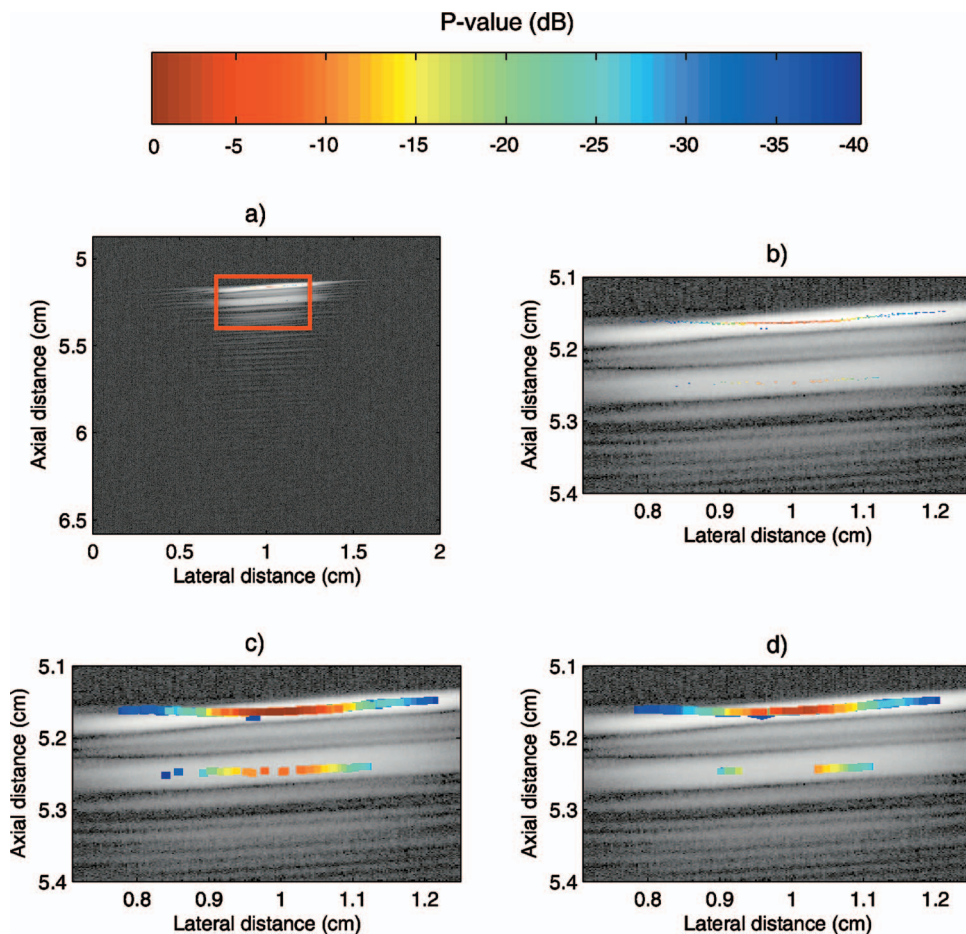


FIG. 8. Images of a single seed in an acoustically transparent gel stand-off pad. (a) P -mode image of a seed in gel pad. (b) Zoomed-in version of red rectangle of (a). (c) Dilation of (b). (d) Median-filtered version of (c).

low (< -30 dB); therefore, the leakage is recognized clearly as noise. In Fig. 9, the largest P value is at the actual seed location, and we can state confidently that, in this image, the seed is localized using our SSA method. However, in the same image, significant P values are scattered along the water-meat interface. In particular, a strong (about -3 dB) P value is displayed at an axial distance of approximately 3.7 cm and a lateral distance of 2.25 cm. In the absence of prior knowledge, this component might be interpreted as a seed, and if so, would represent a false-positive. However, in the present case, prior knowledge of the water-meat boundary allows for easy dismissal of this potential false-positive. The only other possible false positive is located at an axial distance of 4.4 cm and a lateral distance of 2.1 cm. The P values at that location are about -10 dB so that even though significantly lower than those of the actual seed, this connected component possibly might be interpreted incorrectly as a seed.

In practice, *a priori* knowledge would be applied on a three-dimensional (3D) basis using adaptations of current clinical dosimetry-planning software to depict apparent seed locations in 3D. For example, if 40 seeds were implanted and 43 seeds were depicted, but 40 were shown to be within the gland and 3 were shown to be outside the gland, then a reasonable initial determination would be that all seeds were within the gland and their locations would be used to compute the actual dose distribution; a dose correction could be made by adding seeds if necessary. Similarly, if 40 seeds

were implanted and 43 seeds were depicted, but only 37 were within the gland while 6 were shown to be outside it, then an inadequate dose to the gland would be established and a correction could be made immediately.

V. DISCUSSION

The proposed algorithm appears to be an attractive choice for seed localization and imaging because it allows seed detection and depiction to be performed independently of transducer technology or beam-forming algorithms. These advantages obviously are attractive for real-time clinical imaging of the prostate. However, the top-priority remaining task in developing this algorithm is optimization of the design parameters. These parameters include eigenvalue-pair selection [conditions (1)–(3) in Sec. III B], integration frequencies (f_0 and f_1) for the computation of P in Eq. (5), and windowing parameters (N , M , and the overlap percentage, OP).

When making the choice of N , M , and OP, the first obvious engineering consideration is processing time. Smaller values of N and M , and larger values of OP require longer times for the algorithm to complete. In this study, computation time was not considered critical, and the algorithm implementations in MATLAB (Mathworks Inc., Natick, MA) took several seconds for treating a single A line with the initially used window parameters ($N=400$, $M=200$, and $P=87.5\%$). Using the current MATLAB implementation, pro-

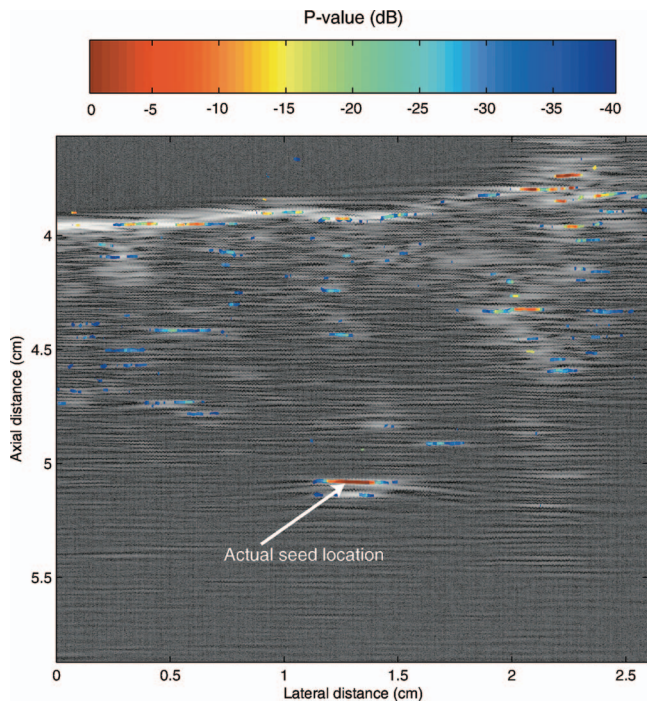


FIG. 9. P -mode image obtained with a 3.5-MHz center-frequency transducer of a seed implanted in a piece of degassed beef.

cessing a complete, conventional, clinical, B -mode image to generate the color-coded P -mode image would require several minutes. We are confident that the algorithm can be implemented in a more-rapid manner (e.g., using C++), and that available computer technology will allow for clinically practical, real-time computation of P -mode images.

More important is making the optimal choice of N and M with respect to the repetition frequencies to which the algorithm is sensitive. N must be sufficiently large to span the anticipated range in repetition period within the N -sample signal; M must be sufficiently large to encompass a single cycle of the repetition. Thus, to properly design the SSA implementation, the nature of repetitions to be detected must be determined with care. A statistical interest also exists in having N and M as large as possible so that the accuracy of the correlation estimation described in Eq. (1) is maximized.

The choice of OP can be related to the SNR of the resulting P line. In particular, a larger OP will lead to a higher SNR. Specifically, if a sufficiently large overlap exists, then a repetition will be contained in several consecutive windows; this effect will produce a high combined P value at the true seed location. If noise is dominant, a very broad noise spectrum would lead to the production of several eigenvalue pairs with spectral peaks near the repetition frequency. However, because of the random nature of noise, the spurious P values arising from noise are unlikely to combine at the same time sample. Thus, the spurious P values are most likely spread randomly and simply lead to an increased SNR of the P line. The tradeoff in the choice of OP is in effect a tradeoff between computational time and SNR in the reconstructed P line.

As demonstrated in the simulations, the eigenvalue-selection threshold is related to noise and numerical analysis. The eigenvalue cut-off threshold can be related to the SNR in the system, and it easily can be tuned to optimize sensitivity and specificity. Furthermore, according to the first set of simulation results, threshold selection is fairly robust on a large range of SNR. (We used a 90% threshold value for all the presented analyses and results.) In the absence of noise, the eigenvalues of a periodic process theoretically are exactly equal if the correlation matrix is estimated accurately. Noise has two effects: it can separate actual pairs, and it can create spurious pairs. We did not experience difficulty in tuning these threshold parameters, but the problem is “discontinuity” in the processing. A discontinuity occurs when an infinitesimal change in the data causes a large change in the results; this occurs if a pair of eigenvalues near the threshold either is selected or discarded thus resulting in a different reconstructed signal and hence a different P value. Therefore, P is not a continuous function of the data, which potentially could cause robustness issues for the algorithm in certain “close” cases.

Finally, the integration window [i.e., f_0 and f_1 in Eq. (5)] must be specified by studying signals from seeds in a controlled environment, as we did in our laboratory experiment. A reasonable possibility is that look-up tables of values for f_0 and f_1 could be designed *in vitro* and then input to medical scanners for *in vivo* imaging.

Frequency-dependent attenuation may affect the robustness of the algorithm because the seed signature may change as a function of depth. The effect of frequency-dependent attenuation currently is under investigation. Nevertheless, we anticipate that the algorithm can be implemented with depth-dependent parameters to correct for possible changes in seed signature as a function of depth.

A remaining step in refining the method is to evaluate its dependence on seed orientation. While the implanted seed axis ideally is orthogonal to the transducer beam axis and scan plane, orthogonality rarely is maintained, and variations exist in seed orientation with respect to the scan plane. The cylindrical geometry and highly specular reflective symmetry of the seed may make the efficacy of the SSA method depend upon the angle between the seed and the beam axis. We have begun to examine cases where the seed orientation was varied, and have found that the repetition period may be a function of the angle between the seed and beam axes. Thus, evaluation of angle dependence of the method is being strongly pursued, and results will be reported in subsequent papers.

The physical origin of the repetition signal backscattered from a seed also is under investigation. We have not yet determined conclusively what the origin of this signal is. However, experiments with a range of frequencies extending from 3 to 25 MHz showed that the repetition signal occurred independent of frequency. These preliminary measurements did not show any statistically significant differences among the seed repetition-signal periods at different insonification frequencies. Previous unpublished studies by Feleppa *et al.* using 5, 10, and 15 MHz suggested that the repetition period is not frequency-dependent. Because all our measurements

led to repetition periods in the narrow range of 0.8–1.2 μ s, we conclude that if the repetition period is a function of frequency, then the dependence is not strong. A physical understanding of the cause of the signal periodicity may permit improvement of the SSA algorithm. Therefore, we are pursuing this understanding actively.

A necessary extension of the completed study is assessing errors made by the algorithm, specifically, an evaluation of the false-positive and false-negative determinations. Certainly, error management can be improved by adding *a priori* knowledge to the detection algorithm. The clinician knows how many seeds are injected. This number can be input into the algorithm to limit the algorithm to detecting a fixed number of seeds within or near the gland. Even more important, although misplacement of seeds occurs because of movement and distortion of the gland, the clinician always knows approximately where the seed should be located. In fact, clinical dosimetry software currently utilizes 3D depictions of the gland shape to compute optimal needle loadings and seed placements. The software-prescribed locations can be exploited to engineer a penalty for finding a seed very far from the prescribed locations.

After algorithm development is complete, the algorithm will be tested over a large number of actual clinical cases to obtain receiver operator characteristic curves that will provide an objective assessment of algorithm performance under actual clinical conditions.

ACKNOWLEDGMENTS

The authors thank Dr. W. C. A. Pereira for many stimulating and informative discussions regarding SSA and Dr. C. S. Wu for his insights and perspectives regarding current methods of seed placement and dosimetry in prostate brachytherapy. We also thank E. Girard for her help in the experimental aspects of these studies. This research is supported in part by NIH Grant No. CA098465.

¹H. H. Holm, N. Juul, J. F. Pedersen, H. Hansen, and I. Stroyer, "Transperineal ¹²⁵Iodine seed implantation in prostatic cancer guided by transrectal ultrasonography," *J. Urol. (Baltimore)* **130**, 283–286 (1983).

²J. C. Blasko, H. Ragde, and D. Schumacher, "Transperineal percutaneous iodine-125 implantation for prostatic carcinoma using transrectal ultrasound and template guidance," *Endocrine. Hypertherm. Oncol.* **3**, 131–139 (1987).

³P. Niehoff, B. Ballardini, C. Polgar, T. Major, J. Hammer, A. Richetti, and G. Kovacs, "Early European experience with the mammosite radiation therapy system for partial breast brachytherapy following breast conserva-

tion operation in low-risk breast cancer," *Breast J.* **15**, 319–325 (2006).

⁴P. Niehoff, C. Polgar, H. Ostertag, T. Major, Z. Sulyok, B. Kimmig, and G. Kovacs, "Clinical experience with the mammosite(r) radiation therapy system for brachytherapy of breast cancer: Results from an international phase ii trial," *Radiother. Oncol.* **79**, 316–320 (2006).

⁵R. R. Patel and R. K. Das, "Image-guided breast brachytherapy: An alternative to whole-breast radiotherapy," *Lancet Oncology* **7**, 407–415 (2006).

⁶J. Crook, M. McLean, C. Catton, I. Yeung, J. Tsihlias, and M. Pintilie, "Factors influencing risk of acute urinary retention after trus-guided permanent prostate seed implantation," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 453–460 (2002).

⁷Z. Wei, M. Ding, D. Downey, and A. Fenster, "3d TRUS guided robot assisted prostate brachytherapy," *Medical image computing and computer-assisted intervention* **8**, 17–24 (2005).

⁸D. H. Brinkmann and R. W. Kline, "Automated seed localization from ct datasets of the prostate," *Med. Phys.* **25**, 1667–1672 (1998).

⁹H. Liu, G. Cheng, Y. Yu, R. Brasacchio, D. Rubens, J. Strang, L. Liao, and E. Messing, "Automatic localization of implanted seeds from post-implant ct images," *Phys. Med. Biol.* **48**, 1191–1203 (2003).

¹⁰M. E. Miquel *et al.*, "Using combined x-ray and mr imaging for prostate i-125 post-implant dosimetry: Phantom validation and preliminary patient work," *Phys. Med. Biol.* **51**, 1129–1137 (2006).

¹¹F. Lizzi, E. Feleppa, N. Jaremko, D. King, and P. Wai, "Liver-Tissue Characterization by Digital Spectrum and Cepstrum Analysis," *Proceedings of the 1981 IEEE Ultrasonics Symposium*, Chicago, IL, 1981, pp. 575–578.

¹²D. L. King, F. L. Lizzi, E. J. Feleppa, P. M. Wai, M. M. Yaremko, M. C. Rorke, and J. Herbst, "Focal and diffuse liver disease studied by quantitative microstructural sonography," *Radiology* **155**, 457–462 (1985).

¹³K. Wear, R. Wagner, M. F. Insana, and T. J. Hall, "Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **40**, 50–58 (1993).

¹⁴K. A. Wear, "Autocorrelation and cepstral methods for measurement of tibial cortical thickness," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50**, 655–660 (2003).

¹⁵T. Varghese and K. D. Donohue, "Mean-scatterer spacing estimates with spectral correlation," *J. Acoust. Soc. Am.* **96**, 3504–3515 (1994).

¹⁶T. Varghese and K. D. Donohue, "Estimating mean scatterer spacing with the frequency-smoothed spectral autocorrelation function," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **42**, 451–463 (1995).

¹⁷R. Vautard and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," *Physica D* **35**, 395–424 (1989).

¹⁸D. Broomhead and G. King, "Extracting qualitative dynamics from experimental data," *Physica D* **20D**, 217–236 (1986).

¹⁹R. Vautard, P. Yiou, and M. Ghil, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signals," *Physica D* **58**, 95–126 (1992).

²⁰W. C. Pereira and C. D. Maciel, "Performance of ultrasound echo decomposition using singular spectrum analysis," *Ultrasound Med. Biol.* **27**, 1231–1238 (2001).

²¹W. C. Pereira, S. L. Bridal, A. Coron, and P. Laugier, "Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancellous bone specimens," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **51**, 302–312 (2004).

²²J. Serra, *Image Analysis and Mathematical Morphology* (Academic, New York, 1982).

Erratum: “Constrained comparison of ocean waveguide reverberation theory and observations” [J. Acoust. Soc. Am. 120(4), 1922–1931 (2006)]

Charles W. Holland

The Pennsylvania State University, Applied Research Laboratory, State College, PA 16801

(Received 21 November 2006; revised 19 December 2006; accepted 20 December 2006)

[DOI: 10.1121/1.2434755]

PACS number(s): 43.30.Gv, 43.28.Fp, 43.10.Vx [RCG]

On Pg. 1923 4th line from bottom $\varepsilon_{m,n}$ should read $\varepsilon_n \varepsilon_m$

On Pg. 1924, first full sentence should read

Another way to say this is that the incoherent range-averaged monostatic reverberation considered in this paper, while very sensitive to the angular dependence of the backscatter, is relatively insensitive to the angular dependence of the vertically bi-static paths.